Modern statistical approaches for epidemiological modelling and uncertainty quantification

Nicholas Steyn

Department of Statistics

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2025

Supervised by:

Prof. Christl A. Donnelly

Dr. Kris V. Parag

Acknowledgements

Christl, you have been the most incredible supervisor, mentor, and friend. Your unwavering support, unlimited patience, extraordinary expertise, and incredible attention to detail are appreciated more than I can express. **Kris**, your readiness to entertain every wild idea I had, and then always finding a way to make them work, has been the spark behind much of this thesis and my enjoyment of it. Thank you both.

Team Donnelly - I cannot imagine a better group of people to have spent the last four years with. Thank you for the laughter, the technical discussions, and your support throughout this journey. **Ruth**, you are the reason for my enduring sanity. **Matt**, **Cathal**, **Tarek**, and **Sarah**, your friendship, local and global adventures, and readiness to help with any of my questions will be forever treasured.

The last four years have been defined by the incredible people I have met. Thank you to **Ettie**, **Seth**, **Robin**, and **Marc** for your insight, encouragement, friendship, and generosity with your time. To **Vik**, for fundamentally shaping how I view statistics. To **Zak**, for letting me be a part of your own doctoral journey. Also thank you to the team in the department for supporting me through endless questions: **Emma**, **Joanna**, **Frédérique**, and **Beverley**. Finally, thank you to my tutorial students at Magdalen College and elsewhere for adding some colour to my time here - meeting you has been a true highlight.

My DPhil journey started long before I arrived in Oxford. Without the immense and continuing support of **Mike Plank** and the Te Pūnaha Matatini COVID-19 team, I would never have applied to this degree. You invested so much time and energy into me and I am forever grateful for your support.

Finally, a huge thank you to my parents **Kay** and **Rudi**, and sister **Josie** for your unwavering support both before and during this journey, and for always welcoming me home when Oxford's weather grew too cold.

DONNELLY STATS 22-25

Thank you to Ben Odbert for letting me include this incredible artwork in my thesis. It captures our group dynamic perfectly (although Christl might object to the throne, I think it is well deserved).

Abstract

Advances in statistical methodology and computing power are enabling the development of increasingly sophisticated models of infectious disease dynamics, transforming our understanding of the spread and impact of these diseases. These advances allow us to perform more robust inference, make better predictions, extract more information from existing data sources, and harness novel data sources. As exemplified by the COVID-19 pandemic, such progress is crucial for our highly connected world, which faces an increasing number of infectious disease threats. The need for adaptable, reliable, and innovative statistical methods for the analysis of epidemiological data is clear.

This thesis responds to this need by introducing a series of methodological contributions to the field of infectious disease modelling. In Chapter 3, we present a decision-theoretic approach to uncertainty quantification, linking statistical first principles directly to epidemiological applications. In Chapter 4, we extend two popular epidemiological models to demonstrably improve their uncertainty quantification. Later chapters develop novel methodology for fitting epidemic models to data (Chapter 5), for robust modelling of epidemic prevalence survey data (Chapter 6), and for leveraging wastewater sampling data alongside traditional data sources (Chapter 7). Chapter 8 examines risk-related behaviours in England during the COVID-19 pandemic, and compares inferences made from traditional survey data to those derived from novel data sources. Finally, in Chapter 9, we revisit the models introduced in Chapters 4, 5, and 6 through the framework introduced in Chapter 3.

Contents

List of abbreviations

Abbreviation	Meaning
ABC	approximate Bayesian computation
APE	accumulated prediction error
BUGS	Bayesian inference Using Gibbs Sampling
BvM	Bernstein-von Mises
CAR	case ascertainment ratio
CDF	cumulative distribution function
CI	confidence interval
COVID-19	coronavirus disease 2019
CrI	credible interval
CRPS	continuous ranked probability score
Ct	cycle threshold
EIG	expected information gain
ESR	Institute for Environmental Science and Research
ESS	effective sample size
EUR	expected uncertainty reduction
EVPI	expected value of perfect information
EVSI	expected value of sample information
FI	Fisher information
GLM	generalised linear model
GP	Gaussian process
HIV	human immunodeficiency virus
HMC	Hamiltonian Monte Carlo
HMM	hidden Markov model
i.i.d.	independent and identically distributed
KL	Kullback-Leibler
MALA	Metropolis-adjusted Langevin algorithm
MAP	maximum a posteriori
MCMC	Markov chain Monte Carlo
MLE	maximum likelihood estimate
NPI	non-pharmaceutical intervention
MRP	multilevel regression and poststratification
NUTS	No-U-Turn Sampler
NZ	Aotearoa New Zealand
ONS	Office for National Statistics

OR	odds ratio
OxCGRT	Oxford COVID-19 Government Response Tracker
POMP	partially observed Markov process
PMF	probability mass function
PMMH	particle marginal Metropolis-Hastings
R_0	basic reproduction number
RAT	rapid antigen test
REACT-1	REal-time Assessment of Community Transmission-1
RIM	random iterative method
RKHS	reproducing kernel Hilbert space
RMSE	root mean square error
R_t	instantaneous/time-varying reproduction number
r_t	instantaneous growth rate
R_t^c	case reproduction number
RT-PCR	reverse transcription polymerase chain reaction
PVE	proportion of variance explained
SARS	severe acute respiratory syndrome
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SEIR	susceptible-exposed-infectious-removed
SIMPLE	Survey Inference Method for Prevalence and other Latent variables in Epidemiology
SBI	simulation-based inference
SIR	susceptible-infectious-removed
SMC	sequential Monte Carlo
SPI-M	Scientific Pandemic Influenza Group on Modelling
UK	United Kingdom of Great Britain and Northern Ireland
UKHSA	UK Health Security Agency
UR	uncertainty reduction
VOI	value of information
WBE	wastewater-based epidemiology
WHO	World Health Organization
WWTP	wastewater treatment plant

Chapter 1

Introduction

1.1 Motivation

Infectious diseases pose one of the largest threats to global health [1]. Epidemics of influenza, HIV, SARS, Ebola, SARS-CoV-2, and Mpox, among others, have repeatedly disrupted societies, strained health systems, and caused large-scale morbidity and mortality [2]. Despite advances in public health and pandemic prevention, the number of emerging infectious disease events has risen significantly in recent years [3, 4], driven by increased human mobility [5], deforestation [6], agricultural intensification [7], climate change [8], and other societal and environmental factors [9]. The burden of these epidemics disproportionately falls on marginalised and poorer populations [10, 11, 12]. While improvements in medicine, vaccination, and sanitation have reduced the burden of some infectious diseases, epidemics remain persistent and unpredictable threats [13].

Statistical epidemiology provides important tools for understanding and predicting the spread of infectious diseases [14, 15]. Statistical models are applied to epidemiological data to estimate key parameters such as the epidemic growth rate and instantaneous reproduction number [16], infer prevalence and incidence [17], identify demographic heterogeneities and inequities [18], and forecast future disease burden [19]. These modelling outputs are useful for both real-time decision-making and retrospective analyses, informing public health interventions [20], healthcare resource allocation [21], and retrospective assessments that improve preparedness for future epidemics [22].

Reflecting the growing threat of infectious disease outbreaks and the value of statistical approaches for epidemic response, the volume of infectious disease modelling research has risen rapidly, particularly during the COVID-19 pandemic, far outpacing growth in the wider scientific literature [23]. Driven by increasing computational power [24], novel data streams [25], and methodological advances [26, 27], this rapid expansion has substantially broadened the scope of the epidemiological toolkit [28]. Yet, the speed and scale of this growth also creates challenges for reproducibility, accessibility, and clarity [29, 30].

In this context, progress in statistical epidemiology depends both on methodological innovation and the careful consolidation of existing approaches. Developing principled statistical frameworks for the unique challenges of infectious disease epidemiology is critical [31, 32]. Further statistical challenges include quantifying and communicating uncertainty [33, 34], ensuring robustness of modelling outputs [35], accounting for biases in observed data [36], and synthesising multiple data sources [24]. The rapid development of modern machine learning approaches also provides both opportunities and challenges for epidemiology [37]. Addressing these challenges will yield more reliable and useful models, strengthening their utility for public health decision-making and ultimately improving global health outcomes [15].

1.2 Thesis objectives and structure

The primary objective of this thesis is to present novel methodology and results that advance the field of statistical infectious disease epidemiology. A distinguishing feature of *statistical* modelling is its explicit consideration of uncertainty, a topic central to this thesis. Additionally, this thesis makes meaningful contributions to the field by prioritising the following principles:

- Consolidation: This thesis reframes, generalises, and advances existing methods within the field.
- Reproducibility and accessibility: All code and data are available in a well-documented, easily reproduced manner tailored to the specific work.
- Interdisciplinarity: Methods and ideas from adjacent fields that can benefit epidemiology are identified and adapted.

Brief descriptions of each chapter and their contribution to these additional principles are given below. This information is also summarised in Table 1.1.

Chapter 2 provides a background and literature review. As each paper in this thesis is self-contained, we provide a background through a historical lens, focusing on how both statistics and epidemiology have developed. We also highlight and define key statistical approaches, epidemiological models, parameters, and sources of data, that are used throughout the thesis.

Chapter 3 presents the first paper: "*A decision-theoretic framework for uncertainty quantification in epidemiological modelling*" [38]. This paper introduces a decision-theoretic approach to uncertainty quantification in epidemic models. By leveraging recent machine learning results and adapting them for epidemiology, we demonstrate a theoretically grounded-yet-practical separation of uncertainty into reducible and irreducible components. This paper focuses on consolidation and interdisciplinarity. Code and data to reproduce the wastewater-based epidemiology example are provided here: https://github.com/nicsteyn2/DecisionTheoreticEpi.

Chapter 4 presents the second paper: "*Robust uncertainty quantification in popular estimators of the instantaneous reproduction number*" [39]. This paper derives likelihoods for the smoothing parameters in two popular estimators of the instantaneous reproduction number, a widely used metric for quantifying epidemic spread. These likelihoods are used to marginalise out parameters, showing that this improves the robustness of the model's uncertainty quantification. This paper focuses on reproducibility (we provide well-documented code of both original methods and our extensions in the Julia programming language) and consolidation (by improving two widely used methods). The online material can be found here: https://nicsteyn2.github.io/RobustRtEstimators/.

Chapter 5 presents the third paper: "*A primer on inference and prediction with epidemic renewal models and sequential Monte Carlo*" [40]. This paper provides a tutorial-style introduction to sequential Monte Carlo methods, focusing on their application to epidemic renewal models. Given recent advances in computational power, these methods have become increasingly appealing as a flexible-yet-conceptually simple way to perform inference and prediction,

particularly in the epidemiological-time-series setting. This paper focuses on reproducibility: all example models, along with additional ones, are provided on a standalone website: https://nicsteyn2.github.io/SMCforRt/.

Chapter 6 presents the fourth paper: "*A Bayesian model for repeated cross-sectional epidemic prevalence survey data*" (accepted at PLOS Computational Biology, preprint [41]), in which we develop and test models for smoothing and performing inference on epidemic prevalence survey data, such as the data collected during the REal-time Assessment of Community Transmission-1 (REACT-1) study. We also provide a detailed comparison with two existing approaches, giving insight into good modelling practices for these data, and intuition for how seemingly small assumptions can strongly affect model outputs. This paper focuses on consolidation and reproducibility. Example notebooks and well-documented code are published here: https://github.com/nicsteyn2/EpidemicSurveySmoothing.

Chapter 7 presents the fifth paper: "*Jointly estimating epidemiological dynamics of COVID-19 from case and wastewater data in Aotearoa New Zealand*" [42], in which we construct a model to estimate COVID-19 dynamics using both reported cases and wastewater sampling data. By combining these two data sources, we reduce biases associated with each source, and enable estimation of quantities such as the relative case ascertainment ratio. This paper focuses on consolidation and reproducibility. All data used and well-documented code are available here: https://github.com/nicsteyn2/NZWastewaterModelling.

Chapter 8 presents the sixth paper: "*Pandemic-risk-related behaviour change in England from June 2020 to March 2022: REACT-1 study among over 2 million people*" [43], in which we analyse survey responses on behaviour during the COVID-19 pandemic. By comparing responses with community-level mobility data, we show that these mobility data often reflect self-reported population mobility. This paper focuses on accessibility; in particular, we make extensive data available online: https://github.com/nicsteyn2/REACTBehaviouralData/.

Chapter 9 re-examines aspects of Chapters 4, 5, and 6 through the lens of Chapter 3. We provide further insight into uncertainty in the renewal model, re-examine the marginalisation technique introduced in Chapter 4, demonstrate how the methods from Chapter 5 can estimate reducible and irreducible uncertainty, and consider survey sampling strategies in Chapter 6.

Chapter 10 provides a summary of contributions, directions for future work, limitations of the thesis, and short concluding remarks.

Chapters 3 through 8 are approximately ordered from most theoretical and general to most applied and specific. This ordering was chosen to facilitate discussion: the lessons from earlier chapters are relevant to later chapters and are used to frame Chapters 9 and 10. It is not coincidental that this ordering is also approximately reverse-chronological (based on when the work was done rather than published) reflecting the aim of generalising lessons from earlier work.

Table 1.1: An overview of the chapters in this thesis. Chapters 3-to-8 are formed of papers that have either been submitted to or accepted in academic journals. Chapters 1, 2, 9, and 10 were written specifically for this thesis. Where papers are not yet published, a reference to the corresponding preprint is provided.

Chapter	**Paper**	**Title**	**Status**	**Reference**
1	-	Introduction	-	-
2	-	Background and literature review	-	-
3	I	A decision-theoretic framework for uncertainty quantification in epidemiological modelling	Submitted for publication	[38]
4	II	Robust uncertainty quantification in popular estimators of the instantaneous reproduction number	Published in *American Journal of Epidemiology*	[39]
5	III	A primer on inference and prediction with epidemic renewal models and sequential Monte Carlo	Published in *Statistics in Medicine*	[40]
6	IV	A Bayesian model for repeated cross-sectional epidemic prevalence survey data	Accepted in *PLOS Computational Biology*	[41]
7	V	Jointly estimating epidemiological dynamics of COVID-19 from case and wastewater data in Aotearoa New Zealand	Published in *Communications Medicine*	[42]
8	VI	Pandemic-risk-related behaviour change in England from June 2020 to March 2022: REACT-1 study among over 2 million people	Submitted for publication	[43]
9	-	Revisiting Chapters 4, 5, and 6 through the lens of Chapter 3	-	-
10	-	Discussion and conclusions	-	-

1.3 Other work

In addition to the work presented in this thesis, I have contributed to several other published works, listed below in reverse chronological order. These include estimates of orphanhood and caregiver loss associated with COVID-19 in Brazil [44], methodological work on simulation-based inference of the instantaneous reproduction number [45], and evaluations of the sensitivity of reverse transcription polymerase chain reaction (RT-PCR) tests for SARS-CoV-2 [46], which informed work in Chapter 6 of this thesis. I also contributed to a case study in the Royal Society's report on the effectiveness of non-pharmaceutical interventions for COVID-19 [47], and two papers from the REACT-1 study [48, 49].

1. **Steyn, N.**, Unwin, H. J. T., Ponmattam, J., *et al.* (2025). Regional and national estimates of children affected by all-cause and COVID-19-associated orphanhood and caregiver death in Brazil, by age and family circumstance: a modeling study. *Accepted in The Lancet Regional Health Americas*. Preprint: https://doi.org/10.1101/2025.01.31.25321479.

2. Ogi-Gittins, I., **Steyn, N.**, Polonsky, J., Hart, W. S., Keita, M., Ahuka-Mundeke, S., Hill, E. M., & Thompson, R. N. (2025). Simulation-based inference of the time-dependent reproduction number from temporally aggregated and under-reported disease incidence time series data. *Philosophical Transactions of the Royal Society A*. https://doi.org/10.1098/rsta.2024.0412.

3. The Royal Society. (2023). *COVID-19: Examining the effectiveness of non-pharmaceutical interventions*[1]. https://www.royalsociety.org/npi-impact-on-covid-19.

4. Elliott, P., Whitaker, M., Tang, D., Eales, O., **Steyn, N.**, *et al.* (2023). Design and Implementation of a National SARS-CoV-2 Monitoring Program in England: REACT-1 Study. *American Journal of Public Health*, 113(5), 545–554. https://doi.org/10.2105/AJPH.2023.307230.

5. Binny, R. N., Priest, P., French, N. P., Parry, M., Lustig, A., Hendy, S. C., Maclaren, O. J., Ridings, K. M., **Steyn, N.**, Vattiato, G., & Plank, M. J. (2023). Sensitivity of Reverse Transcription Polymerase Chain Reaction Tests for Severe Acute Respiratory

[1] My contribution was to the case study on pages 52 to 55.

Syndrome Coronavirus 2 Through Time. *The Journal of Infectious Diseases*. https://doi.org/10.1093/infdis/jiac317.

6. Elliott, P., Eales, O., **Steyn, N.**, *et al.* (2022). Twin Peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 Epidemics in England. *Science*, 376(6600), eabq4411. https://doi.org/10.1126/science.abq4411.

Chapter 2

Background and literature review

This chapter provides background information and a literature review relevant to the papers presented in this thesis. We start with a very brief history of statistical epidemiology, in which we identify key statistical frameworks, epidemiological models, epidemiological parameters, and sources of epidemiological data that have been developed to model and understand infectious disease dynamics. These topics form the next four sections of this chapter. Finally, we present an overview of the COVID-19 pandemic. As each paper presented in this thesis is self-contained and includes its own introduction, we focus on broader underpinning themes, concepts, and historical context.

2.1 A very brief history of statistical epidemiology

Epidemiology and statistics are inextricably linked fields that have co-evolved over time. Early examples of statistical epidemiology in Western scholarship date to the 17th century [50, 51], when John Graunt's seminal work on the Bills of Mortality in London offered novel and valuable insights about public health [52, 53]. Many challenges faced by modern statistical epidemiologists were identified by Graunt, including reporting biases and missing data [54]. He noted that the "searchers", typically older and uneducated women, tasked with determining causes of death were unreliable and could be incentivised by officials to misreport [53]. By arguing that plague deaths were systematically underreported and estimating the extent of this, Graunt's work is a precursor to the modern task of estimating the *case ascertainment ratio*. His work was also groundbreaking for showing that population-level health could be described mathematically, a fact taken for granted today.

At the same time, Blaise Pascal, Pierre de Fermat, Christiaan Huygens, and Antoine Arnauld were developing probability theory [54]. In 1662, the very same year as the publication of Graunt's first analysis, Arnauld published *La logique, ou l'art de penser*, which contained the first known use of the term "probability" in its modern mathematical sense [53]. In 1669, Huygens and his brother Lodewijk Huygens applied the concept of probability to Graunt's work, though their results remained unpublished until 1894 [54].

In 1760, Daniel Bernoulli used probability and calculus to model the effect of smallpox inoculation [55]. Bernoulli justified his chosen values of "two pieces of elementary information" (the probability of catching smallpox and the probability of dying from it), arguing that alternative values would produce data contradicting what was observed. Today, we call these pieces of "elementary information" *parameters*, with these specific parameters being called the *force of infection* and *case fatality ratio* in the modern literature. Identifying parameter values that could plausibly generate observed data underpins modern statistical inference.

Both fields continued to develop over the following centuries. Early statistics were driven in large part by the problem of binomial proportions: if we observe x "successes" from n trials, what can we say about the probability of success p? This is immediately familiar to the epidemiologist: what is the prevalence of a disease? (Of course, calling an infection a "success" requires a level of semantic detachment.) The earliest solutions to this problem were independently developed by Thomas Bayes and Pierre-Simon Laplace [56, 57], using what became known as inverse probability, a precursor to modern Bayesian inference [58]. While treating the true value of p as fixed, a probability distribution was placed over the potential values to express uncertainty. By assuming a probabilistic model for the observed data, they worked backwards using (what we now call) Bayes' theorem, inverting the probability distribution on x to one on p.

19th-century epidemiology is most commonly associated with the complementary and sometimes contradictory works of John Snow and William Farr. Snow is best known for his identification of a specific water pump in London as the source of a cholera outbreak, which he (correctly) argued was caused by contaminated water, refuting the widely held "miasma" theory of disease transmission [59]. Farr, a highly respected statistician, was more focused on the statistical analysis of mortality data and the development of vital statistics [59]. In 1840,

nine years before the aforementioned cholera outbreak, Farr published what became known as "Farr's laws" [60], showing that the incidence of an infectious disease follows a predictable pattern over time. He made the insightful observation that "*if the latent cause of epidemics cannot be discovered, the mode in which it operates may be investigated. The laws of its action may be determined by observation...*" [61]. Farr was effectively arguing that we can make useful mathematical statements about infectious disease dynamics even if we do not fully understand the underlying mechanisms of transmission. This was fortunate, as Farr did not completely accept Snow's correct theory of cholera transmission until the later 1866 epidemic of cholera, which occurred after Snow's death.

The 20th century saw the rapid development of more sophisticated mathematical models, including the Reed-Frost chain-binomial model[1] (although an earlier version of a chain-binomial epidemic model was published by Pyotr Dimitrevich En'ko) [62, 63, 64]. Perhaps the most influential paper in the field, however, is William Ogilvy Kermack and Anderson Gray McKendrick's 1927 paper "*A contribution to the mathematical theory of epidemics*" [65], which is highly cited for the susceptible-infected-removed (SIR) compartmental model. More relevant to this thesis is the often-overlooked fact that Kermack and McKendrick also introduced a form of the *renewal model*, albeit as an intermediate step in their derivations. Denoting infectivity at "age" θ by ϕ_θ (age here refers to the time since infection), "*the number of individuals . . . at the time t who have been infected for θ intervals*" by $v_{t,\theta}$, and the number of susceptible individuals by x_t, the number of individuals who became infected at time t are given by:

$$v_t = x_t \sum_{\theta=1}^{t} \phi_\theta v_{t,\theta}.$$

This was extended to a system of ordinary differential equations by recognising that the continuous-time limiting model (found by taking the length of each time interval to zero) resembles the Lotka-Volterra model for population dynamics [66, 67]. The classic SIR model is recovered as a special case where infectiousness and recovery rates are constant.

[1] Lowell Reed presented his model to the non-scientific audience in a 1951 educational video titled "*Epidemic theory: What is it?*", produced by Johns Hopkins Science Review. This has been published on YouTube by the Johns Hopkins Medical Archives and can be accessed here: https://www.youtube.com/watch?v=OR5uzbPajzM. An interesting physical experiment with coloured marbles is used to demonstrate their model.

Returning to statistics, various criticisms of inverse probability arose over the 19th and 20th centuries. In 1844, Robert Leslie Ellis developed an early frequentist-style framework for statistical inference, arguing that the inverse probability approach violated *ex nihilo nihil* ("out of nothing, nothing"): that no information should lead to no inference at all [68]. Other objections to inverse probability included the conflation of probability as an objective measure of chance with probability as a measure of subjective belief (first attributed to Antoine Augustin Cournot [69]), the fact that the uniform prior distribution artificially adds one success and one failure to the inference (by George Boole and John Venn, among others [70]) and the lack of invariance under reparametrisation (most famously by Harold Jeffreys [71]).

Driven partially by these critiques of inverse probability, as well as the development of large-sample limiting arguments that allowed the consideration of previously intractable problems [72], multiple alternative approaches to statistical inference were formalised in the early 20th century. Most notably, the frequentist school of thought was introduced by Ronald A. Fisher [73, 74], Jerzy Neyman, and Egon Pearson [75]. Rather than seeking a probability distribution over a fixed parameter, the frequentist approach considers the behaviour of estimators under hypothetical repeated sampling, allowing the direct application of results like the central limit theorem. Motivated by both professional and personal disagreements with Pearson, Fisher went on to develop another framework called fiducial inference [76, 77], seeking to marry the benefits of the frequentist and Bayesian (inverse probability) approaches. While elegant in special cases (e.g., the binomial proportion), generalisation proved fraught and internal inconsistencies were identified. By the late 1950s, fiducial ideas had largely fallen out of use, leaving the frequentist school dominant.

The 1940s and 1950s saw the introduction of statistical decision theory. By framing inference as a problem of decision-making under uncertainty, Abraham Wald and others introduced the concepts of loss and risk, directly connecting statistical inference to real-world decision-making [78, 79]. This also provided a common language for the frequentist and Bayesian schools: frequentist approaches could be evaluated based on their long-run risk properties, with procedures often justified by minimax criteria (controlling the worst-case risk), while Bayesian methods were understood as rules that minimised expected loss under a prior distribution [80,

81, 82]. Modern-day epidemiology, particularly in the health economics context, often leverages these decision-theoretic ideas [83].

By the late 20th century, Bayesian methods were experiencing a resurrection, driven by both theoretical and computational advances [84, 85]. While early Bayesians had been constrained by analytic tractability, the development of Markov chain Monte Carlo (MCMC) methods allowed for efficient approximation of posterior distributions in high-dimensional settings [86]. Combined with the rapid increase in readily available computing power, and the development of user-friendly software (e.g., BUGS, Stan [87, 88]), Bayesian methods became increasingly popular across many fields, particularly in infectious disease contexts where mechanistic models are common and data are often limited and noisy [26, 89].

The 21st century has seen the reshaping of statistical epidemiology both conceptually and practically. On the conceptual side, debate about whether statistics should focus on "data modelling" (i.e., fitting mechanistic models) or "algorithmic modelling" (i.e., using flexible black-box methods) [90], and whether the fundamental purpose of statistics is to explain (infer) or to predict [91, 92], has coincided with the development of modern machine-learning methods [27, 93]. On the practical side, growth in the availability of large and novel datasets, computational capacity, and methodological advances have substantially broadened the toolkit available to the statistical epidemiologist. This is an exciting time for statistical epidemiology, with many open questions and opportunities for impactful research.

2.2 Statistical frameworks, models, and methods

Formally, a statistical model is defined as a set of probability distributions on some sample space $\mathcal{S}$ [94]. A parameterised statistical model further contains a parameter space Θ and a function $P : \Theta \to \mathcal{P}(\mathcal{S})$ which maps each parameter value $\theta \in \Theta$ to a probability distribution P_θ on $\mathcal{S}$. The sample space $\mathcal{S}$ consists of all possible datasets, where the goal of a statistical method is to use an observed dataset in $\mathcal{S}$ to make inferences about the "true" parameter value.

Practically, for this thesis, we define statistical modelling as the creation of a mathematical representation of a real-world data-generating process with the goal of making inferences and/or predictions about the underlying process. *Statistical* modelling is distinguished from other

forms of modelling by the use of statistical methods to quantify uncertainty about these inferences and predictions. The extent to which the mathematical representation is a mechanistic reflection of reality can range from a simplified description of the entire system to black-box algorithmic modelling.

2.2.1 Two dominant frameworks: frequentism and Bayesianism

Setting aside fiducial inference, there are two main schools of statistical inference: frequentism and Bayesianism. Fundamentally, the two schools differ in their interpretation of the source of uncertainty [95]. Take the goal of inferring a parameter θ given n observed data points $y_{1:n}$, with subscript $1:n$ referring to observations indexed $i = 1, \ldots, n$. The frequentist views uncertainty as arising from the alternative datasets that could have been observed, denoted by random variable $Y_{1:n}$. The frequentist estimator of θ is a function of the $Y_{1:n}$ and may be colloquially written [96]:

$$\theta(Y_{1:n}).$$

In contrast, the Bayesian conditions on the observed data $y_{1:n}$ and reasons about the missing data that would allow them to calculate, rather than estimate, θ. The Bayesian estimator may be written:

$$\theta(y_{1:n}, Y_{n+1:N}),$$

where N can be either finite or infinite.

An example from Chapter 3 makes this clear. Consider testing n individuals out of a population of size N for some disease, with $y_i = 1$ if individual i has the disease, and 0 otherwise. The goal is to estimate the proportion p of the population that has the disease. The frequentist would note that the sample mean $\hat{p} = \bar{y}_{1:n}$ is a sufficient statistic for p and apply the central limit theorem, indirectly arguing about the alternative datasets that could have been observed [95]. Frequentist confidence intervals reflect the variability in $\hat{p}$ that would arise under repeated sampling. The Bayesian would instead build a predictive model for the remaining $Y_{n+1:N}$, either implicitly by placing a prior distribution on p [97], or explicitly by building a predictive model for the unobserved $Y_{n+1:N}$ [98, 99]. Bayesian credible intervals reflect the posterior uncertainty about the true value of p we would calculate if we were to observe the full dataset.

The two frameworks are also often compared by their objectivity and subjectivity, with frequentism considered objective while Bayesianism is considered subjective, although both require the specification of a model, making them both subjective [100], and many researchers work on developing "objective" Bayesian methods [101]. Reflecting the widespread popularity of Bayesian methods for dynamic modelling in epidemiology, Chapters 3-to-7 are primarily Bayesian in nature. Chapter 8 instead leverages more traditional frequentist reasoning.

2.2.2 Likelihoods and simulation-based inference

The likelihood is central to both frequentist and Bayesian inference. Given observed data $y_{1:n}$ and a parameterised statistical model, the likelihood is defined as the probability of observing $y_{1:n}$ given parameter θ [95]:

$$L(\theta; y_{1:n}) = P_\theta(y_{1:n}).$$

In frequentist inference, the likelihood is typically maximised to obtain the maximum likelihood estimate (MLE) of θ:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta; y_{1:n}).$$

In Bayesian inference, the likelihood is coupled with a prior distribution $P(\theta)$, and Bayes' theorem is used to obtain the posterior distribution:

$$P(\theta|y_{1:n}) \propto L(\theta; y_{1:n})P(\theta).$$

Simulation-based inference (SBI) is used when the likelihood is unavailable or intractable, but where it is possible to simulate data from the model [102]. Rather than explicitly defining $L(\theta; y_{1:n})$, an implicit likelihood is induced by a simulator that generates synthetic datasets. This style of inference is particularly popular in epidemiology, as it allows general epidemic models to be used for inference. Methods such as approximate Bayesian computation (ABC) [26, 103] and neural density estimators [104] have been created to allow for SBI.

2.2.3 The predictive decomposition of the likelihood

One representation that features prominently in this thesis is the predictive decomposition of the likelihood [91], written:

$$L(\theta; y_{1:n}) = P(y_{1:n}|\theta) = P(y_1|\theta) \prod_{i=2}^{n} P(y_i|y_{1:i-1}, \theta). \tag{2.1}$$

While this applies to any sequence $y_{1:n}$ where the stepwise conditional densities are known, it is particularly useful in the time-series setting. Fundamentally, this tells us that the MLE of θ corresponds to the value of θ that maximises the probability of the one-step-ahead predictions. In the Bayesian setting, this approximately applies to the maximum a posteriori estimate (assuming the prior is mildly informative at most). The predictive decomposition of the likelihood connects the two tasks of inference and prediction: *inference* about a parameter θ is implicitly performed by choosing the best overall *predictive* model.

2.2.4 Mechanistic, semi-mechanistic, and non-mechanistic models

Mechanistic models (e.g., the SIR model, agent-based models) are direct mathematical representations of the underlying process that generated the observed data, and typically encode information about the entire system being modelled [105, 106]. Non-mechanistic models (e.g., most machine-learning models) are primarily data-driven, and tend to be more "black-box" in nature. Semi-mechanistic models (e.g., the renewal model, many state-space models) strike a balance between these two extremes, modelling only the relevant parts of the system mechanistically, while abstracting away other parts.

The trade-off between mechanistic and non-mechanistic models is driven by model flexibility, interpretability, and data requirements. Mechanistic models are usually more interpretable, with parameters that have direct real-world interpretations, and require fewer data, as more assumptions are encoded about the real-world process in the model structure itself. In contrast, non-mechanistic models are usually more flexible, but require more data to fit, and tend to be less interpretable. Semi-mechanistic models occupy an intermediate position: incorporating mechanistic structure where empirical knowledge is strong, data are sparse, or interpretability is important, while using non-mechanistic components where empirical knowledge is weak, data are plentiful, or flexibility is important.

Data are naturally limited in epidemiology. We only ever observe a single realisation of a given epidemic. This means that mechanistic models, where the relative lack of data is compensated for by making stronger assumptions about the underlying process, have long been popular in epidemiology. In recent years, however, semi-mechanistic modelling has become increasingly popular (such as through the renewal model [107] and dynamics-informed machine-learning

models [37, 108]). Even non-mechanistic models have gained popularity [19], particularly as machine-learning methods are adapted for epidemiological applications [109]. This trajectory mirrors the broader debate in statistics between the "data-modelling" (i.e., mechanistic) and "algorithmic-modelling" (i.e., non-mechanistic) cultures [90].

2.2.5 Sources of uncertainty

In the wider literature, statistical uncertainty is often classified as either *"aleatoric"* (relating to chance) or *epistemic* ("relating to knowledge") [110]. The term *aleatoric* is used to refer to uncertainty that is inherent to the data, inherent to the model, or cannot be reduced. The term *epistemic* is used to refer to uncertainty that is due to a lack of knowledge (such as from not knowing the true values of parameters), that results from a lack of data, or can otherwise be reduced. These definitions are somewhat vague and refer to a range of different quantities, leading to calls for more precise definitions [110].

The description of the sources of uncertainty in epidemiology also varies between publications. For example, D'Agostino McGowan et al. [111] categorise uncertainty into three types: *data uncertainty*, *stochastic uncertainty*, and *structural uncertainty*. Meanwhile, Zelner et al. [112] categorise uncertainty as *stochastic uncertainty*, *parameter uncertainty*, and *model/structural uncertainty*, and Swallow et al. [113] categorise uncertainty as *observation error/bias*, *stochastic uncertainty*, *parameter uncertainty*, *structural uncertainty*, and *model discrepancy*. The definitions of these categories, even those with shared names, vary by publication, and mathematical descriptions are typically not provided. In Chapter 3, we begin to address this gap by introducing precise yet practical definitions of *reducible* and *irreducible* uncertainty in an epidemiological context.

2.3 Epidemiological models

In this section, we briefly review epidemiological models that feature in this thesis. Further explanation of each model is given in the relevant chapters.

2.3.1 The renewal model

Central to this thesis is the semi-mechanistic renewal model. The modern discrete-time formulation was defined in the literature by Christophe Fraser in 2007 [107], and subsequently

popularised as a tool for real-time estimation of the time-varying reproduction number R_t by Anne Cori et al. in 2013 [114]. Since then, the use of this model in epidemiology has grown rapidly, with many extensions and adaptations being proposed [36, 115, 116]. In addition to R_t estimation, the renewal model can also be used for short-term forecasting [105, 117], estimating elimination probabilities [118], and measuring the effect of non-pharmaceutical interventions (NPIs) [22].

The renewal model directly relates past incidence to current incidence through a generation time distribution. Denoting infection/reported incidence at time t by I_t, the value of the generation time/serial interval (defined in Section 2.4.3) probability mass function by ω_u (where $\sum_u \omega_u = 1$ and u typically represents days), and the time-varying instantaneous reproduction number by R_t (defined in Section 2.4.2), the Poisson renewal model is written:

$$I_t \sim \text{Poisson}\left(R_t \sum_{u=1}^{t-1} I_{t-u}\omega_u\right), \tag{2.2}$$

where I_t can refer to the true (unobservable) infection incidence or to observed incidence such as reported cases or hospitalisations.

The renewal model is semi-mechanistic: it directly models the generation of new incidence from past incidence, while abstracting away other parts of the epidemic process, such as the depletion of susceptible individuals. It is common to pair this model with some form of smoothing on R_t to pool information over time [114, 115]. In fact, many popular implementations of the renewal model differ primarily in their choice of smoothing method.

2.3.2 Hidden-state models

A general hidden-state model consists of time-indexed unobserved hidden states X_t, observed data y_t, and parameter(s) θ [119]. The hidden states typically represent some underlying process of interest from which the observed data are generated. Such a model is defined by two probability distributions: a state-space transition distribution, which governs how the hidden states evolve over time, and an observation distribution, which governs how the observed data are generated from the hidden states. A diagram of a general hidden-state model is shown in Chapter 5 Figure 5.1.

The renewal model above (Equation 2.2) can be written as a hidden-state model, where the hidden states are R_t, the observed data are reported case incidence I_t, and the parameters are the values of the generation time distribution ω_u. In the Bayesian setting, absent any smoothing on R_t, the prior distribution on R_t forms the state-space transition distribution[2], and Equation 2.2 forms the observation distribution.

Hidden-state models are particularly useful in epidemiology as they allow for the explicit modelling of both the epidemic process (through the state-space transition distribution) and the observation process (through the observation distribution). This allows for the incorporation of delays, under-reporting, and other sources of noise in the observed data. Furthermore, we only need to be able to sample from the state-space transition distribution, allowing for simulation-based inference of the typically complicated epidemic process.

Methods for fitting hidden-state models include Kalman filters [119], particle filters [120, 121], MCMC [88, 122], and variational inference [123], all Bayesian methods. In Chapter 5 we describe the use of sequential Monte Carlo (SMC) methods, also known as particle filters, for fitting hidden-state epidemiological models. Frequentist methods also exist, although these are less common in the literature [124].

2.3.3 Logistic regression models

A logistic regression model is a type of generalised linear model (GLM) used for modelling binary outcome data [125]. The model assumes that the log-odds follow a linear function of the predictor variables. Denoting the binary outcome variable for the i^{th} individual by Y_i (where $Y_i = 1$ indicates the presence of some characteristic, and $Y_i = 0$ its absence), a vector of p predictor variables by $\mathbf{X}_i = (X_{1,i}, X_{2,i}, \ldots, X_{p,i})$, and a vector of regression coefficients by $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$, the logistic regression model is written:

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_p X_{p,i},$$

where $\text{logit}(\pi_i) = \log(\pi_i/(1-\pi_i))$. The outcome is then assumed to be distributed $Y_i \sim \text{Bernoulli}(\pi_i)$. The exponentiated coefficient $\exp(\beta_j)$ represents the odds ratio associated with a one-unit increase in the j^{th} predictor variable, holding all other variables constant.

[2]The word "transition" is somewhat of a misnomer here, as R_t does not "transition" over time in this example.

Uncertainty quantification for the regression coefficients can be performed in either a frequentist or Bayesian framework, with the former being more common in practice. Confidence intervals about the coefficients in the frequentist setting (as performed in Chapter 8) reflect the uncertainty arising from sampling variability. Finally, while the logistic regression model is not mechanistic in the traditional sense, it does impose a high level of structure on the data-generating process, and thus (subjectively) falls on the mechanistic end of the spectrum discussed in Section 2.2.4 (i.e., it is an example of the data-modelling culture rather than the algorithmic modelling culture discussed in [90]).

2.3.4 Modelling binomial proportions

A common task in epidemiology is to estimate a proportion, such as the prevalence of a disease (Chapter 6) or the proportion of individuals taking risk-mitigating behaviours (Chapter 8). While binomial proportions are one of the oldest and most studied problems in statistics [58], there is no universally agreed best method for estimating them and their associated uncertainty, with many methods being proposed over the years [126].

Popular frequentist approaches include the Wald interval, Wilson score interval, and Agresti-Coull interval [126, 127, 128]. Each method has strengths and weaknesses, focused around the trade-off between interval width, coverage, and the tendency for intervals to extend outside the $[0, 1]$ range.

Bayesian approaches typically use a conjugate $beta(\alpha, \beta)$ prior distribution for the proportion. The standard uniform prior distribution ($\alpha = \beta = 1$) is a common choice [56, 57], but is equivalent to adding one success and one failure to the observed data, thus introducing bias. The Jeffreys prior distribution ($\alpha = \beta = 0.5$) is another popular choice [71], as it is invariant under reparametrisation, but still introduces bias. The improper Haldane prior distribution ($\alpha = \beta = 0$) does not assume any pseudo-counts, but can lead to improper posterior distributions when there are no observed successes or failures [129]. In fact, the lack of a universally well-behaved Bayesian estimator for binomial proportions was one of the original criticisms of the inverse probability approach to statistics [70].

In Chapter 8, we face the additional challenge of estimating a proportion using weighted survey data. This naturally lends itself to a frequentist approach, as survey weights are typically

designed for use with frequentist estimators [130], and Bayesian methods would require the survey design itself to be modelled [131]. One solution to this problem, and the approach we use in Chapter 8, is to fit a weighted logistic regression model to produce confidence intervals on the log-odds scale, which are then transformed back to the probability scale [132]. This eliminates the issue of intervals extending outside the $[0, 1]$ range, but can still produce biased estimates and poor coverage at small sample sizes.

2.4 Epidemiological parameters

2.4.1 The growth rate

Perhaps the simplest epidemiological parameter is the exponential growth rate r_t [133]. This parameter describes the rate of change of incidence (new infections or cases) or prevalence (current infections or cases). Denoting incidence/prevalence at time t by X_t, the growth rate is defined as the value r_t that satisfies:

$$X_t = X_{t-1}e^{r_t}.$$

We use a general X_t here, as the growth rate can be defined for a variety of epidemiological quantities, including infection incidence, case incidence, hospitalisations, deaths, and prevalence. The growth rate is positive when X_t is increasing, negative when it is decreasing, and zero when X_t is stable. The units of r_t depend on the time units of t, for example, if t is measured in days, then r_t is the per-day growth rate.

2.4.2 The reproduction number

There are three commonly used definitions of the reproduction number: the basic reproduction number, the case reproduction number, and the instantaneous reproduction number. All three measure the average number of secondary infections/cases caused by a primary infection/case, but have different interpretations and uses. Others, like the household reproduction number [107], also exist, but are not discussed here.

The basic reproduction number R_0 was the first reproduction number to be described in the epidemiological literature, and is defined as the average number of secondary infections caused by a typical primary infection in a completely susceptible population [134]. It is a function of

the transmissibility of the pathogen, the contact patterns of the population, and the duration of infectiousness, but does not change in response to the depletion of susceptible individuals as an epidemic progresses. R_0 also features in the classic formula for the herd immunity threshold: $1 - 1/R_0$ [133], the proportion of a population that needs to be immune for epidemic growth to cease. R_0 is typically estimated from early epidemic growth rates [16] or by fitting mechanistic models to data [135].

The case reproduction number R_t^c is defined as the average number of secondary cases caused by a primary case infected at time t, popularised by Jacco Wallinga and Peter Teunis in 2004 [136]. Like R_0, R_t^c depends on the transmissibility of the pathogen, the contact patterns of the population, and the duration of infectiousness, but additionally accounts for the depletion of susceptible individuals as an epidemic progresses. R_t^c can be estimated by using the method of Wallinga and Teunis, or by reconstructing transmission trees [137].

Finally, the instantaneous reproduction number R_t is defined as the average number of secondary infections that would be caused by a primary infection if the primary infection were to expend all of its infectiousness at time t [138]. This definition is perhaps the most nuanced, but also the most useful for real-time epidemic tracking, as it reflects the current state of transmissibility in the population. A value of R_t greater than 1 indicates that the epidemic is *currently* growing, while a value R_t less than 1 indicates that the epidemic is *currently* shrinking, with the proportional change in contact rates required for a stable (flat) epidemic being $1/R_t$. The renewal model (Equation 2.2) is commonly used to estimate R_t from incidence data [114, 115, 116, 139]. R_t is also directly related to the growth rate in incidence r_t through the generation time distribution ω_u [16].

2.4.3 Generation times and serial intervals

The generation time is the time interval between the infection of a primary case and the infection of a secondary case. The serial interval is the time interval between the symptom onset of a primary case and the symptom onset of a secondary case [133, 140]. Both quantities are central to epidemiological models, including the renewal model, as they describe the temporal dynamics of pairs of infections/cases. The generation time is seen as more fundamental, as it directly reflects transmission, but is harder to observe. The serial interval includes additional variability

from the incubation period (the time from infection to symptom onset), but is typically easier to observe.

2.4.4 The case ascertainment ratio

The case ascertainment ratio (CAR) is the proportion of all infections that are detected and reported as cases [141]. This quantity is important for understanding the true burden of disease in a population, as many infections may go undetected due to asymptomatic or mild cases, limited testing capacity, or other factors. The CAR can be estimated by comparing the number of reported cases to the estimated number of infections from seroprevalence surveys, infection prevalence surveys, or other sources. In Chapter 7, we estimate the *relative* CAR over time for SARS-CoV-2 in Aotearoa New Zealand by jointly modelling wastewater surveillance data (which is unaffected by changes in testing behaviour or other reporting biases) and reported case data.

2.5 Epidemiological data

The surveillance of an infectious disease requires the collection and subsequent interpretation of epidemiological data. Surveillance methods are traditionally defined as either passive (where positive tests and suspected cases are reported to a central authority, for example [142]) or active (where public health authorities actively seek out data) [143], although this distinction is becoming increasingly blurred as new data sources become available.

One of the most commonly used sources of data in infectious disease surveillance is reported case counts of notifiable diseases. During the COVID-19 pandemic, daily reports of positive tests were common [144], though weekly aggregated data are typical for other infectious diseases [145, 146]. Other epidemiological data, such as hospitalisations, deaths, case demographics, and contact tracing data, are also often collected. Except for contact tracing, these are all examples of passive surveillance. Each source of data is subject to its own set of biases, requiring careful consideration during interpretation.

In addition to these traditional sources of epidemiological data, alternative data sources, such as genomic sequencing, have been increasingly used during the COVID-19 pandemic [147]. Further relevant examples are described below.

2.5.1 Infection prevalence surveys

Infection prevalence surveys aim to estimate the proportion of a population that is currently infected with a disease. Survey data avoid many of the biases present in passive systems, so are often considered the gold standard for prevalence estimation. Such surveys have been run on a small scale in the past [148, 149]. Most notably, however, during the COVID-19 pandemic, two infection prevalence surveys of unprecedented scale were run in the United Kingdom (UK): the Office for National Statistics (ONS) Coronavirus (COVID-19) Infection Survey [150] and the REal-time Assessment of Community Transmission-1 (REACT-1) study [48].

The REACT-1 study was an infection prevalence survey that tested for the prevalence of SARS-CoV-2 infection in England between May 2020 and March 2022 [48]. Conducted over 19 rounds, more than 2.5 million self-administered throat and nose samples were processed using reverse transcription-polymerase chain reaction (RT-PCR). From January 2021 (REACT-1 study round 8) onwards a subset of samples underwent genomic sequencing to determine the variant of SARS-CoV-2 present. Cycle threshold (Ct) values for both gene targets (N-gene and E-gene) were also recorded. Daily prevalence estimates and sample sizes for all 19 rounds of the REACT-1 study are reported in Figure 2.1. We develop and test a series of models for these data in Chapter 6.

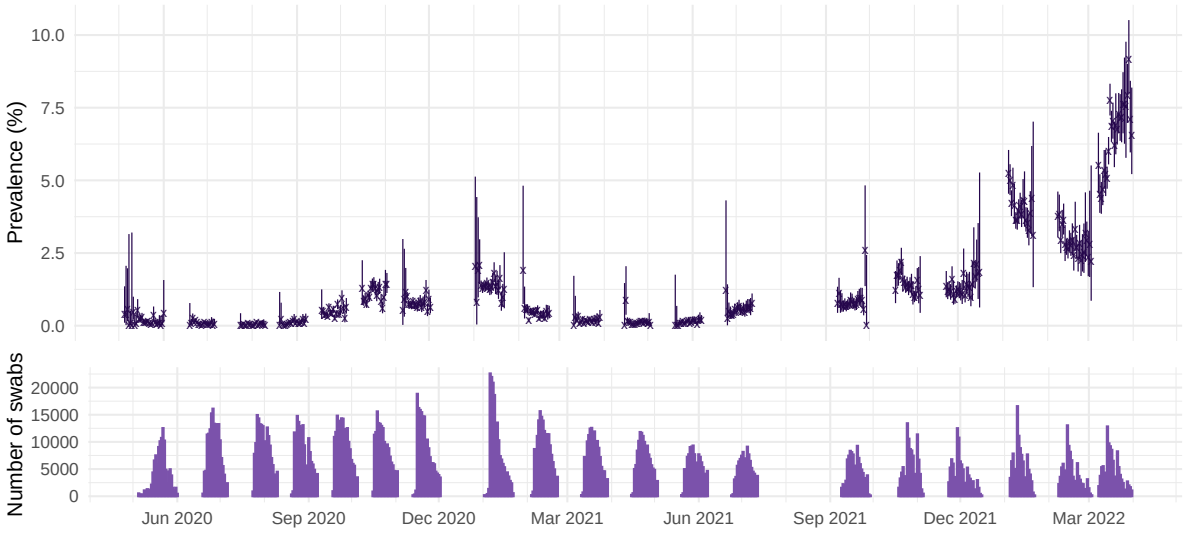

Figure 2.1: Daily SARS-CoV-2 prevalence estimates and 95% confidence intervals in England from the REACT-1 survey (upper) and corresponding daily sample sizes (lower). Weighted 95% confidence intervals for the prevalence estimates were calculated using the logit method in the *svyciprop* function of the *survey* package in R [132].

In addition to returning swabs for testing, participants completed a questionnaire that asked about their demographics, risk tolerances, behaviours, infection history, and vaccination status. This information has been used to supplement our understanding of the COVID-19 pandemic in England, for example, by allowing for the calculation of real-world vaccine effectiveness estimates [151], or variant-specific symptoms [152]. We leverage some of this collected information in Chapter 8 to investigate how individual behaviours changed over the pandemic, and whether these changes were reflected in Google mobility data [153] and the Oxford COVID-19 Government Response Tracker (OxCGRT) stringency index [154].

2.5.2 Wastewater sampling

Wastewater-based epidemiology (WBE) is a rapidly growing field that tests for the presence of biological markers to monitor the spread of infectious diseases in a population [155, 156], a relatively new form of passive surveillance. Wastewater samples are collected from wastewater treatment plants (WWTPs) and tested for the presence of pathogen biomarkers, typically using RT-PCR to detect genomic material. The concentration of pathogens in wastewater samples can then be used to estimate the prevalence of infection in the population served by the WWTP. Other uses for WBE include the early detection of outbreaks and the identification of infection hotspots [157, 158].

Unlike reported case data, the viral load in wastewater is not affected by human testing behaviours; hence wastewater sampling is hoped to provide a relatively unbiased correlate to the prevalence of an infectious disease. The variance of the concentration of genetic material in wastewater samples can be high, however, and may also be subject to several other biases [159, 160]. Even rainfall has been observed to impact wastewater sampling data [161].

While a promising area of research, WBE is still in its infancy [162]. How the concentration of genomic material is related to the prevalence of infection in a community, how this relationship varies between pathogens and over time, and how to account for dilution and degradation of pathogens in wastewater are all active areas of research. Determining the contexts in which WBE can provide valuable insights versus those in which the data may be too noisy for reliable interpretation remains an open question.

2.5.3 Transmission covariates

During the COVID-19 pandemic, many large technology companies published aggregate measures of population mobility. These include the Google COVID-19 Community Mobility Reports [153], Apple Mobility Trends Reports [163], and Facebook Movement During Crisis [164]. Additionally, in some jurisdictions, mobile phone providers also made mobility-indices available [165, 166]. These datasets have been used in a variety of epidemiological models [167, 168, 169, 170].

2.6 The COVID-19 pandemic

The COVID-19 pandemic, caused by the novel SARS-CoV-2 virus, was first detected in Wuhan, China in December 2019. It quickly spread around the world and was formally declared a pandemic by the World Health Organization (WHO) on 11 March 2020 [171]. After more than three years, with over 765 million reported cases, and nearly 7 million reported deaths, the WHO formally announced the end of the public health emergency of international concern on 5 May 2023 [172].

A wide range of responses were implemented by governments around the world. These included NPIs such as social distancing, mask mandates, and border closures [47], as well as pharmaceutical interventions such as vaccines and antiviral drugs. While these interventions should be evaluated in the context of their intended goals and the knowledge at the time of implementation, it is clear that their effectiveness varied widely between countries and over time [173].

The trajectory of the pandemic was further complicated by the emergence of SARS-CoV-2 variants of concern, such as Alpha, Delta, and Omicron, which are credited with causing large waves of infection in the UK and elsewhere [49, 174]. These variants often displayed either increased inherent transmissibility or increased immune evasion, making containment and management more challenging, and thus played a substantial role in resurgences of infection globally, even in countries that had previously eliminated COVID-19, such as Aotearoa New Zealand [175].

Statistical modelling played an integral role in developing an understanding of, and shaping responses to, the COVID-19 pandemic. Early models, such as Imperial College's "Report 9" [176], played key roles in forming early policy responses in the UK, and were reproduced and modified for use in other countries [177]. Alongside a plethora of statistical models, large-scale prevalence studies such as REACT-1 in England were established to monitor the spread of infection [48].

2.6.1 In Aotearoa New Zealand

Aotearoa New Zealand is an island nation in the South Pacific. It has a population of approximately 5.3 million people, with 1.7 million living in the largest city, Auckland [178]. New Zealanders are used to strict biosecurity controls at the border and the country has a history of eliminating invasive species [179, 180]. This experience made New Zealand well suited to the elimination of COVID-19, which formed the basis of the national response to the pandemic in 2020 and 2021.

The first case of COVID-19 in New Zealand was detected on 26 February 2020 [20]. Strict international border controls and local NPIs were used to successfully eliminate COVID-19 [181]. This allowed New Zealanders to largely live without COVID-19 or domestic NPIs until 17 August 2021, when the incursion of the Delta variant of SARS-CoV-2 triggered the end of the elimination strategy on 4 October 2021 [182]. Border controls remained in place until early 2022 - which were loosened after the local establishment of the Omicron variant of SARS-CoV-2 caused New Zealand's first major wave of COVID-19. Figure 2.2 shows the daily reported case numbers in New Zealand between February 2020 and August 2023. Further information about New Zealand's strategy for managing COVID-19 can be found in [183, 184].

2.6.2 In the United Kingdom

The UK is a country situated in northwestern Europe and is composed of four nations: England, Scotland, Wales, and Northern Ireland. With a population of approximately 68 million people [186] in a similar land area to New Zealand, the UK is densely populated, particularly in cities like London. While an island like New Zealand, the UK's role as a major international travel hub and close ties to mainland Europe exposed the nation to early and widespread transmission of SARS-CoV-2.

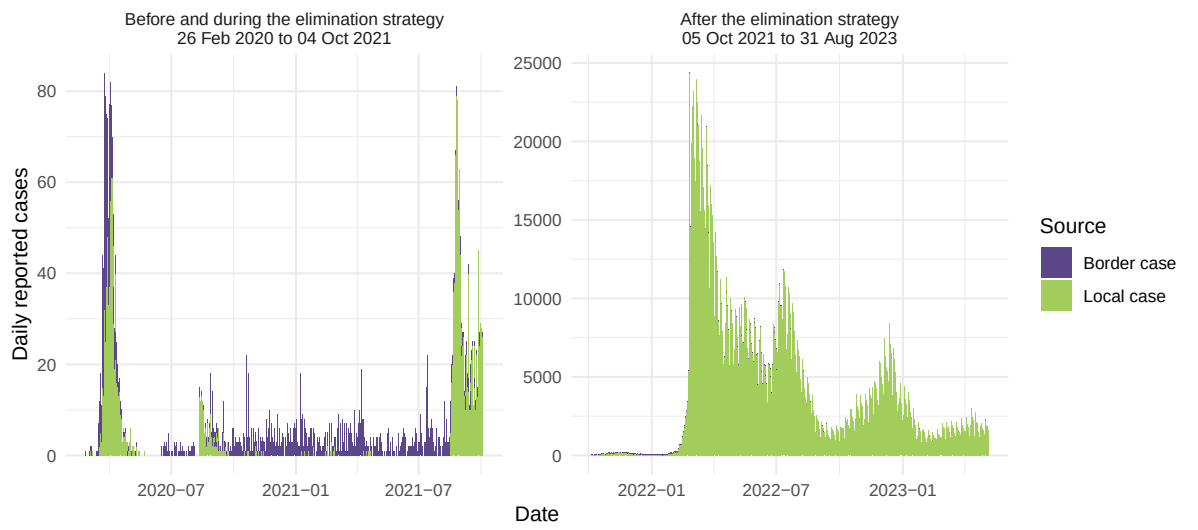

Figure 2.2: Daily reported case incidence in New Zealand between 26 February 2020 and 31 August 2023 [185]. Cases detected at the border are coloured purple while locally acquired cases are coloured green. Counts include both confirmed (82.4% of total) and probable (17.6% of total) cases. Between April 2020 and February 2022 all border cases were in some form of isolation and did not threaten New Zealand's elimination status.

The first case of COVID-19 in the UK was detected on 30 January 2020 [187]. By March, SARS-CoV-2 had spread widely across the country, leading to the implementation of a lockdown on 23 March 2020 to prevent health services being overwhelmed. While elimination was not achieved (unlike in New Zealand, elimination was not a goal of UK public health policy), infections decreased and remained low over the summer months [188]. Reported cases began to increase in autumn 2020, and the nation subsequently experienced multiple waves of infection, driven by new variants of SARS-CoV-2 [188]. Throughout the pandemic, a variety of NPIs were used, including social distancing, mask mandates, and restrictions on gatherings. These rules often varied by region, with a tier system used to help dampen transmission in regions identified as hotspots [189]. Later, vaccination also played an important role in controlling the spread and impact of the virus [190, 191].

The UK has a strong statistical epidemiology community, resulting in the development of multiple comprehensive surveillance systems. For example, multiple groups independently produced estimates of transmission parameters, which were aggregated by a central modelling authority, the Scientific Pandemic Influenza Group on Modelling (SPI-M) [192]. These modelling outputs were directly used in public health policymaking, as well as communicated to the public. In

addition to this, a variety of infectious disease surveillance programmes were set up, including extensive hospital- and community-based testing, digital contact tracing systems, infection prevalence surveys, seroprevalence studies, and wastewater sampling, for example.

Chapter 3

Paper I: A decision-theoretic framework for uncertainty quantification in epidemiological modelling

Status: Submitted for publication. A preprint is available on arXiv here: https://doi.org/10.48550/arXiv.2509.20013.

Full author list: Nicholas Steyn, Freddie Bickford Smith, Cathal Mills, Vik Shirvaikar, Christl A. Donnelly, and Kris V. Parag.

Authorship contributions: N.S. conceived of the study, wrote the software, created the visualisations, and wrote the original draft. N.S. and F.B.S. designed the methodology. All authors validated the methods and reviewed and edited the manuscript. C.A.D and K.V.P. supervised the project.

Supporting information: All code and data to reproduce the wastewater application are available at: https://github.com/nicsteyn2/DecisionTheoreticEpi.

Abstract

Estimating, understanding, and communicating uncertainty is fundamental to statistical epidemiology, where model-based estimates regularly inform real-world decisions. However, sources of uncertainty are rarely formalised, and existing classifications are often defined inconsistently. This lack of structure hampers interpretation, model comparison, and targeted data collection. Connecting ideas from machine learning, information theory, experimental design, and health economics, we present a first-principles decision-theoretic framework that defines uncertainty as the expected loss incurred by making an estimate based on incomplete information, arguing that this is a highly useful and practically relevant definition for epidemiology. We show how reasoning about future data leads to a notion of expected uncertainty reduction, which induces formal definitions of *reducible* and *irreducible* uncertainty. We demonstrate our approach using a case study of SARS-CoV-2 wastewater surveillance in Aotearoa New Zealand, estimating the uncertainty reduction if wastewater surveillance were expanded to the full population. We then connect our framework to relevant literature from adjacent fields, showing how it unifies and extends many of these ideas and how it allows these ideas to be applied to a wider range of models. Altogether, our framework provides a foundation for more reliable, consistent, and policy-relevant uncertainty quantification in infectious disease epidemiology.

3.1 Introduction

Statistical epidemiology seeks to learn about infectious disease dynamics by analysing data, often by fitting models [14]. These models are used for inference (estimating unobservable quantities such as parameters or latent variables) and/or prediction (estimating hypothetically observable data). Such estimates frequently inform real-world decision-making and public health planning [193]. In addition to producing estimates, the modeller typically quantifies uncertainty, for example, by presenting confidence intervals, credible intervals, or other uncertainty measures [34].

It is natural to ask where this uncertainty originates, how to quantify it, and how it might be reduced. Beyond its theoretical interest, a robust understanding of uncertainty is essential for determining which aspects of pathogen transmission to model [194], for guiding real-world decision-making [195], for planning data collection [196], and for communicating results [34]. Decision-theoretic approaches have been shown to enhance the credibility of using uncertain modelling outputs in these scenarios [32, 197].

Despite its importance, there is little agreement on how to systematically conceptualise and categorise uncertainty in epidemiological models, a lack of consensus also seen in other fields [110]. The epidemiological literature on uncertainty is fragmented, with inconsistent definitions, terms, and classifications (for example [111, 112, 113]). Discussions can be qualitative, lacking the mathematical formalism necessary for reproducibility, comparison between studies, and practical application.

This paper assumes uncertainty arises from making an estimate with incomplete information [98, 198]. Any decision would be fully informed if we had access to "complete data", where the precise definition of "complete data" depends on the model and problem setting, but uncertainty fundamentally arises from the fact that such data are never available. This leads to an intuitive first-principles decision-theoretic framework for uncertainty quantification which also induces a practical notion of *reducible* and *irreducible* uncertainty.

Our aim is to connect rigorous theoretical ideas with the practical needs of epidemiological modelling by drawing on concepts from machine learning [110] and adjacent fields. We adapt and integrate these ideas into a framework for uncertainty quantification in epidemiological

modelling by outlining a series of principles while also reflecting on broader conceptualisations. Our framework unifies existing concepts, such as value-of-information (VOI) (from health economics) [199, 200, 201], Bayesian experimental design [196, 202, 203, 204], information theory [205, 206], and scoring rules [32, 207], in an uncertainty-first form tailored towards epidemiology.

We focus on uncertainty arising within a chosen model and do not attempt to measure the mismatch between the model and the true data-generating process. Statements made about uncertainty here apply to the real world only if the model is correctly specified. However, our framework is a necessary step towards broader uncertainty quantification and we signpost extensions in the discussion.

Section 2 presents our decision-theoretic approach to uncertainty quantification in epidemiology as four principles, motivated by two examples: estimation of infection prevalence and estimation of the instantaneous reproduction number R_t, a common measure of epidemic spread [114]. Section 3 applies this framework to a real-world model of wastewater surveillance for SARS-CoV-2 in Aotearoa New Zealand. Section 4 discusses how this framework connects, extends, and generalises ideas from adjacent fields. Section 5 concludes with a discussion.

3.2 A decision-theoretic framework

We view model-based uncertainty as arising from missing information [98, 198]. For inference, this missing information is understood as the (possibly infinite) additional data that, if observed, would allow us to precisely determine any identifiable target quantity. For prediction, this view separates a reducible component of uncertainty that vanishes with unlimited data (a notion of parametric uncertainty) from inherent variability in the data-generating process. We describe our approach by stating four principles, motivating them with two conceptual examples.

This missing information perspective is foundational to Bayesian inference (although our framework is not restricted to Bayesian methods), in which a joint distribution is constructed over all modelled random quantities. By conditioning on the observed data, a posterior distribution over unobserved quantities is obtained [208], leading to the notion that uncertainty "flows from" missing data. The modelling of these missing data can be implicit, usually by updating

beliefs about model parameters [97], or explicit, by directly modelling the missing data [91, 98, 99].

Consider estimating infection prevalence θ (Figure 3.1). If infection statuses for the full population were known, there would be no uncertainty about θ. With only a subset tested, a binomial model treats uncertainty as arising from an infinite unobserved population, while a hypergeometric model treats it as directly arising from the finite unobserved population. In both cases, uncertainty flows from assumed missing information, with the precise nature determined by the model.

Principle 1: Model-based uncertainty arises from decision-making under incomplete information.

Decision theory directly relates the missing data to a real-world decision, giving rise to concrete and problem-grounded measurements of uncertainty. For simplicity, we take the decision to be making an estimate (covering both inference and prediction), where the statistical estimator performs the role of a decision rule that maps observed data to some form of estimate. The framework is readily extended to other types of decisions (such as whether to impose an intervention), where the decision rule maps observed data to a real-world action.

In this framework, the modeller's role is to make an estimate $a \in \mathcal{A}$ of unknown $z \in \mathcal{Z}$, inducing a loss $\ell(a, z)$ to be minimised, where $\mathcal{A}$ is the space estimates (actions) that could be made and $\mathcal{Z}$ are the possible values of z. $\ell(a, z)$ encodes the real-world consequence of making an estimate a when the true value is z, a concept most familiar to machine-learning modellers and statistical forecasters [209, 210]. The unknown quantity z could be a model parameter (e.g., R_t) or a future observation (e.g., future hospitalisations).

The decision-theoretic approach begins by fitting a model $p(z; y_{obs})$ to the observed data y_{obs}. In the prevalence example, this is the posterior distribution $p(\theta|y_{1:n})$. The *Bayes-optimal* estimate a^* minimises the expected loss $E_{p(z|y_{obs})}[\ell(a, z)]$ under the fitted model:

$$a^* = \arg\min_{a \in \mathcal{A}} E_{p(z|y_{obs})}[\ell(a, z)]. \tag{3.1}$$

For point estimates ($\mathcal{A} = \mathcal{Z}$), the quadratic loss $\ell(a, z) = (a - z)^2$ is commonly used, whose Bayes-optimal estimate is the posterior mean. For probabilistic estimates ($\mathcal{A} = \mathcal{P}(\mathcal{Z})$, the set

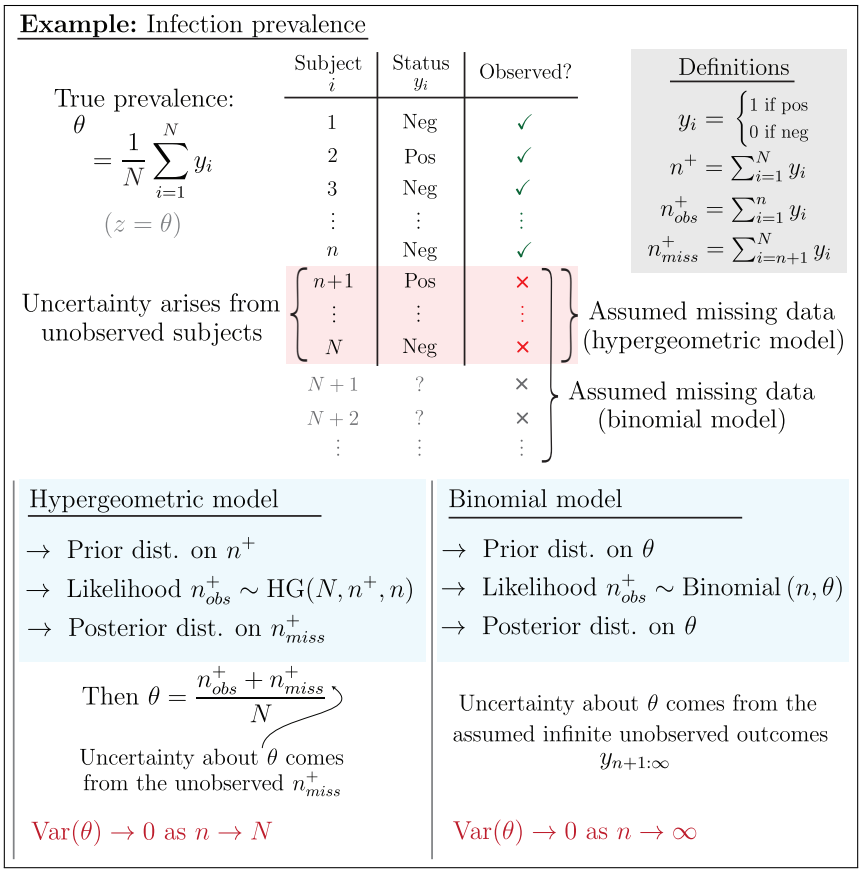

Figure 3.1: Uncertainty associated with estimating infection prevalence (here, $z = \theta$). Assume we test n subjects out of a population of size N and observe n^+_{obs} positive results. The missing data are the number of positive results n^+_{miss} in the $N-n$ subjects that were not tested. A popular approach is to place a prior distribution on θ and model the observed data using a binomial distribution, resulting in a posterior distribution whose variance goes to 0 as $n \to \infty$. In this example, the assumed missing data are the infection status of an infinite number of untested individuals. An alternative approach is to place a prior distribution on the true total number of infected subjects n^+ and model the observed data using a hypergeometric distribution, directly resulting in a posterior distribution on n^+_{miss}. By setting $\theta = \frac{n^+_{obs}+n^+_{miss}}{N}$, we obtain a posterior distribution on θ whose variance (one possible measure of uncertainty) goes to 0 as $n \to N$. In this example, the assumed missing data are the infection status of the finite number of untested individuals.

of probability distributions on $\mathcal{Z}$), the log-loss $\ell(a, z) = -\log a(z)$ is commonly used, whose Bayes-optimal estimate is the posterior distribution itself.

A key insight of DeGroot [211] is that the minimised expected loss provides a general and problem-grounded definition of uncertainty [110, 212]. Borrowing notation from Bickford Smith et al. [110], we define uncertainty h as:

$$\begin{aligned} h\left[p(z|y_{obs})\right] &\coloneqq \min_{a \in \mathcal{A}} E_{p(z|y_{obs})}[\ell(a, z)], \\ &= E_{p(z|y_{obs})}[\ell(a^*, z)]. \end{aligned} \tag{3.2}$$

Uncertainty is measured by examining the expected loss incurred by taking the optimal action given the information at hand.

Multiple popular measures of uncertainty can be derived from this setup [110, 200, 212]. For a point estimate and quadratic loss where a^* is the posterior mean, uncertainty is measured by the posterior variance:

$$\begin{aligned} h\left[p(z|y_{obs})\right] &= E_{p(z|y_{obs})}[(E_{p(z|y_{obs})}[z] - z)^2], \\ &= \mathrm{Var}_{p(z|y_{obs})}(z). \end{aligned} \tag{3.3}$$

For a probabilistic estimate and log-loss where a^* is the posterior distribution, uncertainty is measured by the differential entropy of the posterior distribution:

$$\begin{aligned} h\left[p(z|y_{obs})\right] &= E_{p(z|y_{obs})}[-\log p(z|y_{obs})], \\ &= \mathrm{H}[p(z|y_{obs})]. \end{aligned} \tag{3.4}$$

We illustrate this for the prevalence example in Figure 3.2.

The measure of uncertainty is determined by the loss function, thus alternative functions can capture different consequences of estimation error. When estimating R_t, for example, authorities may want to prioritise avoiding underestimation to be conservative about epidemic control. Here, we could use a loss function that penalises underestimation more than overestimation, or target tail risk with a quantile (pinball) loss function [213]. This yields problem-specific uncertainty measurements, reflecting user-specific risk preferences, that can guide decision-making, data-collection, and model comparison (for example).

Principle 2: The incomplete information driving uncertainty is application-specific and determines the decomposition of reducible and irreducible uncertainty.

In machine learning and other fields, it is common to assume observed data $\{y_i\}_{i=1}^n$ are independent and identically distributed (i.i.d.) from some "training" distribution $y_i \sim p_{train}(y)$. This leads to the notion that irreducible (sometimes called "aleatoric" [33], although we emphasise this term has many definitions in the literature [110]) uncertainty is the uncertainty that would remain after fitting the model to the "complete", often infinite, training data. In this view, the missing information corresponds to the unobserved data $y_{n+1:\infty}$.

In epidemiological transmission modelling, we only ever observe one realisation of a given epidemic, making the empirical quantification of inherent uncertainty difficult [33]. Arguing

Example: Infection prevalence (binomial model continued)

Prior distribution: $\theta \sim \text{Beta}(\alpha_0, \beta_0)$

Likelihood: $n^+_{obs}|\theta \sim \text{Binomial}(n, \theta)$

Posterior distribution: $\underbrace{\theta|n^+_{obs} \sim \text{Beta}(\alpha_0 + n^+_{obs},\ \beta_0 + n - n^+_{obs})}_{\textit{Our model fit to the observed data}}$

	Point estimate $a \in [0,1]$	Probabilistic estimate $a \in \mathcal{P}([0,1])$
Loss function $\ell(a,\theta)$	$(a-\theta)^2$ *(quadratic loss)*	$\log a(\theta)$ *(log loss)*
$\Downarrow$	$\Downarrow$	$\Downarrow$
Optimal estimate a^*	$\frac{\alpha_0 + n^+_{obs}}{\alpha_0 + \beta_0 + n}$ *(posterior mean)*	$\text{Beta}(\alpha_0 + n^+_{obs},\ \beta_0 + n - n^+_{obs})$ *(posterior distr.)*
$\Downarrow$	$\Downarrow$	$\Downarrow$
Measure of uncertainty $E\left[\ell(a^*,\theta)\right]$	$\text{Var}(\theta\|n^+_{obs})$ *(posterior variance)*	$\text{H}\left[\text{Beta}(\cdot,\ \cdot)\right]$ *(posterior entropy)*

Figure 3.2: **Principle 1**. Two measures of uncertainty in the infection prevalence example with the binomial model ($z = \theta$ is the estimand). For simplicity, we only show the binomial model. If we use a conjugate $\text{Beta}(\alpha_0, \beta_0)$ prior distribution, then the posterior distribution is $\text{Beta}(\alpha_0 + n^+_{obs}, \beta_0 + n - n^+_{obs})$. If we make a point estimate of θ under the quadratic loss function, then the Bayes-optimal estimate is the posterior mean and uncertainty can be measured using the posterior variance. If we make a probabilistic estimate of θ under the log-loss, then the Bayes-optimal estimate is the posterior distribution itself and uncertainty can be measured using the posterior (differential) entropy.

about repeating the epidemic multiple times is of limited practical value, as this offers little insight into the uncertainty we could hope to reduce in practice. Uncertainty instead must be reduced by collecting higher-quality, or a greater quantity of, data during the epidemic [205].

We argue that focusing on missing data y_{miss} that *could practically be observed* is more useful. Specifically, we define the uncertainty reduction as the difference between the uncertainty of the model fit to the observed data (y_{obs}) and the uncertainty of the model fit to the full data ($y_{all} = y_{obs} \cup y_{miss}$) [110]:

$$\underbrace{\text{UR}_z(y_{miss})}_{\text{Uncertainty reduction}} = \underbrace{h\left[p(z|y_{obs})\right]}_{\text{Total uncertainty}} - \underbrace{h\left[p(z|y_{all})\right]}_{\text{Irreducible uncertainty}}. \tag{3.5}$$

For example, consider estimating R_t using the Poisson renewal model (Figure 3.3) [114]. Under perfect case reporting, R_t uncertainty could be eliminated only by repeating the epidemic

infinitely many times. Here, all uncertainty is irreducible. In the presence of underreporting, for example, the missing data are unreported cases, and if we knew these, we could reduce R_t uncertainty. Comparing this to the prevalence example, where y_i are assumed to be i.i.d., we see a key difference between epidemiology (and weather forecasting, economics, etc [32, 214]) and other fields such as machine learning: we often only observe a single realisation of the data-generating process and arguing about alternative realisations is of limited practical value.

This construction is purposefully broad. The forms of $p(z|y_{obs})$ and $p(z|y_{all})$ are not specified, allowing any model to be used. We also note that the form of the missing data y_{miss} need not match y_{obs}, enabling the inclusion of different datasets in the model. We illustrate this in section 3.3, considering the inclusion of wastewater data in a model fit only to reported cases.

Principle 3: To quantify reducible and irreducible uncertainty, we reason about the missing data.

In some scenarios, we may know the value(s) of y_{miss}, for example, when including an additional already collected dataset in a model. In many scenarios, however, such as when planning future data collection, or estimating the irreducible uncertainty, we do not know the value(s) of y_{miss} in advance. In these cases, we must reason about the missing data that could be observed, and consider the *expected* uncertainty reduction.

Letting $p(y_{miss}|y_{obs})$ be a predictive model describing our beliefs about the missing data, we average over all possible realisations of y_{miss} by taking expectations of both sides of Equation 3.5, leading to the expected uncertainty reduction:

$$\underbrace{\mathrm{E}_{p(y_{miss}|y_{obs})}\left[\mathrm{UR}_z(y_{miss})\right]}_{\text{Expected uncertainty reduction}} = \underbrace{h\left[p(z|y_{obs})\right]}_{\text{Total uncertainty}} - \underbrace{\mathrm{E}_{p(y_{miss}|y_{obs})}\left[h\left[p(z|y_{all})\right]\right]}_{\text{Expected uncertainty after collecting } y_{miss}}. \tag{3.6}$$

In some scenarios, the appropriate predictive distribution may be obvious. For example, if we are constructing a Bayesian model using the standard prior-likelihood-posterior update and are considering the uncertainty reduction associated with collecting the same type of data, the posterior-predictive distribution is appropriate. We illustrate this in the context of prevalence estimation in Figure 3.4. In other scenarios, this predictive distribution may be less clear. In either case, $p(y_{miss}|y_{obs})$ is necessarily subjective and describes our beliefs about the missing data.

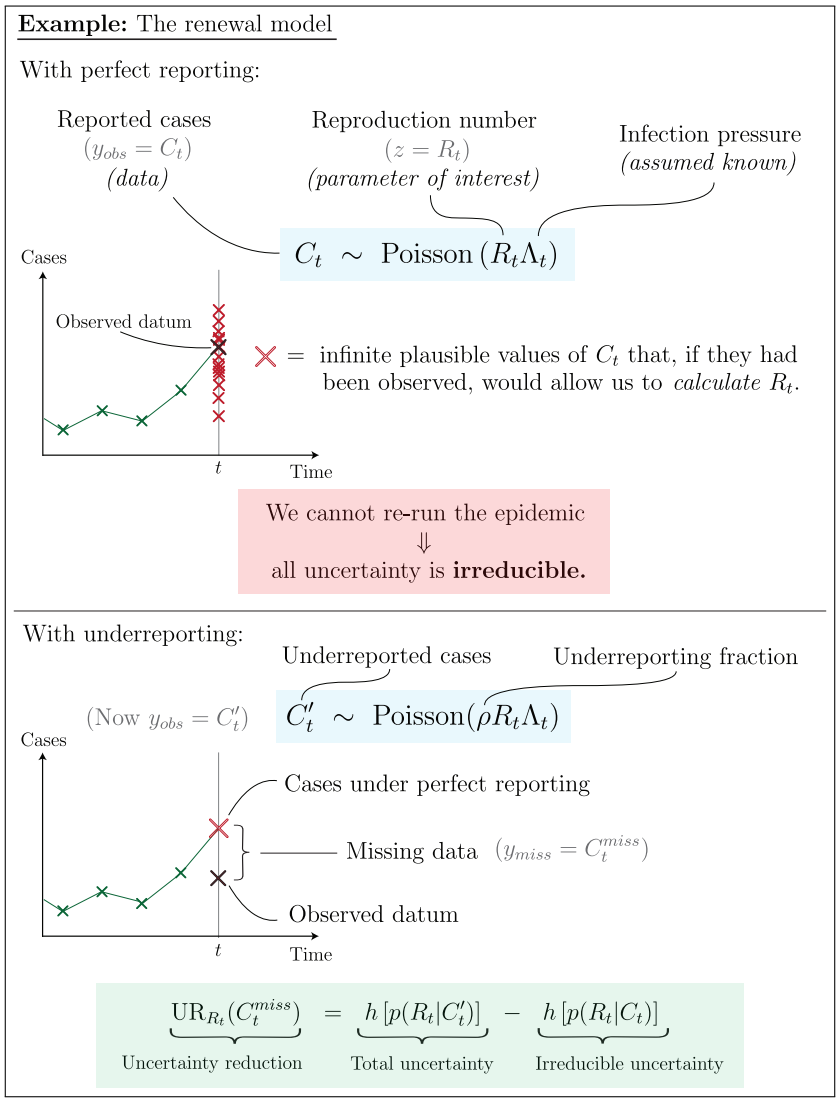

Figure 3.3: **Principle 2.** Reducible and irreducible uncertainty in the renewal model. The instantaneous reproduction number (R_t) is a key quantity in epidemiological modelling, defined as the average number of secondary infections caused by an infected individual at time t if they were to undergo their entire infectious period at this time [138]. It is commonly estimated from reported case data using the Poisson renewal model pictured here [114, 115]. Even if the reported cases C_t perfectly corresponded to the true infections (i.e., no underreporting), uncertainty about $z = R_t$ would still exist. Under this model, uncertainty about R_t could be eliminated only if we were to repeat the epidemic infinitely many times, thus all uncertainty is irreducible. In the presence of underreporting, the missing data are the unreported infections. If we knew these quantities, for example, by improving outbreak surveillance, we could reduce our uncertainty about R_t.

Example: Infection prevalence (binomial model continued)

Consider testing m additional subjects:

Subject i	Status y_i	Observed?
1	Neg	✓
2	Pos	✓
⋮	⋮	⋮
n	Neg	✓
$n+1$	Pos	□
$n+2$	Neg	□
⋮	⋮	⋮
$n+m$	Neg	□
⋮	⋮	⋮
N	Neg	×

$y_{miss} = \sum_{i=n+1}^{n+m} y_i$ *(observable future data)*

Sample 1	Sample 2	Sample 3	Sample 4	⋯
Pos	Neg	Neg	Neg	⋯
Neg	Neg	Pos	Neg	⋯
⋮	⋮	⋮	⋮	⋱
Neg	Neg	Neg	Neg	⋯

Samples from the predictive distribution for the missing data

To estimate the expected uncertainty given y_{all} :

→ Choose a predictive model for y_{miss} :

$$y_{miss} = \sum_{i=n+1}^{n+m} y_i \sim \text{BetaBinom}\left(m, \alpha_0 + n^+_{obs}, \beta_0 + n - n^+_{obs}\right)$$

(the posterior predictive distribution)

→ Sample from this predictive model: $y^{(i)}_{miss} \sim p(y_{miss}|y_{obs})$

→ Calculate uncertainty given these sampled data:

$$h\left[p(\theta|n^+_{obs}, y^{(i)}_{miss})\right] = \text{Var}(\theta|n^+_{obs}, y^{(i)}_{miss})$$ *(assuming point estimate and quadratic loss)*

→ Average over all samples:

$$\text{E}\left[Var(\theta|n^+_{obs}, y^{(i)}_{miss})\right] \approx \frac{1}{S}\sum_{i=1}^{S} \text{Var}(\theta|n^+_{obs}, y^{(i)}_{miss})$$

Expected uncertainty reduction from collecting y_{miss} :

$$\text{E}\left[\text{UR}_\theta(y_{miss})\right] \approx \text{Var}(\theta|n^+_{obs}) - \frac{1}{S}\sum_{i=1}^{S} \text{Var}(\theta|n^+_{obs}, y^{(i)}_{miss})$$

Figure 3.4: **Principle 3.** Estimating the expected uncertainty reduction in the infection prevalence example with the binomial model ($z = \theta$). Consider testing m additional individuals for infection. The posterior-predictive distribution given a beta prior distribution and binomial likelihood is a beta-binomial distribution. We calculate the expected uncertainty reduction about θ after collecting y_{miss} by averaging $h\left[p(\theta|y_{1:n}, y_{miss})\right]$ (the final term in equation 3.6) over this predictive distribution.

Principle 4: Collecting additional data is not guaranteed to reduce uncertainty in practice, but will in expectation if the predictive distribution is *coherent*.

In practice, new data may contradict existing data, leading to an increase in uncertainty about z. While this cannot be prevented with certainty, if the predictive distribution $p(y_{miss}|y_{obs})$ is *coherent* with the fitted model, then additional data will reduce uncertainty *in expectation*.

$p(y_{miss}|y_{obs})$ is coherent with the fitted model $p(z|y_{obs})$ if both are marginal distributions of the same joint distribution $p(z, y_{all})$, implying:

$$p(z|y_{obs}) = \int_{\mathcal{Y}} p(z|y_{all})p(y_{miss}|y_{obs})\ dy_{miss}, \tag{3.7}$$

where $\mathcal{Y}$ is the space of possible observed y_{miss}.

This means that the predictive distribution and the fitted model make the same assumptions about the missing data, and that the expected value of z should not change after including missing data sampled from $p(y_{miss}|y_{obs})$.

To see that coherence leads to an expected uncertainty reduction, define a^* as the Bayes-optimal estimate of z prior to observing y_{miss} (but after observing y_{obs}). After observing y_{miss}, the Bayes-optimal estimate may change, giving the first inequality:

$$\begin{aligned} \min_a \mathrm{E}_{p(z|y_{all})}\left[\ell(a,z)\right] &\leq \mathrm{E}_{p(z|y_{all})}\left[\ell(a^*,z)\right] \\ \implies \mathrm{E}_{p(y_{miss}|y_{obs})}\left\{\min_a \mathrm{E}\left[\ell(a,z)\right]\right\} &\leq \mathrm{E}_{p(y_{miss}|y_{obs})}\left\{\mathrm{E}\left[\ell(a^*,z)\right]\right\} \\ &= \mathrm{E}_{p(z|y_{obs})}\left[\ell(a^*,z)\right]. \end{aligned} \tag{3.8}$$

Taking expectations over $p(y_{miss}|y_{obs})$ gives the second inequality, and the final equality follows from coherence and the law of total expectation. Re-writing this in terms of h gives $\mathrm{E}\left[h(p(z|y_{all}))\right] \leq h(p(z|y_{obs}))$, thus the expected uncertainty reduction is non-negative. Equality holds if and only if z is conditionally independent of y_{miss} given y_{obs}. That is, the expected uncertainty reduction is zero only when the missing data provide no new information about z.

This condition re-asserts the missing data as the source of uncertainty. While we can use any model of the missing data $p(y_{miss}|y_{obs})$ to quantify the expected uncertainty reduction, we may get counterintuitive results if our predictive model is not coherent with the fitted model. Intuitively, coherence ensures that the explicit assumptions about the missing data that we

make when separating uncertainty are consistent with the implicit assumptions made by the fitted model.

In the prevalence example, suppose we believe the prevalence in individuals who do not volunteer for testing may be different from those who do. Reflecting this, we decide to allow for additional variation in the missing data by inflating the variance of the posterior-predictive distribution. If this increase in variance is introduced only in the prediction step, but not in the fitted model, then the predictive model is not coherent with the fitted model, and uncertainty may increase on average.

Coherence is a fundamental property of Bayesian inference and is often used to argue for Bayesian approaches over other inferential approaches. In fact, a model satisfies the coherence condition if and only if it is Bayesian [208, 215] (i.e., it updates beliefs via Bayes' rule). That is, the expected uncertainty reduction defined by Equation 3.6 is non-negative if and only if our model is Bayesian. While this makes using standard Bayesian approaches appealing, this also allows us to use any model that satisfies coherence (Equation 3.7) for Bayesian uncertainty quantification [216].

We emphasise that coherence is not a requirement of our framework and principles (1) to (3) are valid regardless. However, without this assumption, collecting more data is not necessarily expected to reduce uncertainty. All four principles are summarised in Figure 3.5.

3.3 An application to wastewater surveillance

How much can wastewater surveillance reduce uncertainty about R_t?

Wastewater surveillance gained prominence during the COVID-19 pandemic as a potentially less biased source of data than traditional sources (such as reported cases) for tracking real-time pathogen transmission [42, 160]. The concentration of genomic material in wastewater does not map directly onto traditional epidemiological indicators and can be highly variable [217]. This has led some to question the utility of wastewater data in this setting [160].

Watson et al. [42] constructed a joint model of wastewater and reported case data for Aotearoa New Zealand. To examine the utility of wastewater data for estimating R_t, we apply our framework to answer two questions: first, to what extent is uncertainty about R_t reduced by

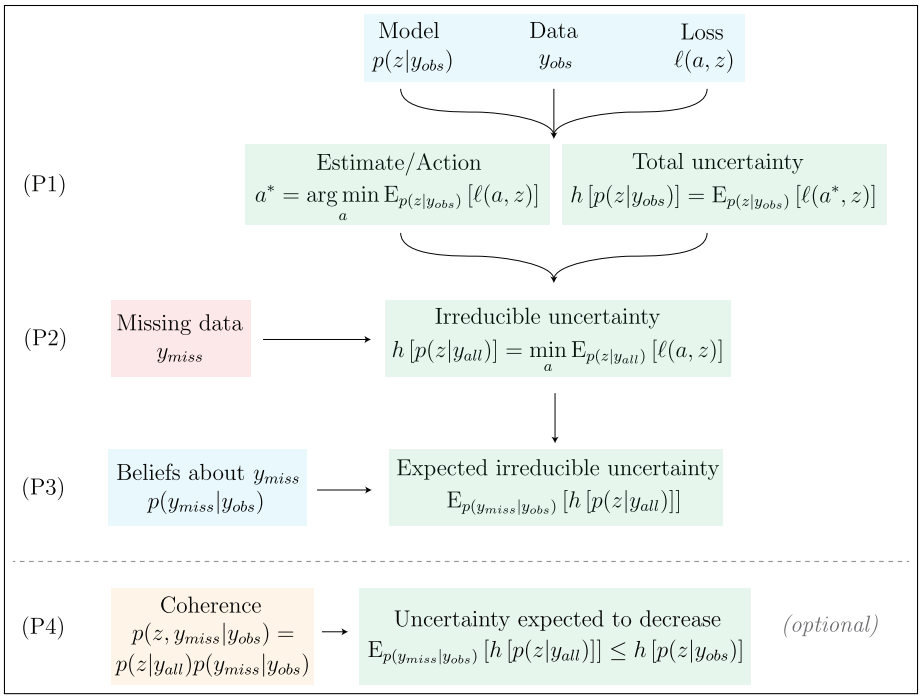

Figure 3.5: An overview of the four principles (P1 to P4) of our decision-theoretic framework. We start with a model $p(z|y_{obs})$ of the unknown quantity z, observed data y_{obs}, and a loss function $\ell(a,z)$ - the three quantities needed to make a decision under uncertainty. P1 uses the loss function to concretely relate the model and data to the real-world act of making an estimate a of z, defining the Bayes-optimal estimate a^* and the corresponding total uncertainty h. P2 defines the irreducible uncertainty as the uncertainty remaining after collecting missing data y_{miss}. As we generally do not know the value of y_{miss}, P3 introduces the expected irreducible uncertainty, which is the irreducible uncertainty averaged over a predictive distribution $p(y_{miss}|y_{obs})$ of the missing data. P4 provides an optional modelling condition, related to the formulation of $p(y_{miss}|y_{obs})$, under which collecting these missing data is expected to reduce uncertainty. Blue shading highlights known inputs, red shading highlights possibly unknown inputs, green shading highlights framework outputs, and yellow shading highlights the optional coherence condition.

including wastewater data in the model (compared to modelling reported cases only)? And second, given current wastewater sampling in Aotearoa New Zealand, how much could uncertainty about R_t be reduced by sampling wastewater catchments covering the entire population every day? We assume decision-makers use point estimates of R_t, with a quadratic loss function deemed suitable, and therefore use variance as our measure of uncertainty.

To assess the benefit of incorporating already-observed data, we compute the daily uncertainty reduction (Equation 3.5). In this setting, y_{miss} are the daily wastewater observations $W_{1:T}$, and y_{obs} are daily reported case data $C_{1:T}$. The uncertainty reduction in R_t at time-step $t \in 1, 2, \ldots T$ is given by $\mathrm{Var}_{p(R_t|C_{1:T})}(R_t) - \mathrm{Var}_{p(R_t|C_{1:T},W_{1:T})}(R_t)$, where $p(R_t|C_{1:T})$ is the

posterior distribution of R_t given only the reported case data, and $p(R_t|C_{1:T}, W_{1:T})$ is the posterior distribution of R_t given both the reported case data and the wastewater data.

To estimate the expected uncertainty reduction in R_t from daily wastewater sampling of the entire population, we first fit the model to the observed data $C_{1:T}$ and $W_{1:T}$, obtaining the joint posterior distribution over infection incidence $I_{1:T}$ and nuisance parameter vector θ, denoted $p(I_{1:T}, \theta|C_{1:T}, W_{1:T})$. We then simulate hypothetical missing wastewater data $W'_{1:T}$ by assuming the daily catchment population equals the missing population and sampling from the posterior predictive distribution $p(W'_{1:T}|C_{1:T}, W_{1:T}) = \int\int P(W'_{1:T}|I_{1:T}, \theta)P(I_{1:T}, \theta|C_{1:T})dI_{1:T}d\theta$. We fit the model to the simulated dataset (reported cases and the daily population-weighted average of $W_{1:T}$ and $W'_{1:T}$), calculate the associated uncertainty, and repeat this 100 times to obtain a Monte Carlo estimate of the expected uncertainty reduction.

We fit the model to data between 1 January 2023 and 1 March 2023 (Figure 3.6-A), sourced from [42]. Wastewater data were collected on 41 out of these 60 days (Figure 3.6-B), with daily catchment coverage ranging from 31,000 to 3.6 million individuals (out of 5.15 million total, Figure 3.6-C).

Including wastewater data reduces expected daily uncertainty about R_t by an average 16.9%, with the greatest reduction being 54.4% on 7 January 2023, though some increases (up to 12.6% on 31 January 2023, 12.1% on 17 and 18 February 2023) occur on some days (Figure 3.6-D). Sampling wastewater data from the entire population every day is expected to further reduce daily uncertainty by 14.6% (or 29.4% when compared to estimates from reported case data only), representing a plausible bound on the expected reduction in R_t uncertainty achievable through wastewater sampling.

Missing wastewater data may (once collected) differ systematically from observed data, for example if existing sampling is biased towards sites expected to yield more reliable measurements. In this case, we might wish to allow the variance of the observation distribution for the wastewater data to be larger in the hypothetical-data setting than in the observed data. To maintain coherence and ensure more data reduces our model uncertainty, we would also need to include this assumption in the model itself.

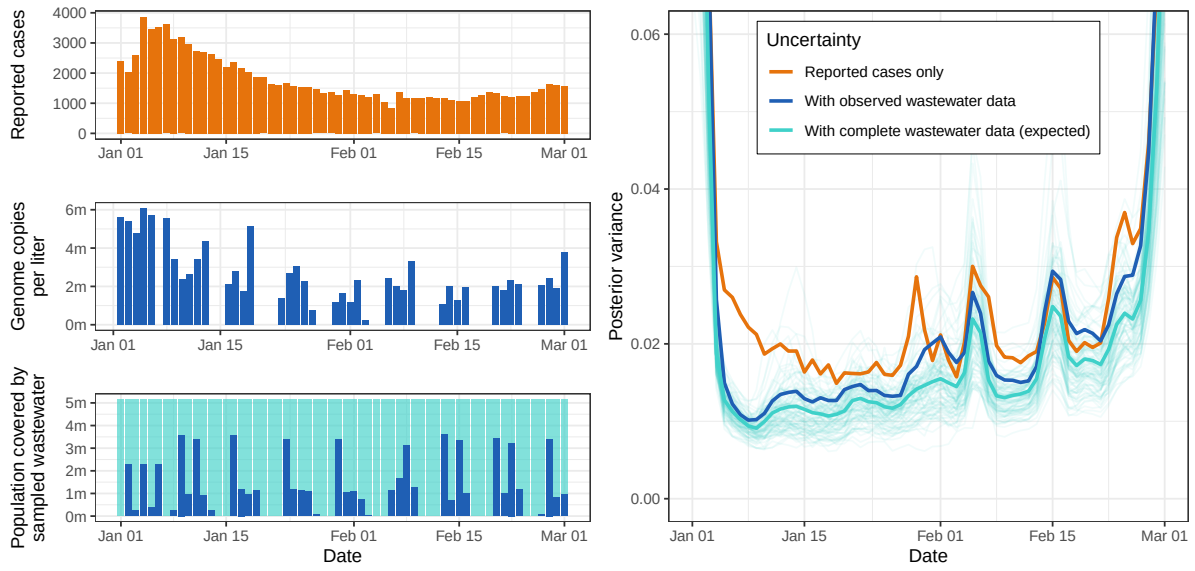

Figure 3.6: Reported case data (A), wastewater data (B), and population covered by wastewater catchment areas (C) for Aotearoa New Zealand between 1 January 2023 and 1 March 2023. The light blue region in panel (C) represents the population from which missing data could be sampled. The daily uncertainty about R_t from the model fit to reported case data only (orange), from the model fit to both reported case data and observed wastewater data (dark blue), and the expected uncertainty after sampling wastewater data from the entire population every day (light blue) are shown in panel (D). Thin light blue lines show the posterior variance of R_t from individual simulations of hypothetical data, demonstrating that even though the expected uncertainty reduction is positive (as we satisfy the coherence requirements), uncertainty about R_t could still plausibly increase. Uncertainty increases at the start and end of the time period as a result of fewer data points being available to fit the model.

3.4 Related frameworks

Our decision-theoretic approach captures multiple popular methodologies in epidemiology under a single umbrella. We outline some of these here.

The **value of information** (VOI) framework, widely used in health economics, measures the expected reduction in loss from obtaining additional data [81, 199, 200, 201, 218, 219]. A central concept in the VOI framework is the *expected value of sample information* (EVSI), which quantifies the expected reduction in loss from collecting new data y_{new}, written in our notation as:

$$\text{EVSI}(y_{new}) = h\left[p(z|y_{obs})\right] - \text{E}_{p(y_{new}|y_{obs})}\left[h\left[p(z|y_{obs}, y_{new})\right]\right]. \tag{3.9}$$

The expectation is taken with respect to the posterior predictive distribution for the new data $p(y_{new}|y_{obs})$, thus the EVSI is equivalent to our Equation 3.6 in a standard Bayesian

setting. Our framework extends the VOI literature beyond the traditional Bayesian setting by allowing for a wider class of predictive distributions and models to be used for valid uncertainty quantification.

The VOI framework is also applied to decision problems, where the loss function is defined in terms of the cost of making a decision based on the model. Our framework adds a layer of interpretation to this setting: the minimised expected cost of the decision is a direct measure of uncertainty about the decision being made.

Bayesian experimental design and **active learning** are focused on data collection strategies, arguing that experiments should be designed to maximise their expected utility, commonly framed in terms of the *expected information gain* (EIG) [220]. Letting ξ denote the experimental design that collects data y_{new} (a yet-to-be-observed random variable), the EIG about z is defined (in our notation) as:

$$\begin{aligned} \mathrm{EIG}_z(\xi) &\coloneqq \mathrm{E}_{p(y_{new}|\xi)}\left[\log p(z \mid y_{new}, y_{obs}, \xi) - \log p(z \mid y_{obs})\right] \\ &= \mathrm{E}_{p(y_{new}|\xi)}\left[\mathrm{H}\left[p(z \mid y_{obs})\right] - \mathrm{H}\left[p(z \mid y_{new}, y_{obs}, \xi)\right]\right], \end{aligned} \tag{3.10}$$

where $\mathrm{H}\left[p(z \mid y_{obs})\right] - \mathrm{H}\left[p(z \mid y_{new}, y_{obs}, \xi)\right]$ is the information gain about z from observing y_{new} from experiment ξ. This can be viewed as a special case of Equation 3.6 under a log-loss function. By Principle 4, an experiment ξ is only guaranteed to reduce uncertainty if it is equivalent to sampling from the posterior predictive distribution, opening a discussion on model specification which we leave for future work. Adaptive learning extends the method to allow for updating ξ as data are collected.

Case et al. [196] apply experimental design principles in epidemiology, specifically in designing tick surveillance strategies, although they do not directly target the EIG. Instead, they target two criteria: Bayesian d-optimality (approximately equivalent to maximising the EIG [221]) and a bespoke loss function that minimises the maximum standard deviation of estimates in "high-risk" regions. Our framework encourages viewing the expected value of their bespoke loss function as the relevant measure of uncertainty in this setting.

Fisher information approaches can also be captured by our framework. For example, Parag et al. [205] use information-theoretic methods to quantify the information in noisy epidemic curves, which they define as the Fisher information (FI) of R_t. The FI is approximately equivalent to

the inverse-variance of the maximum likelihood estimate, so this approach effectively measures uncertainty in terms of the variance of a point estimate. The equivalent approach under our framework is the use of a quadratic loss function, examining the expected uncertainty (in this example, variance) reduction about R_t after reducing reporting delays and noise.

Finally, **scoring rules** have gained popularity in epidemiology for measuring the quality and value of probabilistic estimates [32, 39, 105, 207, 209]. Estimation of future data is distinct from estimation of model parameters in that we eventually observe the data, and thus we can eventually *calculate* the loss (score) associated with a given prediction. In our framework, applying a scoring rule to each forecast-outcome pair can be viewed as an operationalisation of the loss, thus giving a setting in which we can empirically estimate the uncertainty.

3.5 Discussion

We have presented a decision-theoretic framework for conceptualising and quantifying uncertainty in epidemiological models. Starting from the idea that uncertainty arises from missing data [98], our approach directly links theory to practical application and yields a separation of reducible and irreducible uncertainty. We then demonstrated that coherence ensures collecting additional data will reduce uncertainty on average, reinforcing missing data as the source of uncertainty. The framework was illustrated using a published model of SARS-CoV-2 wastewater surveillance in Aotearoa New Zealand [42], where the practical limit of wastewater data in reducing R_t uncertainty was estimated.

By generalising, extending, and improving existing frameworks in epidemiology, we broaden their applicability, clarifying how divergent perspectives on uncertainty often overlap. For example, while VOI is usually applied to traditional Bayesian models, our results allow a wider range of models to be used, and Principle 4 provides guidance on when the EVSI is expected to be non-negative. The VOI literature also introduces concepts such as the *expected value of perfect information* (EVPI), which quantifies the loss reduction if the model parameters were known exactly. This also induces a reducible/irreducible split, though we argue this split is less useful in our context than focusing on reductions available through data collection.

This work lays the foundation for potentially high-impact future research, including (i) application-specific loss functions, (ii) general coherent predictive models, and (iii) the evaluation of model

specification. For (i), adopting losses that reflect the decision context, or even the decision itself (instead of an estimate), could lead to more relevant notions of uncertainty. In particular, there are many cases in epidemiology where being above a certain threshold is of greater concern than being below it [222]. For (ii), while we focused on Bayesian examples, the framework supports a much broader class of models, including black-box machine-learning models. Ongoing work on martingale predictive distributions and conditionally identically distributed sequences is building our understanding of the precise coherence properties necessary to support valid uncertainty quantification, a promising line of research for epidemiological applications [98, 99, 216, 223]. For (iii), as epidemiological models are typically fit to a single data realisation, model structure necessarily plays a greater role in producing estimates than in many other fields. There is consequently greater potential for model misspecification in epidemiology, addressable by extending our framework. Potential research directions include the use of reference distributions [110], parametric extensions [224], and model averaging [224, 225].

In many epidemiological applications, the predictive distribution for y_{miss} will take the form of a simulator. In this setting, the expected uncertainty reduction can be estimated by sampling from this simulator, re-fitting the model, calculating the uncertainty, and repeating this multiple times. Averaging the uncertainties gives a Monte Carlo estimate of the expected uncertainty after collecting y_{miss}. This can be computationally expensive, particularly if the simulator is complex. Existing techniques, such as Gaussian process surrogates or neural density estimators [104, 113, 226], could accelerate this process.

By framing uncertainty as a consequence of estimation under incomplete information, we provide a theoretically grounded and practical approach to uncertainty quantification in epidemiological modelling, particularly useful for planning studies and directing surveillance efforts. Our framework supports consistent, reproducible, and policy-relevant model-based inference in infectious disease epidemiology.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

Title of Paper	A decision-theoretic framework for uncertainty quantification in epidemiological modelling
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Steyn, N., Bickford Smith, F., Mills, C., et al. (2025). A decision-theoretic framework for uncertainty quantification in epidemiological modelling. Submitted for publication.

Student Confirmation

Student Name	Nicholas Steyn
Contribution to the Paper	Conceptualisation, methodology, software, visualisation, validation, writing (original draft), and writing (review and editing).

Signature NSteyn	Date 22/09/2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description above is accurate.

Supervisor name and title	Dr. Kris V. Parag
Supervisor comments	

Signature Kris Parag	Date 22/09/2025

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 4

Paper II: Robust uncertainty quantification in popular estimators of the instantaneous reproduction number

Status: Accepted in *American Journal of Epidemiology* [39].

Authors: Nicholas Steyn and Kris V. Parag.

Authorship contributions: N.S. conceived of the study, wrote the software, and created the visualisations. Both authors designed the methodology, validated the methods, wrote the original draft, and reviewed and edited the manuscript. K.V.P. supervised the project.

Supporting information: Supplementary material is provided in Appendix A. In addition, all code is available on the corresponding GitHub repository. This repository contains all code necessary to reproduce the results in this paper, as well as Julia implementations of both EpiEstim and EpiFilter. Tutorials for the use of these methods are provided on the corresponding website.

Abstract

The instantaneous reproduction number (R_t) is a key measure of the rate of spread of an infectious disease. Correctly quantifying uncertainty in R_t estimates is crucial for making well-informed decisions. Popular R_t estimators leverage smoothing techniques to distinguish signal from noise. Examples include EpiEstim and EpiFilter, which are both controlled by a "smoothing parameter" that is traditionally selected by users. We demonstrate that the values of these smoothing parameters are unknown, vary markedly with epidemic dynamics, and show that data-driven smoothing is crucial for accurate uncertainty quantification of real-time R_t estimates. We derive novel model likelihoods for the smoothing parameters in both EpiEstim and EpiFilter and develop a Bayesian framework to automatically marginalise these parameters when fitting to epidemiological time-series data. This yields marginal posterior predictive distributions which prove integral to rigorous model evaluation. Applying our methods, we find that default parameterisations of these widely-used estimators can negatively impact R_t inference, delaying detection of epidemic growth, and misrepresenting uncertainty (typically producing overconfident estimates), with implications for public health decision-making. Our extensions mitigate these issues, provide a principled approach to uncertainty quantification, improve the robustness of real-time R_t inference, and facilitate model comparison using observable quantities.

4.1 Introduction

The instantaneous reproduction number R_t, defined as the average number of infections generated per primary infection at time t, is a popular measure of epidemic spread [138]. A value of $R_t < 1$ indicates a declining epidemic while $R_t > 1$ indicates a growing epidemic. This quantity is useful for policymakers as it gives the change in transmission required to control the epidemic, informing decisions about public health interventions [227, 228, 229, 230, 231]. As a stark example, in June 2020, an R_t estimate of 1.01 was used to justify continued school closures in Greater Manchester, England [232, 233]. Additionally, estimates of R_t are also used for forecasting, scenario analysis, and understanding the impact of interventions [234].

Many models exist to estimate R_t, including compartmental models, agent-based models, and the renewal model [193], the latter of which underlies most contemporary methods for real-time estimation of R_t from reported case time-series, including EpiEstim [114] and EpiFilter [115]. Reflecting a community drive to improve existing methods and make public health pipelines more reliable, we focus on uncertainty quantification in these two models, although our approach generalises to any real-time model, and reflects approaches used in state-of-the-art frameworks for retrospective estimation [139, 235].

Correct quantification of the uncertainty of R_t is crucial for making well-informed decisions. It is expected that the true value of R_t should fall within a 95% credible interval 95% of the time. Undercoverage occurs when this happens less often than expected, leading to over-confident and biased decision-making. Overcoverage occurs when this happens more often than expected, leading to under-confident and highly uncertain decision-making. A model that produces correct coverage is termed well-calibrated.

Despite the importance of uncertainty quantification, calibration is often neglected in epidemiological models. For example, during the COVID-19 pandemic, SPI-M in England pooled estimates from multiple groups to produce consensus R_t estimates. However, pooling proved difficult due to models "*providing estimates with lower levels of uncertainty that are not fully accounting for inherent uncertainties*" [192]. Even when correct coverage is explicitly targeted, epidemic models frequently fail to achieve it. For instance, nearly all models submitted to the *open challenge to advance probabilistic forecasting for dengue epidemics* [19] produced

over-confident predictions across various forecasting targets. A baseline model (included for comparison) demonstrated superior calibration compared to all 16 submitted models when predicting the peak of the epidemic.

Smoothing assumptions, which include penalised likelihoods [236], piecewise constant/trailing window models [22, 114, 116, 237], and latent-space models [115, 139, 238, 239], are a key source of model miscalibration [235]. Oversmoothed estimates result in delayed and overconfident estimates, while undersmoothed estimates are noisy and lack precision. Even with perfect case reporting (i.e. no observation noise), inherent stochasticity in the transmission of infectious diseases necessitates the use of smoothing to distinguish signal from noise. All popular estimators of R_t employ some type of smoothing.

Despite the importance and ubiquity of these assumptions, the philosophical and practical treatment of smoothing parameters varies by method. Some methods treat these parameters as unknown quantities to be estimated alongside R_t [139, 235, 240, 241], while others treat the choice of smoothing parameter(s) as a model selection problem and seek to find a fixed, optimal parameter value [124, 235, 242]. Some methods allow the user to choose their own values or provide heuristic default values [114, 236, 241].

We argue that, because the true dynamics of R_t are always unknown and depend on both the pathogen and the population in which the pathogen is spreading, uncertainty about the nature of these dynamics should not be ignored. This uncertainty factors in both the choice of the dynamic model itself (structural uncertainty), and the parameters associated with the chosen dynamic model (parametric uncertainty). We focus on parametric uncertainty in this paper, which on its own can cause substantial model miscalibration, and demonstrate that correct marginalisation of smoothing parameters generally improves model robustness. This result holds even if the structure of the dynamic model does not accurately reflect reality, as the increased flexibility of the model allows it to better adapt to the data.

In this paper, we derive novel likelihoods for smoothing parameters in two popular models: Cori et al. [114] and Parag [115], commonly known as EpiEstim and EpiFilter respectively[1]. We use these likelihoods to marginalise the smoothing parameters, presenting estimates of

[1]Technically, the terms EpiEstim and EpiFilter refer to the associated software packages. We follow popular convention and use these terms to describe the method.

R_t that appropriately account for uncertainty in these parameters. We also derive predictive posterior distributions and demonstrate their use in model comparison via the continuous ranked probability score (CRPS) [209], emphasising the importance of model comparison using observable quantities. We validate our methods on both simulated data, where model estimates can be compared to ground truths, and real-world data, where we investigate the practical implications on decision-making during the COVID-19 pandemic in New Zealand. For each dataset considered, we fit four models: EpiEstim and EpiFilter with default parameters and with our marginalised algorithm.

4.2 Methods

We summarise our methods here but provide full mathematical derivations and technical details in Supplementary Section A.1.

4.2.1 Background

Both EpiEstim and EpiFilter leverage the *Poisson renewal model* for R_t estimation. Letting C_s be the number of cases reported at time s and w_u be the probability that a secondary case is reported u days after the primary case (often approximated by the serial interval). The *total infectiousness* is defined as $\Lambda_t = \sum_{u=1}^{t-1} C_{t-u} w_u$. Given Λ_t and the current value of R_t, the Poisson renewal model considers the number of cases at time t to be Poisson distributed:

$$C_t \sim \text{Poisson}(R_t \Lambda_t). \tag{4.1}$$

EpiEstim assumes that, on each day t, R_t has been fixed for a trailing window of k days. Larger values of k imply that R_t has been fixed for a longer period, resulting in smoother estimates. The likelihood of observing cases between time-steps $t-k+1$ and t (denoted $C_{t-k+1:t}$) is the product of the daily Poisson likelihoods (Equation 4.1) along the trailing window. A conjugate Gamma(α, β) prior distribution is assumed for R_t, resulting in a Gamma$(\alpha_{t,k}, \beta_{t,k})$ posterior distribution for R_t given $C_{1:t}$, where the shape-parameter $\alpha_{t,k}$ and rate-parameter $\beta_{t,k}$ are:

$$\alpha_{t,k} = \alpha + \sum_{s=t-k+1}^{t} C_s, \quad \beta_{t,k} = \beta + \sum_{s=t-k+1}^{t} \Lambda_s. \tag{4.2}$$

Alternatively, EpiFilter assumes that R_t follows a Gaussian random walk with standard deviation equal to $\eta\sqrt{R_{t-1}}$. Larger values of η allow R_t to vary faster, resulting in less smooth

estimates. A grid-approximation to the exact Bayesian filtering equations is used to derive the posterior distribution of R_t given $C_{1:t}$. While EpiFilter also allows for the estimation of the smoothing distribution (R_t given past and future data), we focus on real-time estimation (R_t given data up to time t). Further discussion is included in Supplementary Section A.2.

4.2.2 Model likelihoods

We use the same framework to derive likelihoods for the smoothing parameters of both methods. Model-specific derivations are included in Supplementary Section A.1. Letting θ denote an arbitrary smoothing parameter, we begin with the predictive decomposition of the likelihood:

$$\log P(C_{1:T}|\theta) = \sum_{t=1}^{T-1} \log P(C_{t+1}|C_{1:t}, \theta). \tag{4.3}$$

The one-step-ahead likelihood can be written as:

$$P(C_{t+1}|C_{1:t}, \theta) = \int P(C_{t+1}|R_{t+1}, C_{1:t}) P(R_{t+1}|C_{1:t}, \theta)\ dR_{t+1}, \tag{4.4}$$

where $P(C_{t+1}|R_{t+1}, C_{1:t})$ is the renewal model (Equation 4.1). Thus, deriving model likelihoods relies on the derivation of the one-step-ahead predictive distribution for R_{t+1}: $P(R_{t+1}|C_{1:t}, \theta)$.

For EpiEstim, R_t depends on reported cases only on days $t-k+1$ to t, however the predictive distribution explicitly ignores data on day t, so EpiEstim's predictive distribution for R_t is Gamma-distributed with shape $\alpha_{t-1,k-1}$ and rate $\beta_{t-1,k-1}$. In this case, Equation 4.4 is a Gamma-Poisson mixture, hence $C_{t+1}|C_{1:t}$ follows a negative binomial distribution with parameters $r = \alpha_{t,k-1}$ and $p = \frac{\beta_{t,k-1}}{\beta_{t,k-1}+\Lambda_{t+1}}$. The likelihood of EpiEstim's k is then the sum of log-negative binomial probability mass functions for each day t.

For EpiFilter, the predictive distribution of R_{t+1} is a by-product of the Bayesian filtering equations, found by propagating the estimated distribution of R_t given $C_{1:t}$ forward according to the assumed Gaussian random walk. The one-step-ahead likelihood of C_{t+1} is found by taking a weighted average of the Poisson likelihood of C_{t+1} with respect to the predictive distribution of $R_{t+1}|C_{1:t}$.

Figure 4.1 provides a schematic of the model likelihood calculation. Implementations of these methods, alongside worked examples, are provided in the GitHub repository.

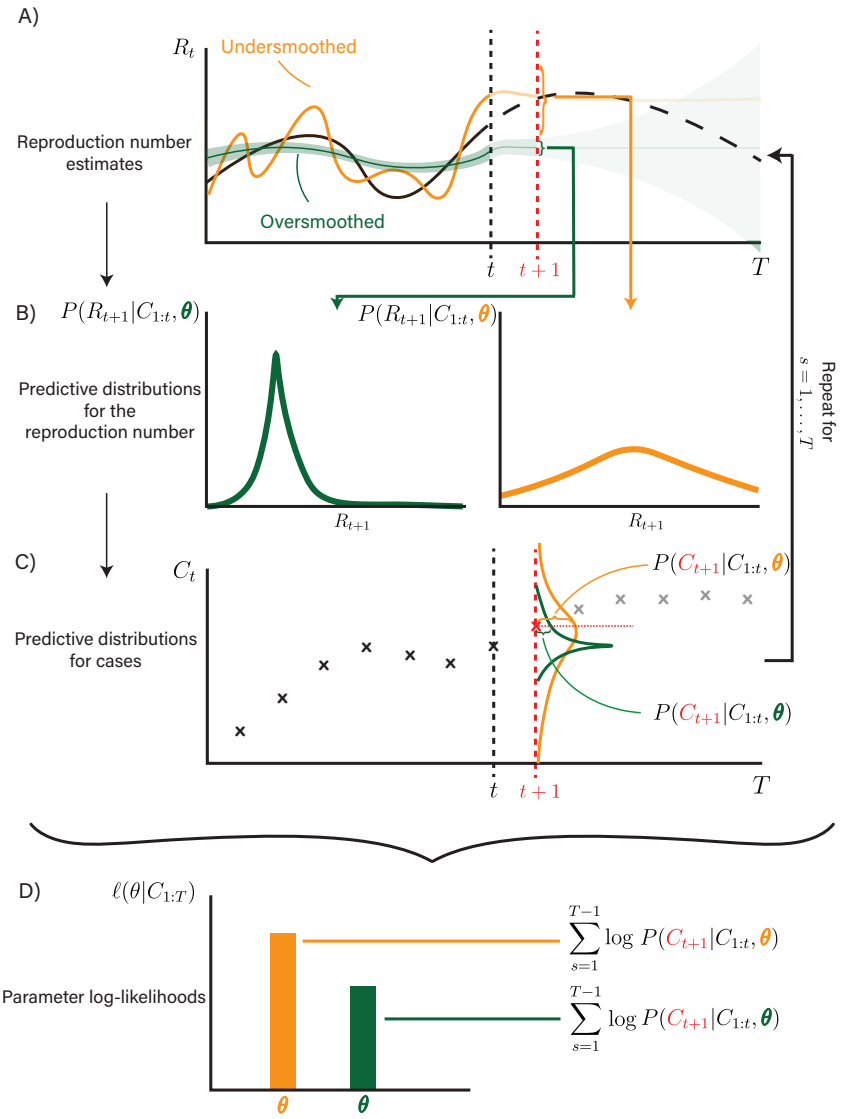

Figure 4.1: Schematic demonstrating the derivation of parameter likelihoods using the predictive decomposition (Equation 4.3). R_t estimates are projected forward one time-step according to the relevant dynamic model (panel A), giving predictive distributions for R_{t+1} (panel B). These are combined with the renewal model to give the probability of observing C_{t+1} conditional on past data $C_{1:t}$ and the chosen smoothing parameter θ (panel C). In this time-step, the undersmoothed model is more likely than the oversmoothed model. Repeating this process and summing the log-predictive probabilities for all time-steps produces the model likelihood (Panel D). These log-likelihoods are later used to marginalise out uncertainty about θ, allowing for robust reporting of uncertainty about R_t. The form of the R_t estimates and predictive distributions for R_t depend on the specific model.

4.2.3 Posterior distributions

These methods admit log-likelihoods for k and η given $C_{1:t}$, denoted $\ell(k|C_{1:t})$ and $\ell(\eta|C_{1:t})$. We use these to derive posterior distributions (denoted $P(k|C_{1:t})$ and $P(\eta|C_{1:t})$), typically using a uniform prior distribution over $k \in \{1, 2, \ldots, 30\}$ (covering daily to monthly dependence) and

$\eta \in [0, 1]$ (covering no noise to Poisson-type diffusion).

We are ultimately interested in estimates of R_t that account for uncertainty about k or η. To achieve this, we marginalise these parameters from the posterior distribution of R_t, a procedure that is rare in the literature. For EpiEstim, we leverage the discrete nature of k to write exactly:

$$P(R_t|C_{1:t}) = \sum_{k=1}^{30} P(R_t|C_{1:t}, k)P(k|C_{1:t}), \tag{4.5}$$

whereas for EpiFilter, we use a grid-approximation ($\eta \in \mathcal{E}$, Supplementary Section A.1.1):

$$\begin{aligned} P(R_t|C_{1:t}) &= \int P(R_t|C_{1:t}, \eta)P(\eta|C_{1:t})\ d\eta \\ &\approx \sum_{\eta\in\mathcal{E}} P(R_t|C_{1:t}, \eta)P(\eta|C_{1:t}). \end{aligned} \tag{4.6}$$

We can also marginalise the smoothing parameter from the predictive distributions for C_t (Equation 4.4). This generates marginal one-step-ahead predictive distributions for C_t under both models, useful for model comparison and probabilistic forecasting.

4.2.4 Model evaluation

We argue that a "good" model is one that maximises precision (i.e. minimises the width of uncertainty intervals), subject to being well-calibrated [209]. Choosing the "best" model thus involves a trade-off: how much miscoverage are we willing to accept in exchange for more precise estimates? We use the Continuous Ranked Probability Score (CRPS), which measures the distance between the estimated predictive distribution and the empirical distribution of the data, to quantify this trade-off. Smaller distances signify closer alignment between the model's predictive uncertainty and observed data variability. Full details of the CRPS calculation are provided in Supplementary Section A.1.4.

4.2.5 Data

We test our methods on three simulated datasets, each assuming a different dynamic model for R_t: a Gaussian random walk (matching the dynamic model assumed by EpiFilter), a sinusoidal curve, and a step-change model. These models cover a range of smooth to sharp changes in R_t. We also compare model outputs on real-world data from the COVID-19 pandemic in New Zealand [243], chosen as an example of high-quality data with limited reporting biases. We explicitly relate real-world decision-making to the inferences made by our models.

A common serial interval from the COVID-19 literature, a Gamma distribution with a mean of 6.5 days and standard deviation of 4.2 days [118, 176], is used for the simulation study, while a Weibull distribution with a mean of 5.0 days and standard deviation of 1.9 days is used when fitting to real-world data, matching the serial interval used in official modelling [244].

Further information on simulated data is provided in Supplementary Section A.3. We also develop additional marginalisation routines to handle uncertainty in serial intervals, and test sensitivity to biases in the serial interval, in Supplementary Section A.4.

4.3 Results

4.3.1 Simulation study

Fitting EpiEstim and EpiFilter to a single realisation from each of the three simulated epidemics (Figure 4.2) demonstrates that default parameterisations of both models result in oversmoothed estimates of R_t, relative to the level of smoothing estimated from the data. The exception is EpiFilter in the random walk simulation, where the true value of η is deliberately chosen to match EpiFilter's default $\eta = 0.1$.

Using these default parameters results in credible intervals for R_t that typically undercover the true value. This is more noticeable for EpiEstim, where coverage of the 95% credible intervals for R_t in the default model ranges from just 8.9% in the sinusoidal simulation, to 74.4% in the step-change simulation. Marginalising out k dramatically improves coverage in these models to 81.1% and 94.4% respectively. Default EpiFilter is generally more robust, partially as a result of the default $\eta = 0.1$ being less extreme with respect to the posterior distribution of η, although marginalising this parameter still improves coverage of R_t from 81.1% to 92.2% in the sinusoidal model.

The one-step-ahead predictive coverage of reported cases is also improved by marginalising the smoothing parameter. This is true for all models and simulations considered, but the effect is more pronounced in EpiEstim.

Finally, marginalising the smoothing parameter generally improves (decreases) the CRPS, suggesting that marginalised models produce more accurate predictive distributions of cases than default models. The sole exception is EpiFilter in the step-change simulation, where the CRPS

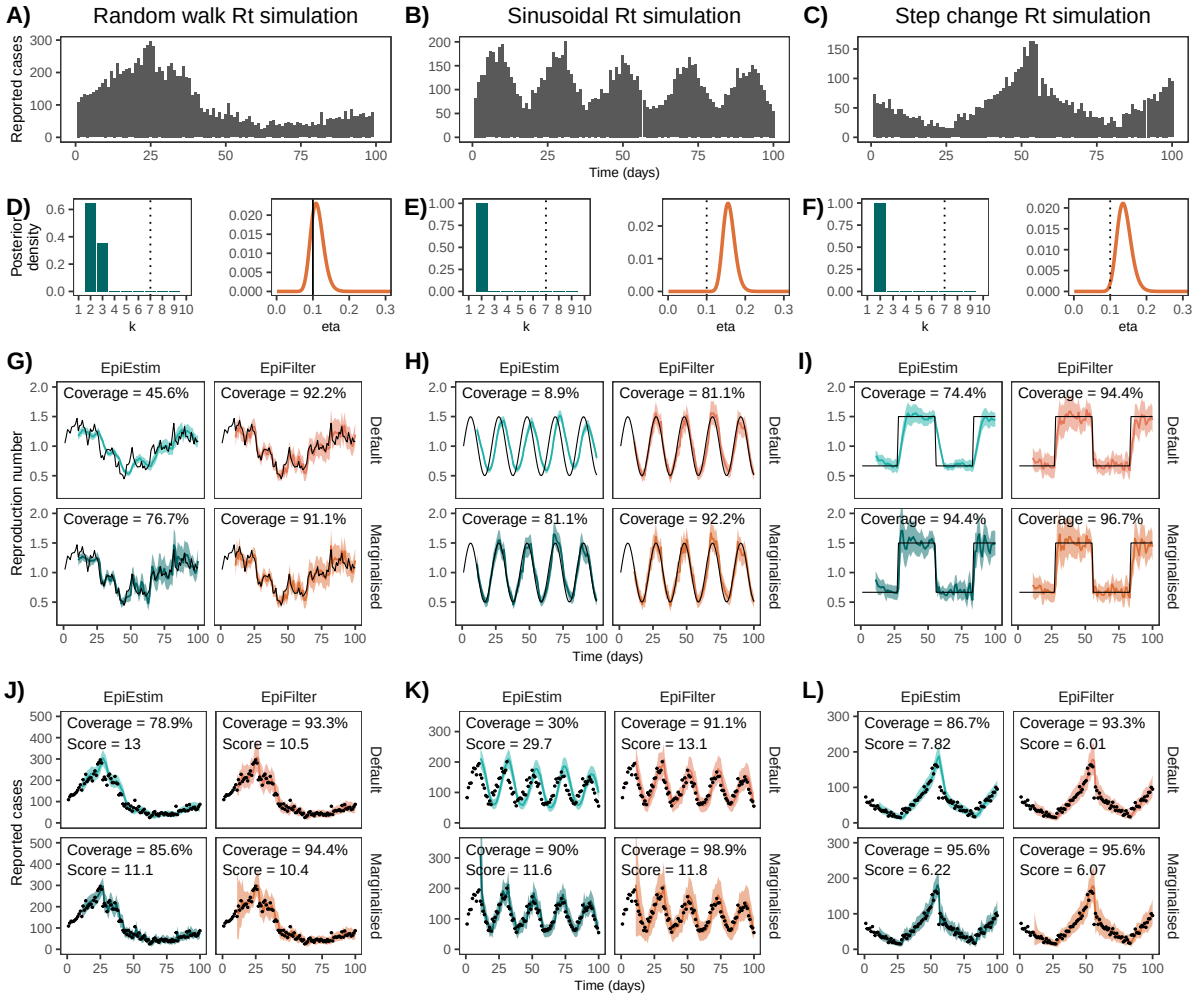

Figure 4.2: Simulated case data (A, B, C), posterior distributions for smoothing parameters at $t = 100$ (D, E, F), estimates of R_t (G, H, I), and estimates of predictive cases (J, K, L) for one realisation of each simulated epidemic. Estimates of R_t and predictive cases are shown for all four models (default EpiEstim, default EpiFilter, marginalised EpiEstim, marginalised EpiFilter). The first column (A, D, G, J) shows results for the Gaussian random walk simulation with $\eta = 0.1$, a dynamic model that precisely matches default EpiFilter. The second column (B, E, H, K) shows results for the sinusoidal simulation, and the third column (C, F, I, L) shows results for the step-change simulation. In panels (D, E, F), vertical dotted lines indicate default parameter values and the vertical solid line indicates both the default and true parameter value. In panels (G-L), solid coloured lines show central estimates (posterior means) and shading shows 95% credible intervals. Black lines (in R_t estimates) and black dots (in predictive C_t estimates) show the true values of R_t and C_t respectively. Predictive coverage of the 95% credible intervals (closer to 95% is better) and the CRPS (lower is better) are shown as text within each figure.

worsens (increases) slightly, although we show in Supplementary Section A.5 that, on average, marginalisation also improves EpiFilter when fit to step-change simulations. The higher CRPS in this specific simulation suggests that the narrower credible intervals produced by default EpiFilter may trade off favourably against its lower predictive coverage of cases. We observe

a similar effect in the sinusoidal simulation when comparing EpiEstim and EpiFilter, where the CRPS is slightly lower in EpiEstim (indicating it as the better model), despite EpiFilter having better coverage of predictive cases.

These results depend on the simulated ground truth, and we consider different simulations in Supplementary Section A.5, revealing how the appropriate level of smoothing depends on the underlying epidemic dynamics. Marginalisation becomes more important as the standard deviation of the simulated random walk increases, the frequency of the sinusoidal curve increases, or the step-change becomes more frequent. In all these scenarios, the true R_t is more dynamic and the default smoothing choices are progressively worse at adapting to these changes.

We also find, contrary to intuition, that fitting to greater numbers of daily cases does not necessarily improve inference quality. Supplementary Figure A.4 demonstrates that, while marginalised EpiFilter is largely robust to sample size, EpiEstim's coverage worsens as sample size increases in both the default and marginalised models. This occurs as guaranteed misspecification (R_t cannot be constant on $[t-k+1, t]$ and then constant at a different value on $[t-k+2, t+2]$) results in the model becoming more confident in the incorrect estimate. For similar reasons, if observation noise is large, both models are also expected to degrade as sample size increases.

4.3.2 The COVID-19 pandemic in New Zealand, August-December 2021

After largely containing the spread of COVID-19, in August 2021 an outbreak of the delta-variant was detected in Auckland, New Zealand, triggering strict non-pharmaceutical interventions. The outbreak featured an initial peak in late August, followed by a subsequent period of decline, and then a second peak in mid-November (Figure 4.3). R_t was repeatedly cited during decision-making, including by the Prime Minister and the Director-General of Health [228, 229, 245, 246, 247, 248, 249].

We fit all four models to reported cases (smoothed using a 5-day moving average to decrease reporting noise; Supplementary Section A.6 tests sensitivity to this assumption) between 11 August and 1 December 2021 (Figure 4.3). Both EpiEstim and EpiFilter exhibit improved CRPS after marginalisation, and EpiEstim exhibits improved predictive coverage after marginalisation. As R_t is unknown, we cannot evaluate its calibration.

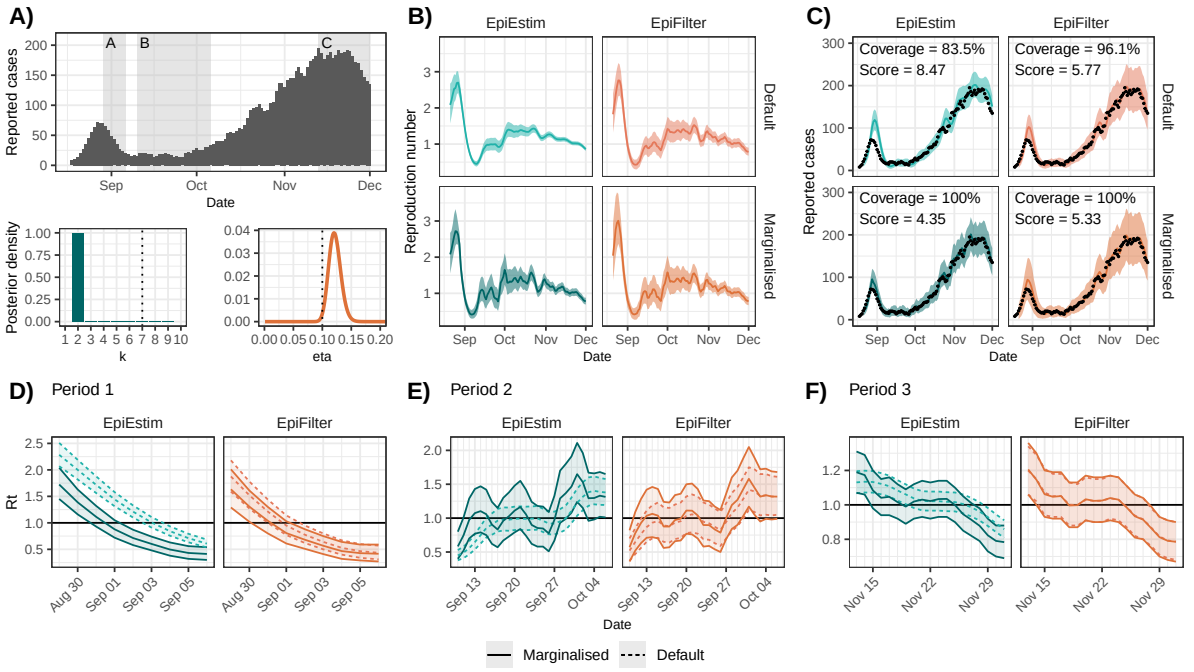

Figure 4.3: Reported case data and parameter posterior estimates (A), reproduction number estimates (B), predictive cases (C) and reproduction number estimates for selected periods (D, E, F) from fitting the four models to reported case data from the outbreak of the delta-variant of SARS-CoV-2 in Auckland, New Zealand between 11 August 2021 and 1 December 2021. In panel A, vertical dashed lines indicate default parameter values while coloured curves show parameter posterior distributions. In panels (B, C), solid lines show central estimates (posterior means) and shading shows posterior 95% credible intervals. Black points show observed reported cases. In panels (D, E, F), middle lines show central estimates (posterior means) while shading and outer lines show posterior 95% credible intervals. Results for default and marginalised EpiFilter in period 3 (Panel F) are nearly identical and overlap considerably. Our methods improve predictive coverage of EpiEstim and the CRPS of both models.

As in the simulations, default models oversmooth relative to marginalised models. This is most notable for EpiEstim where almost all posterior mass is on $k = 2$. For EpiFilter, the maximum a posteriori (MAP) value of η was 0.12 with a 95% credible interval of $(0.102, 0.143)$, which nearly contains the default value. In this case, default EpiFilter is better calibrated than default EpiEstim. However, Table 4.1 still highlights some delays in detection of epidemic growth or decline.

On 31 August, marginalised EpiFilter first signalled $R_t < 1$. Among the four models, default EpiEstim signalled this change last (3 days later). Notably, both marginalised models were confident that $R_t < 1$ by 2 September, before the possibility of decline had been detected by default EpiEstim. This was a period characterised by daily press conferences and strict stay-at-home orders, estimated to have cost the city NZ$56m per day [250].

Table 4.1: Dates of the first detection of growth or decline (defined as the upper bound on the 95% credible interval of R_t crossing 1 for growth, or the lower bound crossing 1 for decline), the date that the central estimates of R_t first crossed 1, and the date that the models are first confident in growth or decline (defined as the lower bound on the 95% credible interval of R_t crossing 1 for growth, or the upper bound crossing 1 for decline) for the COVID-19 outbreak in Auckland, New Zealand that started in August 2021. The number of days after the first detection is shown in parentheses, highlighting the substantial delay often observed in default models.

	EpiEstim		EpiFilter	
	Default	Marginalised	Default	Marginalised
Period 1				
First detection	3 Sep (+3)	31 Aug (+0)	1 Sep (+1)	31 Aug (+0)
Expected $R_t < 1$	4 Sep (+3)	1 Sep (+0)	1 Sep (+0)	1 Sep (+0)
Confidence in $R_t < 1$	4 Sep (+2)	2 Sep (+0)	2 Sep (+0)	2 Sep (+0)
Period 2				
First detection	15 Sep (+4)	11 Sep (+0)	13 Sep (+2)	11 Sep (+0)
Expected $R_t > 1$	30 Sep (+17)	13 Sep (+0)	20 Sep (+7)	13 Sep (+0)
Confidence in $R_t > 1$	1 Oct (+1)	30 Sep (+0)	1 Oct (+1)	30 Sep (+0)
Period 3				
First detection	20 Nov (+5)	15 Nov (+0)	15 Nov (+0)	15 Nov (+0)
Expected $R_t < 1$	27 Nov (+9)	19 Nov (+1)	19 Nov (+1)	18 Nov (+0)
Confidence in $R_t < 1$	29 Nov (+1)	28 Nov (+0)	28 Nov (+0)	28 Nov (+0)

On 13 September, the Government announced that because restrictions had *"...reduced that value [R_t] down to consistently below one"*, they would be relaxed the following week on 21 September [247]. There was considerable interest in R_t over that week, as a resurgence could have triggered prolonged restrictions. This announcement occurred after the first detection of a potential resurgence by the marginalised models (11 September), while default EpiFilter first detected this resurgence on the day of the announcement, and default EpiEstim two days later. The marginalised models first produced central estimates of $R_t > 1$ on the same day as the announcement, while it was 7 days and 17 days later that default EpiFilter and default EpiEstim estimated $R_t > 1$, respectively.

Daily reported cases continued to increase until mid-November, before appearing to plateau. While the marginalised models (and default EpiFilter) became uncertain about continued epidemic growth around 15 November, default EpiEstim was still confident that $R_t > 1$ until 20 November (for 5 days longer), a clear example of oversmoothed models being overconfident.

We test our methods on additional real-world datasets in Supplementary Section A.6. These examples highlight that real-world data feature many unknowns with no known ground truth. In every case but one, marginalisation decreased the CRPS, emphasising that the most robust models are those that are able to adapt to these unknowns, as well as the importance of validation on observable quantities.

4.4 Discussion

We have derived and analysed novel likelihoods for two popular R_t estimators. We used these likelihoods to develop posterior distributions for the corresponding smoothing parameters, and marginalised out this parametric uncertainty from estimates of R_t. Our algorithms generally improve uncertainty quantification for R_t, allowing increased confidence in the calibration of reported credible intervals.

Robust uncertainty quantification is crucial for real-world decision-making. Our methods provide one of the first principled and computationally efficient ways of ensuring the robustness of reported uncertainty in two popular models: EpiEstim [114] and EpiFilter [115], aligning these methods with state-of-the-art smoothing approaches [139, 235]. Furthermore, our methods can be applied to other sequential models (i.e., any model that estimates R_t using $C_{1:t}$ and then R_{t+1} using $C_{1:t+1}$ and so on), such as [238, 251, 252], all of which currently assume fixed values of smoothing parameters.

In addition to real-time decision-making, estimates of R_t are also used in models investigating the impact of non-pharmaceutical interventions [173, 253, 254, 255, 256, 257, 258, 259, 260], the effect of climate [261], or the relationship between mobility and transmission [167]. These models all use EpiEstim to estimate R_t, most using $k = 7$. While some correct for smoothing-induced delays by shifting estimates by $k/2$ days, this deterministic correction still ignores uncertainty about k. The Bayesian nature of our methods allows for the propagation of uncertainty in k through to estimates of R_t and thus to downstream models, such as the consensus estimates of R_t produced by SPI-M in England during the COVID-19 pandemic.

While we focus on parametric uncertainty, by fitting our models to simulated data from a range of dynamic models, our work reveals that marginalising smoothing parameters improves model robustness, even when ignoring structural uncertainty. This is observed by improved

coverage of predictive cases and decreased CRPS in simulations and on real data, by improved coverage of R_t in simulations, and is a key finding. As R_t is always unknown, it is not possible to compare coverage of this quantity on real data, and we rely on the CRPS on observables as a proxy for model performance (justified by our results on simulated data).

We used two metrics to evaluate model performance: coverage of the 95% credible intervals and CRPS. The former measures calibration only, while the latter also factors in precision (and simultaneously considers all credible intervals, not just the 95% level). Calibration and precision are often a trade-off: improved calibration can be obtained by decreasing precision. The CRPS is a principled way of balancing these two goals. Alternative scoring rules may be appropriate depending on context [262]. The documentation of [263] summarises scoring rules in an epidemiological context.

While our results suggest that improved CRPS implies improved estimates of R_t even when the model is misspecified, there is no guarantee of performance in such situations. Since misspecification is inevitable when modelling infectious diseases, any scoring rule should be interpreted with caution. One example faced by almost all R_t estimators is serial interval misspecification, addressed in Supplementary Section A.4. Like existing literature [116, 124, 264], we find that misspecified serial interval distributions lead to biased estimates of R_t, although estimates near $R_t = 1$ are more robust. Another example of model misspecification comes from observation noise: EpiEstim and EpiFilter assume that reported cases follow the Poisson renewal model and that the appropriate level of smoothing is fixed over time (Supplementary Section A.7). We demonstrate that our methods help with robustness to observation noise in Supplementary Section A.5.3, but the lack of explicitly representing such noise remains a limitation of both models. We chose the New Zealand dataset as an example of high-quality data with limited reporting biases to reduce the impact of this limitation.

Alternative parameter selection procedures for EpiEstim have previously been proposed. In the supplement of [114], the authors suggest selecting k such that the window contains sufficient cases to reduce the posterior coefficient-of-variation to a desired level. Depending on philosophy, this is either a subjective decision about the bias-variance trade-off, or a way of choosing parameters to obtain a desired confidence level. In either case, choosing a value of k with low likelihood will lead to poor model calibration. Alternatively, Parag and Donnelly

[242] proposed an information-theoretic approach to selecting k called APEestim. While their approach is principled, it results in the selection of k shifted by one unit compared to the MAP value from our method, and does not allow for the marginalisation of uncertainty when there are multiple plausible values of k (Supplementary Section A.8). EpiEstim and EpiFilter are Bayesian estimators, thus marginalising smoothing parameters is a more justified approach than selecting a single value. This is supported by Supplementary Section A.2, where two implementations of a different R_t estimator (EpiLPS [235]) are compared: one implementation optimises the smoothing parameters while the other performs marginalisation. We find marginalisation continues to offer improved performance and robustness (Supplementary Table A.1 and Supplementary Figure A.1).

Other methods approach the smoothing problem differently. For example, EpiNow2 [139] models R_t using a Gaussian process. The smoothness of this model is determined by the covariance kernel, which is estimated alongside R_t. Comparisons with additional methods (EpiNow2 [139], EpiLPS [235], and rtestim [124]) are included in Supplementary Section A.2. These methods include features such as explicitly representing observation noise, at the cost of increased mathematical and computational complexity (typically requiring Monte Carlo methods for marginalisation of parameters), when compared to EpiEstim and EpiFilter (where smoothing parameters are marginalised with our deterministic approach). We also provide novel methodology for finding posterior predictive distributions and calculating CRPS values for these models, providing model comparison techniques using observed data. While EpiEstim is often outperformed by these other methods, there is value in both the interpretability of the sliding window and in the simplicity of a conjugate prior-posterior for R_t. Understanding the nuances of various approaches to smoothing and hence inference-based decision-making will form a future study.

The August 2021 outbreak of SARS-CoV-2 in New Zealand provides a pertinent example of the practical importance of smoothing assumptions. R_t was first reported by officials as being less than 1 on 29 August, two days before any of our models. While official models were also based on the renewal model, among other differences (e.g., accounting for asymptomatic infections), they approached the smoothing problem differently, assuming R_t was fixed prior to the lockdown on 18 August, and then step-changed to a different fixed value after the lockdown

[244]. These piecewise-constant assumptions allowed more data to inform each estimate of R_t, reducing uncertainty. However, if these assumptions were incorrect then uncertainty about R_t will have been underestimated.

It is often argued that public health decision-making should be "data-driven", with R_t frequently featuring as an example of such "data" [265]. However, without an accurate representation of uncertainty, estimates of R_t risk being influenced more by assumptions than the underlying data. As demonstrated on both simulated and real-world data, public health decisions made using oversmoothed estimates of R_t will be delayed and overconfident relative to decisions made using estimates with more robust uncertainty quantification. Fortunately, our methods provide a simple and computationally efficient way to improve the robustness of these estimates, and to benchmark the uncertainty surrounding smoothing assumptions.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

Title of Paper	Robust uncertainty quantification in popular estimators of the instantaneous reproduction number
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Steyn, N. & Parag, K. V. (2025). Robust uncertainty quantification in popular estimators of the instantaneous reproduction number. *American Journal of Epidemiology*, kwaf165, https://doi.org/10.1093/aje/kwaf165.

Student Confirmation

Student Name	Nicholas Steyn
Contribution to the Paper	Conceptualisation, methodology, software, visualisation, validation, writing (original draft), and writing (review and editing).

Signature NSteyn	Date 22/09/2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description above is accurate.

Supervisor name and title	Dr. Kris V. Parag
Supervisor comments	

Signature Kris Parag	Date 22/09/2025

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 5

Paper III: A primer on inference and prediction with epidemic renewal models and sequential Monte Carlo

Status: Published as a *Tutorials in Biostatistics* article in *Statistics in Medicine* [40].

Full author list: Nicholas Steyn, Kris V. Parag, Robin N. Thompson, and Christl A. Donnelly.

Authorship contributions: N.S. conceived of the study, designed the methodology, wrote the software, and wrote the original draft. N.S. and R.N.T. created the visualisations. All authors validated the methods and reviewed and edited the manuscript. C.A.D. and K.V.P. supervised the project.

Acknowledgements: We would like to thank Cathal Mills for helpful discussion and feedback throughout the development of this paper and the corresponding website, Ben Cooper and Ben Lambert for comments on an earlier version of this paper, and members of the Infectious Disease Modelling group at the Mathematical Institute in Oxford for helpful discussions throughout the project. We would also like to thank Alice Chen and Otto Arends Page for highlighting additional benefits of this framework when compared to existing methods.

Supporting information: All code required to reproduce these results, and further discussion on a range of topics, is provided online at https://nicsteyn2.github.io/SMCforRt/.

Abstract

Renewal models are widely used in statistical epidemiology as semi-mechanistic models of disease transmission. While primarily used for estimating the instantaneous reproduction number, they can also be used for generating projections, estimating elimination probabilities, modelling the effect of interventions, and more. We demonstrate how simple sequential Monte Carlo methods (also known as particle filters) can be used to perform inference on these models. Our goal is to acquaint a reader who has a working knowledge of statistical inference with these methods and models and to provide a practical guide to their implementation. We focus on these methods' flexibility and their ability to handle multiple statistical and other biases simultaneously. We leverage this flexibility to unify existing methods for estimating the instantaneous reproduction number and generating projections. A companion website *SMC and epidemic renewal models* provides additional worked examples, self-contained code to reproduce the examples presented here, and additional materials.

5.1 Introduction

Modern epidemiology relies on statistical models to track and project the spread of infectious diseases, inform public health policymaking, and estimate disease burden [266]. A commonly used model for these purposes is the semi-mechanistic renewal model, which relates past incidence to current incidence through a simple renewal process [107, 267, 268]. While this model is most commonly used for estimating the instantaneous reproduction number R_t [114, 115, 139], it can also be used for generating projections [105, 117, 139, 269], estimating elimination probabilities [118, 270, 271], and modelling the effect of non-pharmaceutical interventions [22, 173], for example. Renewal models are frequently modified to account for various biases and limitations in epidemiological data [205, 239, 264, 272]. These modifications often necessitate bespoke methods for fitting the model to data, which can be difficult to implement and become increasingly complex with additional modifications.

Reviews of reproduction number estimation highlight key practical considerations to bear in mind when fitting the renewal model to epidemiological data. Gostic et al. [36] highlight generation interval misspecification, reporting delays, right-truncation due to observation delays, incomplete observation, and smoothing windows as particular concerns. Nash, Nouvellet, & Cori [272] identify 54 papers that make modifications to EpiEstim (arguably the most popular R_t estimator) to account for such biases. They note delayed and missing reported case data, weekly reporting noise, alternative smoothing methods, imported cases, and temporal aggregation of reported data as common considerations. These adjustments are just as necessary for generating projections and other epidemiological analyses as they are for R_t estimation.

In this primer, we present a pedagogical overview of how sequential Monte Carlo (SMC) methods [119, 120], which are well established in other contexts, can be used as flexible and general-purpose tools for fitting renewal models to epidemiological data. The overall approach can be viewed as a distillation of the methods employed by Watson et al. [42]. We focus on how these methods can handle multiple biases simultaneously and provide a unified framework for a range of tasks including reproduction number estimation and generating projections.

SMC methods, often called particle filters, are a class of algorithms that are used to fit general hidden-state models to observed data. They use a collection of samples (called particles)

to represent the posterior distribution of "hidden states" at each time step (e.g., the instantaneous reproduction number or daily infection incidence) given some observed data (e.g., reported cases). These particles are projected forward according to a state-space transition model (typically an epidemic model of some kind) and weighted by an observation model (typically representing a noisy observation process). Such methods are particularly powerful as they only require simulations from the epidemic model, rather than evaluations of a likelihood, allowing for simulation-based inference [273]. Accounting for the aforementioned delays and biases simply requires incorporating these in the state-space transition and observation models, allowing a wide range of models to be fit using a single framework.

Despite being well suited to fitting epidemiological models and having a relatively straightforward and intuitive implementation, and calls for their further use [274], SMC methods are not widely used when fitting epidemic renewal models. They have seen greater use in fitting compartmental models [275, 276, 277, 278, 279, 280, 281], while examples of their application to agent-based models [282], Hawkes processes [238], and renewal models [42, 251, 252] are more limited. This discrepancy likely arises from the Markovian nature of many compartmental models, a property which is leveraged by many popular SMC methods, whereas the renewal model is explicitly non-Markovian. We are aware of two reviews of SMC methods in epidemiology [283, 284], both of which are high quality and provide complementary information to this primer, though neither covers methods applicable to the renewal model. To aid comparison with these papers, a discussion of similarities and differences with this primer is provided online.

Methods for fitting renewal models that feature comparable flexibility typically rely on Hamiltonian Monte Carlo (HMC) [139, 285, 286] (as implemented in Stan [88]), often coupled with an approximate Gaussian process state-space model (such as in EpiNow2 [139], which is available as a plug-and-play software package). However, HMC-based methods do not support discrete state spaces or parameters, can be comparatively computationally intensive, and cannot be updated online as new data are observed. Additionally, Gaussian process approximations introduce substantial mathematical complexity. Our goal is to enable a researcher to rapidly implement their own methods, ensuring they understand the process from start to finish. An alternative flexible approach is the Laplacian P-spline method of EpiLPS [235], though its fully

Bayesian implementation is also gradient-based, again limiting it to continuous state spaces and parameters, and the mathematical complexity of this method is also high. Finally, approximate Bayesian computation methods could also be used to fit renewal models, although (as for SMC) the literature is largely focused on compartmental models [287, 288].

We aim to familiarise the reader with both the discrete-time renewal model and SMC methods. We provide simple implementations of SMC-type algorithms, including a bootstrap filter with fixed-lag resampling and a particle marginal Metropolis-Hastings (PMMH) algorithm, for fitting renewal models to data. Their use is demonstrated in several epidemiologically interesting scenarios using national data from the COVID-19 pandemic in Aotearoa New Zealand. We also provide guidance on model evaluation and comparison. Our aim is not to provide a model for every scenario, but to equip the reader with the tools to construct and fit their own models.

This paper contains the technical details needed to construct and fit discrete-time epidemic renewal models with SMC methods, consolidating established methods into a single, accessible framework. Some readers may find it helpful to consult the example code online while reading, which is available as downloadable notebooks from *SMC and epidemic renewal models*. This website also contains additional examples, tutorials, and longer explanations of the topics considered here.

5.2 Hidden-state models

A typical hidden-state model consists of time-indexed unobserved hidden states X_t, observed data y_t, and parameter(s) θ. These may be scalar or vector-valued. The hidden states represent the underlying process of interest (e.g., the true infection incidence and reproduction number) from which the observed data (e.g., reported cases) are generated. The goal is to infer the value of the hidden states X_t and/or parameters θ given the observed data. SMC methods, discussed later, are a class of algorithms for fitting general hidden-state models to observed data [120].

A hidden-state model is defined by two probability distributions. The first is the **state-space transition distribution**, which governs how hidden states evolve over time:

$$P(X_t|X_{1:t-1}, \theta) \quad \text{(state-space transition distribution)}.$$

The notation $X_{s:t}$ denotes the values of this quantity at all time steps from s to t, inclusive. In epidemiological applications, this distribution is typically implied by an epidemic model describing how the epidemic evolves over time. Stochastic compartmental models are a popular choice here [283, 284], although other models (such as the renewal model) can be used.

The second probability distribution is the **observation distribution**, which relates observed data to the hidden states:

$$P(y_t|X_{1:t}, y_{1:t-1}, \theta) \quad \text{(observation distribution)}.$$

This distribution can model the effect of various biases in the data (e.g., we may model underreporting using a binomial or beta-binomial distribution). Alternatively, non-mechanistic observation models (e.g., a Gaussian observation model) can be used to allow for general noise in the data.

As these distributions model the state-space transition and observation processes, we also refer to *state-space transition models* and *observation models* throughout this primer. Together, they define a generative model of the epidemic, a simplified version of which is visualised in Figure 5.1. This structure is particularly convenient in epidemiology: the underlying epidemic process forms the hidden states, which are assumed to drive the observed data.

Other names for hidden-state models include Hidden Markov Models (HMMs), when the model is Markovian [289]; partially observed Markov processes (POMPs), also for Markovian models [290]; and latent variable models, where X_t are referred to as latent variables [97]. The renewal model we examine is not Markovian, as infection incidence today depends on a convolution of infection incidence across multiple previous time steps, thus it cannot be treated as an HMM or POMP. Augmenting past states into X_t, i.e., setting $X_t = (\tilde{X}_t, \tilde{X}_{t-1}, \ldots)$ can transform a non-Markovian model into a Markovian one, at the cost of an enlarged hidden-state space. We do not consider this further in this primer.

Many popular epidemic models can be framed as a hidden-state model. EpiFilter, which can be used to estimate the instantaneous reproduction number R_t [115] or infer the elimination probability of an epidemic [118], is explicitly formulated this way. In this model, R_t is assumed to follow a Gaussian random walk (the state-space transition model) and observed cases C_t are

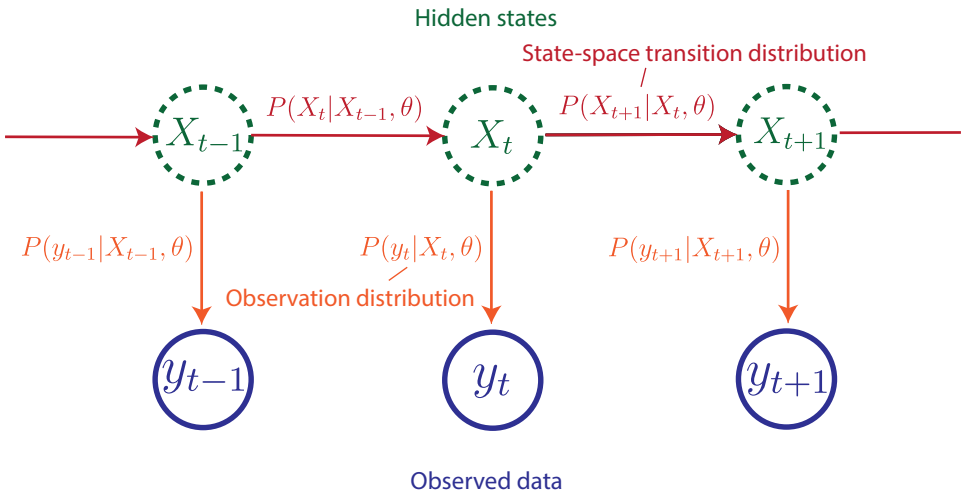

Figure 5.1: A diagram of a simple hidden-state model. The hidden states X_t evolve over time according to the state-space transition distribution, from which observed data y_t are generated via the observation distribution. Both distributions may depend on the parameter vector θ. In this example, hidden states X_t depend only on the previous hidden state X_{t-1}, and observed data y_t depend only on the hidden state at time t, making this particular model Markovian - an example of an HMM or POMP. In contrast, the more general non-Markovian models we focus on in this primer would require many more arrows in the diagram.

assumed to follow the Poisson renewal model (the observation model). A less obvious example of a hidden-state model is that by Fraser et al. [237], which shares EpiFilter's observation model but assumes R_t is piecewise constant on intervals of 10 days. Further examples include the various implementations of EpiNow2 [139], and a model for temporally aggregated and underreported data by Ogi-Gittins et al. [45]. These models all rely on the renewal model and employ bespoke inference methods, such as a grid-based approximation to the Bayesian filtering and smoothing equations in EpiFilter, maximum likelihood estimation in the model by Fraser et al. [237], or HMC in EpiNow2 [139].

5.3 The renewal model

While SMC methods can be used to fit various hidden-state models, and many epidemic models can be formulated as hidden-state models, we focus on the discrete-time epidemic renewal model in this primer. The renewal model was introduced by Leonhard Euler in 1767 and popularised in the contemporary continuous-time formulation by Alfred Lotka in 1907, who used it to model age-structured population dynamics [67]. This continuous-time formulation was first introduced into epidemiology alongside the susceptible-infected-recovered (SIR) model by Kermack and

McKendrick [65] and gained wider popularity during the 1970s [267]. The modern discrete-time formulation was proposed by Fraser [107] (see also [291]) and subsequently popularised as a Bayesian estimator of R_t in Cori et al. [114]. Since then, the discrete-time renewal model has become a popular semi-mechanistic model of disease transmission [268].

Denoting reported cases at time t by C_t, the reproduction number by R_t, and the probability that a secondary case is reported u time steps after the primary case by ω_u (often approximated by the probability mass function (PMF) of the serial interval), the renewal model relates reported cases at time t to previously reported cases through the renewal equation:

$$E[C_t] = R_t \sum_{u=1}^{u_{max}} C_{t-u}\omega_u. \tag{5.1}$$

If C_t follows a Poisson distribution with mean $E[C_t]$, then we have the popular Poisson renewal model. However, other distributions can be used [193], often to account for overdispersion in transmission or "superspreading" [292]. The serial interval is the time between the symptom onset of the primary case and the symptom onset of the secondary case, the PMF of which is represented by $\{\omega_u\}_{u=1}^{u_{max}}$, where u_{max} is the assumed maximum possible serial interval. The term $C_{t-u}\omega_u$ in the summation represents the expected contribution to current reported cases by reported cases from u time steps ago. This relationship is visualised in Figure 5.2.

In a simple example, R_t follows a Gaussian random walk (the state-space transition model), and daily reported cases have an expected value defined by Equation (5.1) (the observation model). If a Poisson observation distribution is used, then this is the model defined by EpiFilter [115], which we also adopt in the first example model.

A more realistic model might assume that underlying infection incidence follows the renewal model: Equation 5.1 is modified by replacing observed C_t with unobserved I_t, and the serial interval distribution with the generation time distribution. Reported cases are then treated as noisy observations of infection incidence. In this setup, the renewal model now defines part of the state-space transition model, and the observation process is modelled separately. Explicitly modelling infection-to-infection, instead of reported case-to-reported case, has the added benefit of allowing for potentially negative serial intervals, which would otherwise be difficult to account for, especially when fitting models to paired data with observed negative

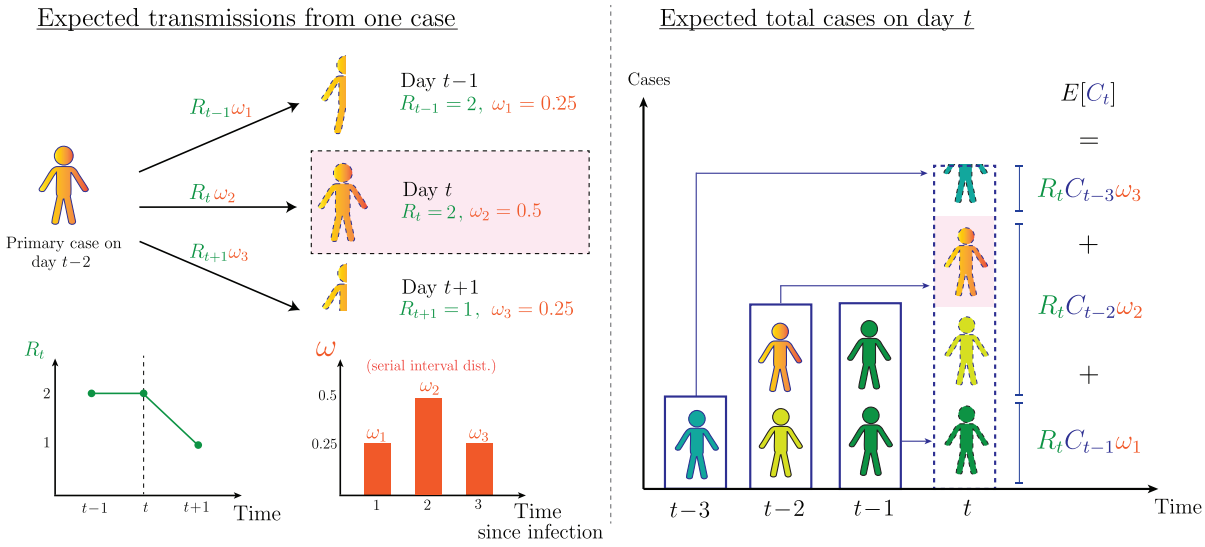

Figure 5.2: Diagram of the renewal model. In this example, the serial interval takes values $\omega_u = 0.25, 0.5, 0.25$ for $u = 1, 2, 3$ (a maximum serial interval of three time steps) and the reproduction number is assumed to take values $R_{t-1} = 2$, $R_t = 2$, and $R_{t+1} = 1$. On the left, the expected number of secondary cases produced by a primary case who was reported at time $t-2$ is shown (0.5 cases at time $t-1$, a single case at time t, and 0.25 cases at time $t+1$), with their expected contribution to total cases at time t highlighted. On the right, the expected total cases at time t is shown as the sum of the expected cases produced by primary cases reported $u = 1, 2,$ and 3 time steps ago, defining the renewal equation.

intervals. This is the approach adopted in the second and third example models. Crucially, SMC methods can be used in both cases, whereas EpiFilter is limited to the case where the renewal model defines the observation distribution.

The discrete-time nature of our model necessitates the specification of an appropriate time-step duration. The key constraint can be seen in Equation 5.1, which assumes that primary cases (or infections if we model the underlying infection incidence) cannot produce secondary cases (secondary infections) on the same day they were reported (infected). The discretisation should be fine enough that the probability of same-day transmission is negligible [293]. For COVID-19, daily time steps were popular, as this coincided with the frequency of reported case data.

In some scenarios, data may be reported on a longer timescale than infections occur. For example, weekly reporting is common in influenza surveillance, while typical influenza generation times are measured in days [45]. Published literature accounts for this using simulation-based methods [45, 293], expectation-maximisation [294], or Gaussian processes [139] to infer infec-

tion incidence at a finer timescale before applying the renewal model. The SMC approach introduced in this primer does not assume data are observed on every time step, allowing an appropriate discretisation to be chosen to suit the pathogen being modelled, independently of how frequently data are reported. This is examined in a practical example in Section 5.6.3. We also demonstrate this on an influenza dataset online.

5.4 Sequential Monte Carlo methods

We introduce two SMC-type algorithms for fitting the renewal model to data. The first is a bootstrap filter, which takes observed data $y_{1:T}$, parameter vector θ, and initial state distribution $P(X_0)$ as inputs, and targets the posterior distribution $P(X_t|y_{1:T}, \theta)$ of the hidden states X_t at each time step t. For example, this might be the posterior distributions of R_t (at each t) given observed cases $C_{1:T}$ and some smoothing parameter η. The second algorithm is PMMH, which replaces the fixed θ with a prior distribution $P(\theta)$ and targets the corresponding posterior distribution $P(\theta|y_{1:T})$.

These two algorithms can be used together to estimate the posterior distribution of the hidden states after marginalising over θ, denoted $P(X_t|y_{1:T})$. In the example above, this is the posterior distribution of R_t given $C_{1:T}$, accounting for uncertainty about η. All three of these posterior distributions can be useful, depending on the model and its intended purpose. An overview of these algorithms is provided in Figure 5.3.

SMC methods are particularly useful in epidemiology as they allow for simulation-based inference. The bootstrap filter only requires sampling from the state-space transition distribution, rather than evaluating its density. This makes it possible to fit complicated models where the likelihood is only defined implicitly by a simulator. We still need to be able to evaluate the observation model, but this is often much simpler.

When estimating the value of hidden states, it is common to distinguish between the *filtering* and *smoothing* distributions. The former is the posterior distribution of X_t given data up to time t: $P(X_t|y_{1:t}, \theta)$, while the latter is the posterior distribution of X_t given both past and future data: $P(X_t|y_{1:T}, \theta)$ - note the different subscripts on y. The uncertainty associated with the smoothing distribution is generally lower than that associated with the filtering distribution as additional data are leveraged, although we are restricted to the filtering distribution in

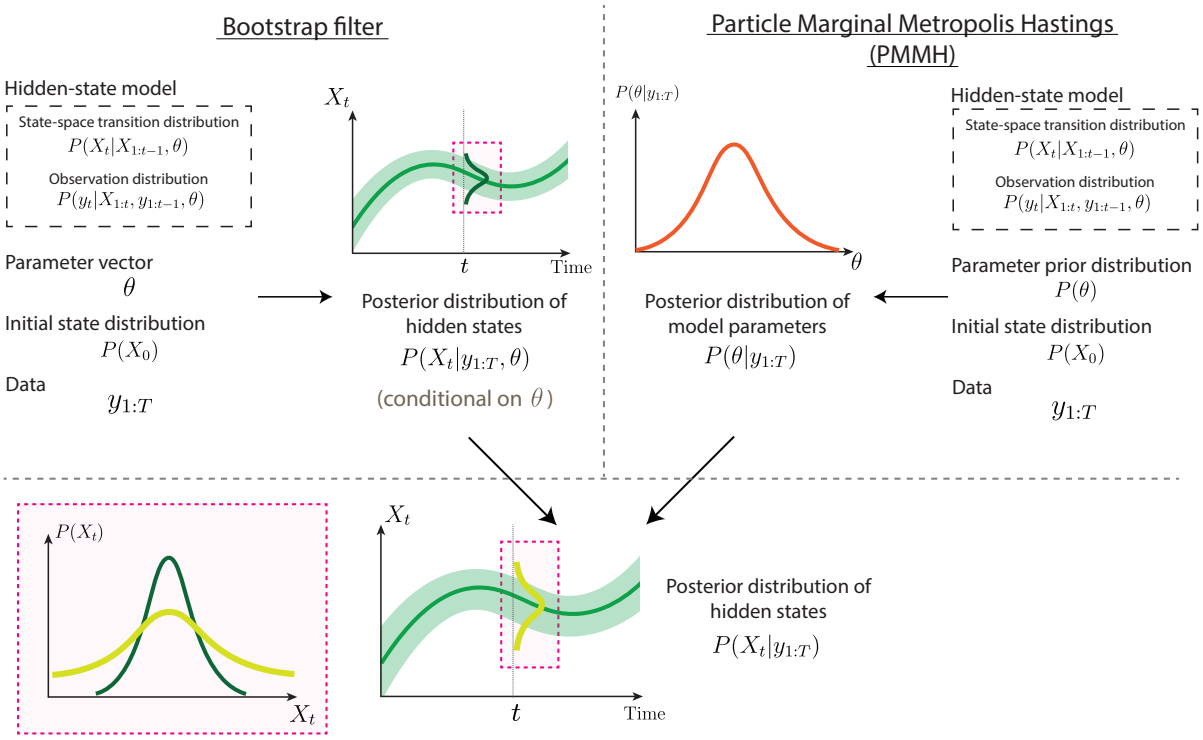

Figure 5.3: An overview of the algorithms presented in this primer. The bootstrap filter (Algorithm 1) and PMMH (Algorithm 2) both take a hidden-state model, an initial state distribution, and data as inputs. The bootstrap filter additionally requires the specification of parameter vector θ and returns samples from the posterior distribution of the hidden states at each time step, conditional on θ, shown here as central estimates and credible intervals over time (green curves). PMMH instead requires a prior distribution for θ and returns samples from the posterior distribution of θ, shown here as the orange curve. These two algorithms can be combined (by running the bootstrap filter at multiple samples of θ from PMMH) to find the posterior distribution of the hidden states, after accounting for uncertainty about θ (Algorithm 3). The pink inlay shows the posterior distribution of X_t at time t before and after marginalisation, illustrating that uncertainty about X_t is generally greater after accounting for uncertainty about θ. When the bootstrap filter is run with fixed-lag resampling, an additional hyperparameter L is required as an input.

real-time analysis. The bootstrap filter presented below can be used to sample from either distribution.

5.4.1 The bootstrap filter

The goal of the bootstrap filter [295] is to generate samples from the filtering and/or smoothing distributions, conditional on the observed data $y_{1:T}$ and parameter vector θ. This is a similar goal to EpiFilter, for example, which targets the posterior distribution of R_t given observed cases $C_{1:T}$ and parameter η [115]. The bootstrap filter is one of the simplest particle filtering methods, although more advanced methods exist and may be preferable in some settings [296].

The similarity to EpiFilter, which produces analytical solutions, offers a convenient way to check a bootstrap filter implementation: fitting the same model to the same data allows outputs to be compared directly.

The algorithm starts with an initial set of N "particles" $\{x_0^{(i)}\}_{i=1}^N$ sampled from an arbitrary initial-state distribution $P(X_0)$. These particles are projected forward by sampling from the state-space transition distribution $\tilde{x_t}^{(i)} \sim P(X_t|x_{0:t-1}, \theta)$, representing equally likely one-step-ahead projections of the hidden states before any data are observed. The projected particles are then weighted by the observation model $w_t^{(i)} = P(y_t|\tilde{x_t}^{(i)}, y_{1:t-1}, \theta)$, quantifying how likely each projected particle is once the data at time step t are observed. The particles are then resampled with replacement according to these weights, resulting in a new set of particles $\{x_t^{(i)}\}_{i=1}^N$ that represent equally-weighted samples from the posterior distribution at time step t. This process is described in Figure 5.4. Repeating for $t = 1, \ldots, T$ completes the algorithm, described formally in Algorithm 1. Further intuition in the language of sequential importance sampling is provided online. Readers familiar with alternative SMC methods may expect all weights to be reset to $1/N$ following resampling. This is unnecessary as we resample at every step, thus we always treat the resampled particles as equally-weighted samples from the target posterior distribution and past weights do not feature in calculations at future time steps (that is, there is nothing to "reset"). Furthermore, the unnormalised weights $w_t^{(i)}$ are later used when estimating the likelihood in Algorithm 2, so it is helpful to store these as initially calculated.

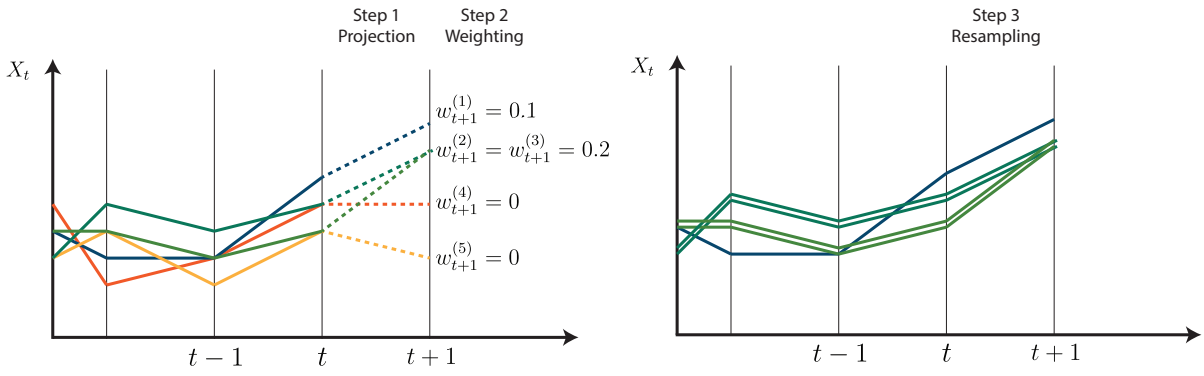

Figure 5.4: A single step of the bootstrap filter with resampling. Particles are projected forward by sampling according to the state-space transition distribution, weighted by the observation distribution, and resampled to form a new set of equally-weighted particles. In this example, the orange and yellow trajectories receive zero weight, so do not feature in the right-hand panel. The remaining trajectories feature in the right-hand panel with relative frequency proportional to their weight.

Algorithm 1 The fixed-lag bootstrap filter. Statements indexed by i should be taken to mean for i in $\{1, 2, \ldots, N\}$. At time step t, $\tilde{x}^{(i)}_{t-L:t}$ is taken to mean the aggregation of particle history $x^{(i)}_{t-L:t-1}$ and newly projected particles $\tilde{x}^{(i)}_t$.

Input: Number of particles N, fixed lag L, parameter vector θ, data $y_{1:T}$, state-space transition distribution $P(X_t|X_{1:t-1}, \theta)$, observation distribution $P(y_t|X_{1:t}, y_{1:t-1}, \theta)$, and initial state-space distribution $P(X_0)$
Sample $x^{(i)}_0 \sim P(X_0)$
for $t = 1$ to T **do**
  $\tilde{x}^{(i)}_t \sim P(X_t|x^{(i)}_{0:t-1}, \theta)$ ▷ Sample from the state-space transition distribution
  $w^{(i)}_t \leftarrow P(y_t|\tilde{x}^{(i)}_{t-L:t}, y_{1:t-1}, \theta)$ ▷ Calculate the observation weights
  $x^{(i)}_{t-L:t} \sim \text{Multinomial}\left(\{\tilde{x}^{(j)}_{t-L:t}\}_{j=1}^N, \left\{\frac{w^{(j)}_t}{\sum_{k=1}^N w^{(k)}_t}\right\}_{j=1}^N\right)$ ▷ Resample the particles
end for
Return: Particle values $x^{(i)}_t$ and observation weights $w^{(i)}_t$ for $i = 1, \ldots, N$ and $t = 1, \ldots T$.

If only the particles $x^{(i)}_t$ at time step t are resampled, then equally-weighted samples from the filtering distributions $P(X_t|y_{1:t}, \theta)$ are returned (for each t). If all particle histories $x^{(i)}_{1:t}$ are resampled at each time step (as in Figure 5.4), then equally-weighted samples from the smoothing distributions $P(X_t|y_{1:T}, \theta)$ are returned. Intuitively, this is because resampling effectively re-weights past particles according to current data.

When T is large, resampling entire particle histories leads to particle degeneracy [119], where few unique particles remain at early time steps. This is visible in Figure 5.4 as a decreasing number of unique particle trajectories after resampling, which results in poor approximations of the smoothing distribution. To mitigate particle degeneracy, we use fixed-lag resampling [296], resampling the past L time steps $x^{(i)}_{t-L:t}$ rather than the full state histories, which yields an approximation to the smoothing distribution: $P(X_t|y_{1:t+L}, \theta)$. If X_t is conditionally independent of data more than L steps in the future, given $y_{1:t+L-1}$, then this approximation is exact. In renewal models, L should exceed the maximum serial interval u_{max}, and, if reporting delays are present, L should also exceed the maximum reporting delay.

Thus far, we have assumed we are targeting the marginal posterior distribution at time t. It may also be useful to consider the joint distribution of the hidden states over multiple time steps, $P(X_{s:t}|y_{1:T}, \theta)$ for some $s < t$. This is useful if we are estimating the timing of epidemic peaks (such as peak infections or peak R_t), for example. If full state histories are resampled at each time step, then the resulting trajectories $x^{(i)}_{1:T}$ are jointly distributed according to

$P(X_{1:T}|y_{1:T},\theta)$. If fixed-lag resampling is used, then the final L time steps of each trajectory $x^{(i)}_{T-L:T}$ are jointly distributed according to $P(X_{T-L:T}|y_{1:T},\theta)$, where T is the time step at which the algorithm is halted.

Practical considerations

The choice of fixed lag L depends on the modelled maximum serial interval (or generation time) u_{max}. In many cases, such as in the COVID-19 examples considered, u_{max} (and thus L) can be chosen to be sufficiently small to ensure algorithmic efficiency without introducing noticeable bias. Some diseases, such as tuberculosis [297], feature serial intervals and generation times that are measured in weeks, months, or even years. While this poses a computational problem when using a daily discretisation (requiring a very large L), the longer serial intervals/generation times allow for the pathogen to be modelled on a weekly (or even monthly) discretisation, thus L can be kept small. The impact of L can be checked by refitting the model multiple times at different values of L and comparing the outputs. This is demonstrated online, where we find that reasonable choices of L have very little impact on an example model.

A greater number of particles N improves the accuracy of the posterior distribution approximation but increases computation time. The appropriate value of N depends on the complexity of the model, how well the model fits the data, and the purpose of the bootstrap filter, and thus is difficult to determine a priori. In practice, we find that $N = 100{,}000$ is sufficient (i.e., produces stable central estimates and credible intervals) for all models considered in this primer. This can be decreased (typically to around $N = 1{,}000$) when the algorithm is used for likelihood estimation (see Section 5.4.2 for further discussion) or when the algorithm is used to find the marginal smoothing posterior distribution (Section 5.5.1).

The choice of N for a specific model can be guided by comparing hidden-state estimates across multiple runs: if the estimates are stable (i.e., they do not noticeably change between runs), then N is likely sufficiently large. If a programmatic heuristic is required, the particle weights (line 5 of Algorithm 1) can be used to calculate the effective sample size[298], which is inversely proportional to the variance of the estimate of the posterior mean and credible interval width. The calculation of effective sample size is demonstrated on an example model online. Further validation can be performed by simulating data from the model and checking that the model

recovers the simulated truth, or by fitting a similar model with a known analytical solution (such as EpiFilter) and comparing results to the known ground truth. Theoretical convergence results [121] developed for Markovian state-space models can be applied by augmenting past states into X_t and considering the model on the expanded state space.

In Algorithm 1, for simplicity, we use the state-space transition distribution as the proposal distribution. This is known to be suboptimal, particularly because it does not use information about the observed data y_t when proposing new particles. More complex proposal distributions can be employed and the weights adjusted to account for the fact that the proposal distribution is not the true state-space transition distribution. One such example is the auxiliary particle filter [299], although we find this unnecessary for the models fit in this primer.

Due to fixed-lag resampling, estimates are usually robust to reasonable choices of the initial state distribution $P(X_0)$. In the absence of clear prior information, the initial state distribution should allow for any plausible value of X_0. The effect of the initial state distribution can be examined through sensitivity analysis.

A large proportion of SMC literature focuses on HMMs. In this primer we apply SMC methods to non-Markovian models (i.e. those where X_t depends on $X_{1:t-1}$ rather than just X_{t-1}). To ensure past particle trajectories are consistent, fixed-lag resampling must be used when fitting non-Markovian models, even when only targeting the filtering posterior distribution. In these cases, L should be chosen such that X_t does not depend on hidden-state values prior to $t - L$. For example, when the state-space transition model includes the renewal model, L should be chosen greater than the maximum serial interval.

5.4.2 Particle marginal Metropolis-Hastings

The goal of the PMMH algorithm [300] is to find the posterior distribution of parameter vector θ, given observed data $y_{1:T}$. We do this by employing a Metropolis-Hastings algorithm [301], while using the bootstrap filter (Algorithm 1) to estimate the likelihood of the proposed parameters.

The log-likelihood estimator is given by the logarithm of the geometric mean of the unnormalised particle weights $\bar{w}_t$ from the bootstrap filter (Algorithm 1) at each time step [284,

302]:

$$\hat{\ell}(\theta|y_{1:T}) = \frac{1}{T}\sum_{t=1}^{T} \log \bar{w}_t. \tag{5.2}$$

Equation 5.2 follows immediately from the predictive decomposition of the likelihood:

$$\ell(\theta|y_{1:T}) = \log P(y_{1:T}|\theta) = \sum_{t=1}^{T} \log P(y_t|y_{1:t-1}, \theta), \tag{5.3}$$

where a Monte Carlo estimator of the one-step-ahead predictive likelihood is given by the average of the unnormalised weights of the projected particles, denoted $\bar{w}_t$:

$$\begin{aligned} P(y_t|y_{1:t-1}, \theta) &= E_{X_{1:t}|y_{1:t-1}}\left[P(y_t|X_{1:t}, y_{1:t-1}, \theta)\right] \\ &\approx \frac{1}{N}\sum_{i=1}^{N} P\left(y_t|\tilde{x}_t^{(i)}, y_{1:t-1}, \theta\right) \\ &= \bar{w}_t. \end{aligned} \tag{5.4}$$

Parameter inference then proceeds by using $\hat{\ell}(\theta|y_{1:T})$ in the acceptance probability of an otherwise standard Metropolis-Hastings algorithm (Algorithm 2).

Algorithm 2 Particle marginal Metropolis Hastings (PMMH). The bootstrap filter $\mathcal{M}$ written here accepts proposed parameters θ' and returns unnormalised weights $\{w_t^{(i)}\}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The average of all weights at time t is denoted $\bar{w}_t$. The data $y_{1:T}$ and other bootstrap filter options are assumed to be included in $\mathcal{M}$.

Input: Number of parameter samples M, bootstrap filter $\mathcal{M}$, parameter prior distribution $P(\theta)$, parameter proposal distribution $q(\theta'|\theta)$.
Initialise: $\theta_1 \sim P(\theta)$
for $j = 2$ to M **do**
  $\theta' \sim q(\theta'|\theta_{j-1})$ ▷ Sample proposed parameters
  $\{w_t^{(i)}\}_{i=1,\ldots,N,\ t=1,\ldots,T} \leftarrow \mathcal{M}(\theta')$ ▷ Run the bootstrap filter (Algorithm 1) at θ'
  $\hat{\ell}(\theta') =\leftarrow \frac{1}{T}\sum_{t=1}^{T} \log \bar{w}_t$ ▷ Calculate log-likelihood estimate (Equation 5.2)
  $\alpha \leftarrow \min\left(\frac{\exp \hat{\ell}(\theta')}{\exp \hat{\ell}(\theta_{j-1})} \frac{P(\theta')}{P(\theta_{j-1})} \frac{q(\theta_{j-1}|\theta')}{q(\theta'|\theta_{j-1})}, 1\right)$ ▷ Calculate acceptance probability
  if $U \sim \text{Uniform}(0, 1) < \alpha$ **then**
    $\theta_j \leftarrow \theta'$ ▷ Accept the proposal
  else
    $\theta_j \leftarrow \theta_{j-1}$ ▷ Reject the proposal
  end if
end for
Return: Sampled parameter values $\{\theta_j\}_{j=1}^{M}$.

Practical considerations

To enable the calculation of convergence diagnostics, multiple chains (we find 4 works well) of the PMMH algorithm should be run. We also encourage the standard practice of discarding an

initial portion of each chain as a burn-in period, ensuring the retained samples are not influenced by initial values outside the stationary distribution of the Markov chain. Diagnostics used in the examples below are the Gelman-Rubin statistic $\hat{R}$ [303] and effective sample size (ESS) [301], which are reported by standard Markov chain Monte Carlo (MCMC) analysis software. We use the MCMCChains.jl package [304] in our examples.

The standard Metropolis-Hastings algorithm uses exact evaluations of the model log-likelihood $\ell(\theta)$, whereas Algorithm 1 admits a stochastic estimator $\hat{\ell}(\theta)$. The standard deviation of this estimator is a function of the number of particles N used in the bootstrap filter. It has been shown that the optimal standard deviation of $\hat{\ell}(\theta)$ is 1.2-1.3 [289, 305], which can be used to guide the choice of N. The standard deviation of the log-likelihood estimate can itself be estimated by refitting the model multiple times, we demonstrate this online. For all examples in this primer, we use $N = 1,000$ particles when estimating model likelihoods. We also note that, while larger standard deviations of $\hat{\ell}(\theta)$ result in slower convergence, model outputs are still valid.

The efficiency of the PMMH algorithm depends on the parameter proposal distribution $q(\theta'|\theta)$. We find that the heuristic multivariate normal proposal distribution with covariance matrix $\Sigma = (2.38^2/d)\hat{\Sigma}$, where $\hat{\Sigma}$ is the sample covariance matrix of previous samples and d is the number of parameters being fit, generally performs well [306].

In the examples presented in this primer, we begin the PMMH algorithm with a diagonal covariance matrix and update it every 100 iterations until it stabilises (once $|\Sigma|$ changes by less than 20% between iterations), and then begin the primary sampling. Primary sampling is also performed in chunks of 100 samples, with the $\hat{R}$ and ESS calculated at the end of each chunk. We stop the algorithm when $\hat{R} < 1.05$ and ESS > 100 for all parameters. Complete details of these procedures are provided in the model files in the GitHub repository.

5.5 General framework

Thus far we have developed an algorithm for estimating the posterior distribution of hidden states X_t given observed data $y_{1:T}$ and parameter vector θ, and an algorithm for estimating the posterior distribution of θ given $y_{1:T}$. In this section we demonstrate how these two algorithms can be used to perform inference and generate projections.

5.5.1 Robust hidden-state inference

We can couple Algorithms 1 and 2 to estimate the marginal smoothing distribution $P(X_t|y_{1:T})$, representing our beliefs about X_t after accounting for uncertainty about θ (Algorithm 3). Intuitively, we use the bootstrap filter to fit the model at multiple parameter samples $\theta^{(i)} \sim P(\theta|y_{1:T})$ and average the results:

$$\begin{aligned} P(X_t|y_{1:T}) &= E_{\theta|y_{1:T}}\left[P(X_t|y_{1:T},\theta)\right] \\ &\approx \frac{1}{N}\sum_{j=1}^{N} P(X_t|y_{1:T},\theta^{(j)}) \quad \text{where} \quad \theta^{(j)} \sim P(\theta|y_{1:T}) \end{aligned} \tag{5.5}$$

In practice, this is achieved by sampling N_θ parameter samples from the output of the PMMH algorithm (Algorithm 2), and running the bootstrap filter (Algorithm 1) at each of these samples. N particles are used in each bootstrap filter, resulting in a total of $N_\theta N$ particles approximating the marginal smoothing posterior. In the examples below, we use $N = 1{,}000$ and $N_\theta = 100$. These values are chosen to ensure that (a) a sufficiently diverse set of parameter samples is used (selection of N_θ), and (b) the posterior distribution of the hidden states is well approximated (selection of $N_\theta N$). We assess these criteria by checking that the marginal smoothing distribution is stable across multiple runs of the algorithm.

Algorithm 3 Marginal smoothing distribution sampler The bootstrap filter $\mathcal{M}$ written here accepts a parameter vector θ and returns a matrix of hidden state values. Note that inputs must satisfy $N_p = N_\theta N$.

1: **Input:** Parameter samples from PMMH $\{\theta_j\}_{j=1}^{M}$, number of unique parameter samples N_θ, per-filter posterior samples N, target posterior samples N_p, and bootstrap filter $\mathcal{M}$.
2: **Initialise:** Pre-allocate matrix X of size $N_p \times T$.
3: **for** $i = 1, \ldots, N_\theta$ **do**
4: $\quad \theta' \sim \{\theta_j\}_{j=1}^{M}$ ▷ Sample a parameter vector θ'
5: $\quad \mathcal{I} \leftarrow \{(i-1)N+1, \ldots, iN\}$ ▷ Specify the indices of the ith block of X
6: $\quad X_{\mathcal{I},1:T} \leftarrow \mathcal{M}(\theta')$ ▷ Run the bootstrap filter (Algorithm 1) at θ'
7: **end for**

5.5.2 Predictions

Thus far we have focused on model inference, both for hidden states X_t and parameter vector θ. We may also be interested in interpolating missing data Y_t, extrapolating future data Y_{T+k}, or projecting future hidden states X_{T+k}.

Predictive posterior distribution of observed data

Various predictive posterior distributions can be targeted. For example, the one-step-ahead predictive posterior distribution $P(Y_t|y_{1:t-1}, \theta)$ can be targeted within the bootstrap filter by sampling from the observation distribution given the projected particles. That is, sampling $y_t^{(i)} \sim P(Y_t|\tilde{x}_{1:t}^{(i)}, \theta)$, where $\tilde{x}_{1:t}^{(i)}$ are the projected particles as defined on line 4 of Algorithm 1.

Alternatively, the smoothing posterior predictive distribution $P(Y_t|y_{1:T}, \theta)$ can be targeted within the bootstrap filter by sampling from the observation distribution given the projected particles at each time step t as above (setting $\tilde{y}_t^{(i)} \sim P(Y_t|\tilde{x}_{1:t}^{(i)}, \theta)$). Defining $\tilde{y}_{t-L:t}^{(i)}$ as the aggregation of the newly sampled $\tilde{y}_t^{(i)}$ and previously stored $y_{t-L:t-1}^{(i)}$, and resampling $y_{t-L:t}^{(i)}$ from $\tilde{y}_{t-L:t}^{(i)}$ alongside $x_{t-L:t}^{(i)}$ in step 6 of the Algorithm 1 results in a set of particles $\{y_t^{(i)}\}$ that can be viewed as unweighted samples from the smoothing posterior predictive distribution. This can be clearly seen if we consider the predictive variable as part of an extended hidden-state space. To ensure consistency between the weights calculated in step 5 of Algorithm 1 and the predictive particles, it is important to use the same indices when resampling both $x_t^{(i)}$ and $y_t^{(i)}$. Computationally this is equivalent to sampling some indices from Multinomial $\left(\{1, 2, \ldots, N\}, \left\{\frac{w_t^{(j)}}{\sum_{k=1}^{N} w_t^{(k)}}\right\}_{j=1}^{N}\right)$, and then using these to resample both the hidden states and predictive variable. The parameter vector θ can be marginalised out by including and storing the relevant samples in Algorithm 3.

Missing data

Missing data can be handled in a variety of ways depending on the type of missingness (missing at random, missing not at random, etc. [307]). Where data are missing by design (for example, wastewater sampling data may only be collected on certain days [42]), the observation distribution becomes $P(y_t \text{ is missing}|X_{1:t}, y_{1:t-1}, \theta) = 1$, irrespective of the values of the hidden states. In practice, we also skip the resampling step in the bootstrap filter on such days, as all hidden states are equally likely. More complex models, where the relationship between hidden states and missingness is explicitly modelled, are also possible. Prediction of missing data can be made by sampling from the observation model, conditional on the value of the hidden states (another type of predictive posterior distribution).

Projections

The projection of future hidden states (such as future R_t) is handled similarly, effectively treating future data as missing by design, and projecting X_T forward by repeatedly sampling from the state-space transition distribution, generating projections of hidden states X_{T+k}. Observed data Y_{T+k} can then be projected made by sampling from the observation distribution conditional on the projected hidden states. We demonstrate this in the third example below.

Probability of elimination

Elimination of an infectious disease is generally defined as "the reduction to zero incidence of a certain pathogen in a given area" [308]. For the purposes of this primer, we say that elimination has occurred at time t if no new infections occur within $[t, t + 28]$, an appropriate window for the SARS-CoV-2 examples we consider, although a variety of other mathematical statements matching this definition are possible [118, 252]. This turns the problem of estimating the probability of elimination into a projection problem: we project the number of new infections in the next 28 days, and the proportion of particle trajectories that contain zero new infections provides an estimate of the probability of elimination. We demonstrate this in the second example below.

5.5.3 Model evaluation and selection

Model evaluation and selection are crucial components of any modelling exercise. We outline three metrics: root mean square error (RMSE) of the posterior predictive distribution, the coverage of posterior predictive credible intervals, and the continuous ranked probability score (CRPS) [209] of the posterior predictive distribution.

The RMSE of the posterior predictive distribution quantifies the average discrepancy between observed data and the model's predictions. It is calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}\ , \tag{5.6}$$

where $\hat{y}_t = E[Y_t|y_{1:T}]$ is the expected value of the posterior predictive distribution at time t. Smaller values indicate a better fit to the data.

The coverage of posterior predictive credible intervals measures the proportion of observed data that falls within the model's credible intervals. It is calculated as:

$$\text{Coverage}_\alpha = \frac{1}{T}\sum_{t=1}^{T}\mathbb{I}(y_t \in \text{CrI}_{\alpha,t}), \tag{5.7}$$

where $\text{CrI}_{\alpha,t}$ is the $(1-\alpha)$-level credible interval of the posterior predictive distribution at time t. A well-calibrated model yields credible intervals whose empirical coverage is close to the nominal level on average, for all values of α. For example, if $1-\alpha = 0.95$, then the posterior predictive credible intervals should contain the observed data approximately 95% of the time.

Finally, the CRPS of the posterior predictive distribution quantifies the average discrepancy between the predictive cumulative distribution function (CDF) and the empirical CDF. It is defined as:

$$\text{CRPS} = \frac{1}{T}\sum_{t=1}^{T}\left(\int_{-\infty}^{\infty}(P(Y_t \le y|y_{1:T}) - \mathbb{I}(y_t \le y))^2 dy\right). \tag{5.8}$$

Smaller values of CRPS indicate a posterior CDF that more closely matches the empirical CDF of the observed data.

In practice, we use a particle approximation to the expectation-definition of the CRPS [309]:

$$\begin{aligned}
\text{CRPS} &= \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{2}E|Y_t - Y_t'| - E|Y_t - y|\right] \quad \text{where} \quad Y_t, Y_t' \sim P(Y_t|y_{1:T}) \\
&\approx \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N}\sum_{j=1}^{N}|Y_t^{(j)} - y_t| - \frac{1}{2}\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k=1}^{N}|Y_t^{(j)} - Y_t^{(k)}|\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N}\sum_{j=1}^{N}|Y_t^{(j)} - y_t| - \frac{1}{N^2}\sum_{j=1}^{N}(2j - N - 1)\tilde{Y}^{(j)}\right].
\end{aligned} \tag{5.9}$$

where $Y_t^{(j)}$ are samples from the posterior predictive distribution at time t and $\tilde{Y}^{(j)}$ are the samples sorted in ascending order. Using sorted samples reduces the computational complexity of the CRPS from $\mathcal{O}(N^2)$ (line 2) to $\mathcal{O}(N \log N)$ (line 3) [309]. The CRPS is particularly useful in the model selection process, as it provides a principled trade-off between precision (how narrow the credible intervals are) and calibration.

5.6 Example: COVID-19 in Aotearoa New Zealand

We demonstrate these methods by fitting three models to national data from the COVID-19 pandemic in Aotearoa New Zealand [185], reported in Figure 5.5. Two time periods are

considered: 26 February 2020 - 4 June 2020 and 1 April 2024 - 9 July 2024. The former is characterised by a single epidemic wave with high-quality case reporting and a large proportion of imported cases, during which elimination of transmission was the primary public health policy aim. The latter is characterised by widespread transmission with a clear day-of-week effect and high levels of reporting noise and bias, with modelling primarily used to inform health-service resource allocation.

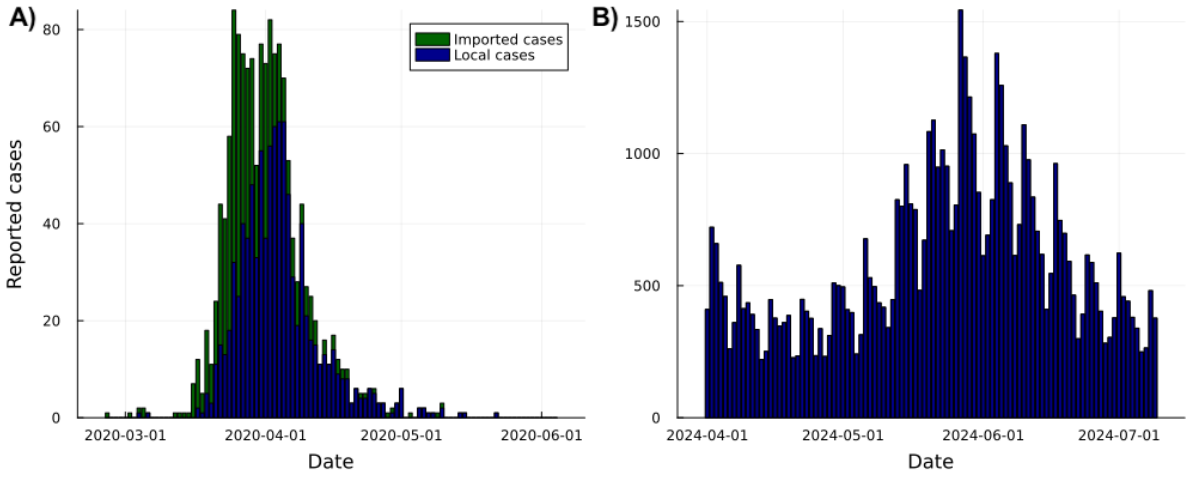

Figure 5.5: Reported cases from the COVID-19 pandemic in New Zealand, separated by locally-acquired and imported infections, for two time-periods: (A) 26 February 2020 - 4 June 2020 and (B) 1 April 2024 - 9 July 2024 [185]. The first period is the first wave of COVID-19 in New Zealand, characterised by a single epidemic wave which ended in temporary elimination of local transmission. The second period is characterised by widespread transmission with much higher numbers of reported cases.

5.6.1 Model 1: Simple example

As a first example, we implement a model very similar to EpiFilter [115], assuming that R_t follows a log-normal random walk (the state-space transition model):

$$\log R_t | \log R_{t-1} \sim \text{Normal}(\log R_{t-1}, \sigma), \tag{5.10}$$

and observed cases follow the Poisson renewal model (the observation model):

$$C_t | R_t, C_{1:t-1} \sim \text{Poisson}\left(R_t \sum_{u=1}^{u_{max}} C_{t-u}\omega_u\right). \tag{5.11}$$

In notation from the hidden-state models section, the hidden states are $X_t = R_t$, the observed data are $y_t = C_t$, and the parameter vector is $\theta = \sigma$. We could additionally consider the values

of the serial interval PMF $\{\omega_u\}_{u=1}^{u_{max}}$ as model parameters, but for simplicity we treat this is as fixed and part of the model structure, a typical assumption in such models.

The serial interval is assumed to follow a Gamma distribution with mean 6.5 days and standard deviation 4.2 days. For simplicity, this is discretised by evaluating the density at $t = 1, 2, \ldots, T$ days and normalising.

First we use PMMH (Algorithm 2) to estimate the posterior distribution of σ. Assuming a uniform prior distribution on $(0, 1)$, we find a posterior mean of 0.24 (95% Credible Interval (Cr.I.) 0.17, 0.33). This parameter governs the smoothness of R_t, though its value is not directly interpretable on its own. Convergence (in this case, $\hat{R} = 1.004$ and ESS = 171) was obtained in 200 iterations, taking approximately 10 seconds on a 2021 M1 MacBook Pro.

The focus of this example is on hidden-state inference: the estimation of R_t. We produce daily estimates of R_t and corresponding 95% credible intervals using Algorithm 3, effectively averaging over the posterior uncertainty about σ. These estimates are shown in panel A of Figure 5.6.

We also present the posterior predictive distribution of reported cases in Figure 5.6-B. We assess the calibration of the credible intervals for reported cases by comparing the observed data to the posterior predictive intervals and find that the model fits the data well, with 97.9% of all daily reported cases and 95.4% of non-zero daily reported cases in the dataset falling inside the daily 95% posterior predictive credible intervals.

By halting the SMC algorithm on 5 April 2020 and using $L = 30$ as the resampling lag, we can generate samples from the joint marginal distribution $P(R_{\text{7 March:5 April}}|C_{\text{26 Feb:5 April}})$, presented in Figure 5.6-C. These samples allow us to estimate the joint posterior distribution of the peak in transmission (the greatest value of R_t) and the timing of this peak. This is reported as a heatmap of the log-posterior density in Figure 5.6-D, where brighter colours represent more likely combinations of these values. The marginal posterior distributions of peak R_t and the date of this peak are shown in panels E and F of Figure 5.6. We estimate that R_t peaked at 6.8 (95% Cr.I. 4.9, 10.3) on 17 March 2020 (95% Cr.I. 16 March, 22 March). Estimates of peak R_t and the timing of this peak are useful when evaluating the effect of public health interventions, although we caution that reporting delays should be considered when interpreting these results.

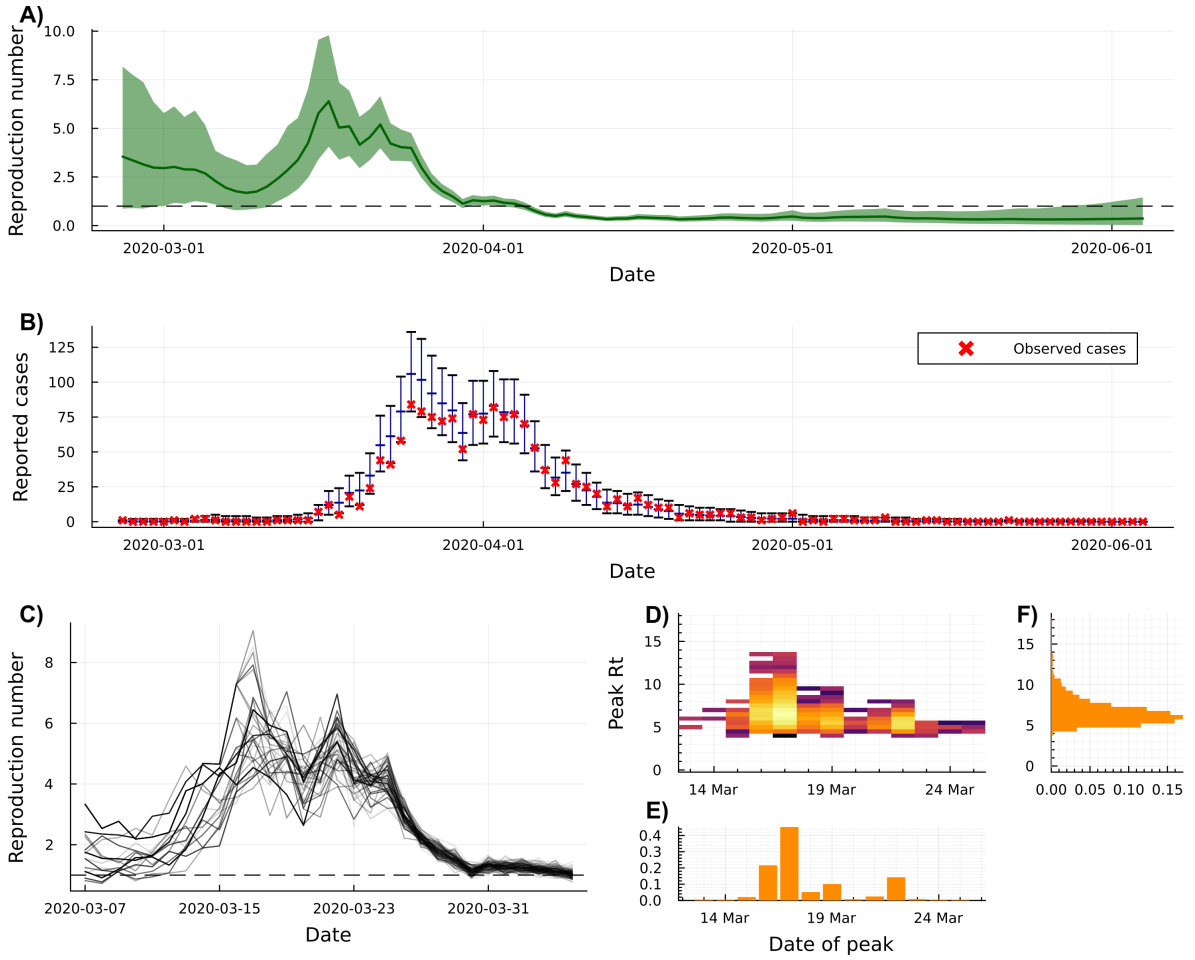

Figure 5.6: Results from fitting model 1 to data from the COVID-19 pandemic in New Zealand between 26 February 2020 and 4 June 2020. The marginal posterior means and 95% credible intervals of R_t (panel A) demonstrate high uncertainty in the early stages as well as a high peak estimate of R_t in late-March 2020. The marginal posterior predictive distributions of reported cases (panel B) shows that the model fits the data well, with 95.4% of non-zero reported daily cases in the dataset falling inside the 95% posterior predictive credible intervals. Panel (C) shows equally-likely samples from the joint posterior distribution of R_t between 7 March and 5 April 2020. These samples can be used to find the joint posterior distribution of peak R_t and the date of this peak, reported as a log-density heatmap (panel D), where brighter colours represent more likely combinations of these values. The marginal posterior distributions of peak R_t and the date of the peak are shown in panels E and F, respectively. Panels D and E suggest that the peak in R_t (the point at which the SARS-CoV-2 virus was spreading most rapidly) was most likely around 17 March 2020, although a second mode is visible with lower peak R_t on 22 March 2020.

Two distinct modes can be seen in the log-posterior density heatmap: one with a peak on 17 March 2020 and the other with a peak on 22 March 2020. The peak on 17 March is associated with a higher value of 7.3 (5.2, 10.6) (if we condition on this peak date), while a peak on 22 March 2020 is estimated to be lower, at 5.8 (4.8, 7.1). That is, there is evidence for either an

early and high peak, or a later and slightly lower peak in transmission.

5.6.2 Model 2: Reporting noise, imported cases, and elimination probabilities

Surveillance of COVID-19 in New Zealand was generally considered to be of high quality, although it can still be desirable to account for certain known biases. Firstly, infection incidence is not directly observed. Instead, we observe reported cases which are subject to reporting noise. Secondly, a large proportion of reported cases in the early phase of the pandemic in New Zealand were imported cases, those infected outside of New Zealand. Failure to account for these may lead the model to overestimate R_t, as imported cases are attributed to local transmission.

During this period, elimination of SARS-CoV-2 was an explicit public-health policy goal in New Zealand. Models explicitly estimating the probability of elimination were of key interest to policymakers [20], particularly as repeated days of zero cases were observed. We demonstrate how these methods can be used to estimate the probability of elimination, defined here as the probability that the true infection incidence I_t is zero for the next 28 days (i.e. $I_{t:t+28} = 0$), while simultaneously accounting for reporting noise and imported cases.

We introduce infection incidence as an additional hidden state, which is assumed to evolve according to the renewal model. The expected number of local infections I_t at time t is a function of R_t, the PMF of the generation time distribution ω_u, past local infections $I_{1:t-1}$, and past imported cases $M_{1:t-1}$:

$$\begin{aligned} \log R_t | \log R_{t-1} &\sim \text{Normal}(\log R_{t-1}, \sigma), \\ I_t | R_t, I_{1:t-1} &\sim \text{Poisson}\left(R_t \sum_{u=1}^{u_{max}} (I_{t-u} + M_{t-u})\omega_u\right). \end{aligned} \tag{5.12}$$

For simplicity, we also represent the generation time PMF by $\{\omega_u\}_{u=1}^{u_{max}}$ and use the same distribution as the last example. In practice, this should be changed to reflect that we are now modelling infection-to-infection rather than case reporting-to-case reporting.

We then assume that reported cases C_t follow a negative binomial distribution with mean I_t and variance $I_t + \phi I_t^2$:

$$C_t | I_{1:t-1} \sim \text{Negative Binomial}\left(r = \frac{1}{\phi}, p = \frac{1}{1 + \phi I_t}\right). \tag{5.13}$$

Parameter ϕ controls the level of observation noise, which we estimate alongside σ using PMMH. Like σ, we use a uniform prior distribution on $(0, 1)$ for ϕ.

The probability of elimination at time t is calculated by projecting the hidden states forward in time, conditioning on $M_{t:t+28} = 0$ (no imported cases), and calculating the proportion of particle trajectories that contain zero new local infections. This is repeated at each time step to produce a time series of the probability of elimination.

Using PMMH (Algorithm 2) we estimate a posterior mean and 95% credible interval of σ of 0.16 (95% Cr.I. 0.09, 0.25) and ϕ of 0.018 (95% Cr.I. 0.0005, 0.065). The decreased estimate of σ (compared to model 1) reflects the re-attribution of some noise in the data from the epidemic process to the observation process. The estimate of ϕ is small, suggesting that the observation process may be adequately modelled by a Poisson distribution, instead of the negative binomial distribution we have assumed (this could be tested using the model selection metrics outlined in Section 5.5.3). Convergence was obtained in 800 iterations, taking approximately 50 seconds on a 2021 M1 MacBook Pro.

Daily posterior means and 95% credible intervals of R_t are reported in Figure 5.7-A. Despite being fit to data from the same outbreak as model 1, the estimates of R_t are very different. By halting the algorithm on 5 April 2020, we estimate that R_t peaked at 1.6 (95% Cr.I. 1.2, 2.2) on 23 March 2020 (95% Cr.I. 6 March, 27 March), much lower and somewhat later than estimates from the simple model. This decrease is primarily a result of distinguishing locally-acquired from imported cases, although the re-attribution of noise from the epidemic process to the observation process also plays a role.

Figure 5.7-B demonstrates that the model fits the data well, although 100% of observed cases fall inside the 95% posterior predictive credible intervals, suggesting that the model is allowing for too much observation noise. Plotting the histogram of posterior samples of ϕ (Figure 5.7-D, from an extended implementation of PMMH with a minimum ESS of 1000) suggests that $\phi = 0$ (i.e. Poisson observation noise) is a plausible value. The uniform prior distribution places little mass around $\phi = 0$, which may be the cause of the overly-wide posterior predictive credible intervals.

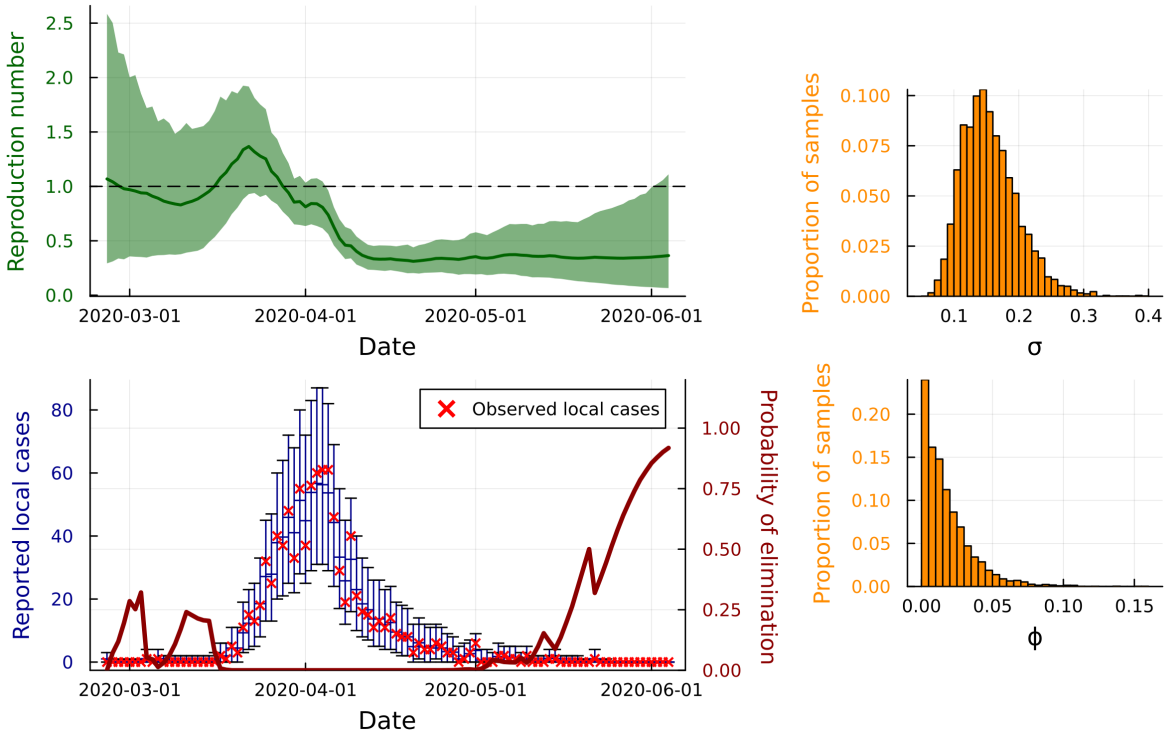

Figure 5.7: Results from fitting model 2 to data from the COVID-19 pandemic in New Zealand between 26 February 2020 and 4 June 2020. The marginal posterior means and 95% credible intervals of R_t (panel A) demonstrate a much lower and somewhat later peak in March 2020 than implied by model 1, reflecting the re-attribution of noise to the observation process and the distinction between locally-acquired and imported cases. The marginal posterior predictive distributions of reported cases (panel B) show that the model fits the observed data well, with observed data often lying close to posterior predictive means, although the model is overcovering the observed data. The probability of elimination increases steadily from early May and is estimated to reach 84.3% on the final time step 4 June 2020. The histogram of parameter samples (panels C and D) show a high posterior probability that ϕ is near 0, suggesting that the observation process may be adequately modelled by a Poisson distribution instead of the overdispersed negative binomial distribution.

The probability of elimination increases steadily from early May and is estimated to reach 91.8% on the final time step 4 June 2020. Conditioning on no new imported cases is equivalent to assuming that any imported infections cause no local transmission, a reasonable assumption given the strict border controls in place at the time.

5.6.3 Model 3: Reporting biases, temporal effects, and projections

For the final example, we consider reported cases of COVID-19 in New Zealand between 1 April 2024 and 9 July 2024, a period characterised by widespread transmission, decreased testing rates, and increased reporting delays and noise. A pronounced day-of-week effect is evident

in the data, with fewer cases reported on weekends. This can be accounted for by explicitly including a day-of-week effect or by modelling temporally aggregated data. We demonstrate both here.

We use a similar state-space transition model as the previous example, although ignore imported cases for simplicity:

$$\begin{aligned} \log R_t | \log R_{t-1} &\sim \text{Normal}(\log R_{t-1}, \sigma), \\ I_t | R_t, I_{1:t-1} &\sim \text{Poisson}\left(R_t \sum_{u=1}^{u_{max}} I_{t-u}\omega_u\right). \end{aligned} \tag{5.14}$$

To account for reporting delays, the expected number of cases μ_t at time t is modelled as a function of past incidence $I_{1:t-1}$ and an incubation period PMF d_u, assumed to be Gamma-distributed with mean 5.5 days and standard deviation 2.3 days (an incubation period distribution previously used for COVID-19 modelling in New Zealand[20], also discretised by evaluating at $t = 1, 2, \ldots$):

$$\mu_t = \sum_{u=1}^{u_{max}} I_{t-u} d_u. \tag{5.15}$$

The ability to handle a wide range of approaches to modelling reporting delays is a key strength of the SMC approach. We discuss this further online.

For the day-of-week model, we introduce seven new parameters: $c_i, i = 1, \ldots, 7$, representing the relative reporting rate on day i. These parameters are subject to the constraint that $\sum_{i=1}^{7} c_i = 7$, enforced by estimating $c_1, \ldots, c_6$ and setting $c_7 = 7 - \sum_{i=1}^{6} c_i$. If $c_i > 1$, then more cases are reported on day i than on average (and vice-versa). The observation distribution in this case is given by:

$$C_t | \mu_t, \phi \sim \text{Negative Binomial}\left(r = \frac{1}{\phi}, p = \frac{1}{1 + \phi c_{\text{mod}(t,7)+1} \mu_t}\right), \tag{5.16}$$

where $c_{\text{mod}(t,7)+1}$ is the day-of-week effect for day t. The $\text{mod}(t, 7)$ term is the modulo operator, which allows us to cycle through the days of the week.

For the temporally-aggregated model, we calculate particle weights and resample on a weekly basis. When $\text{mod}(t, 7) = 0$, we define $C'_t = \sum_{i=t-6}^{t} C_i$ as the non-overlapping weekly aggregated data, and the observation distribution is given by:

$$C'_t | \mu_{t-6:t}, \phi \sim \text{Negative Binomial}\left(r = \frac{1}{\phi}, p = \frac{1}{1 + \phi \sum_{i=t-6}^{t} \mu_i}\right), \quad \text{if mod}(t, 7) = 0. \tag{5.17}$$

We compare these models to a third "naive" model, which fits to daily cases while ignoring the day-of-week effect, using the following observation distribution:

$$C_t|\mu_t, \phi \sim \text{Negative Binomial}\left(r = \frac{1}{\phi}, p = \frac{1}{1+\phi\mu_t}\right). \tag{5.18}$$

Using PMMH (Algorithm 2), we find that all three models produce similar estimates of σ (Table 5.1). The day-of-week and temporally aggregated models produce comparable estimates of ϕ, while the estimated overdispersion in the naive model is much higher. This is expected, as the naive model attributes day-of-week noise to overdispersion in the observation process, while the other models explicitly account for this. Estimates of the day-of-week relative reporting rates from the day-of-week model are given in Figure 5.8-E, highlighting statistically significantly greater reporting rates on weekdays (particularly Monday) than on weekends.

Table 5.1: Parameter estimates from fitting the day-of-week model, the temporally aggregated model, and the naive model to data from the COVID-19 pandemic in New Zealand between 1 April 2024 and 9 July 2024. 95% credible intervals are shown in parentheses. Estimates of the day-of-week relative reporting rates are shown in Figure 5.8-E.

Model	σ	ϕ
Day-of-week	0.074 (0.043, 0.132)	0.011 (0.007, 0.015)
Temporally aggregated	0.070 (0.039, 0.132)	0.008 (0.001, 0.038)
Naive	0.065 (0.039, 0.116)	0.073 (0.053, 0.092)

Figure 5.8-A presents the daily posterior mean and 95% credible intervals of R_t for all three models. All three models produce similar estimates of R_t, although the day-of-week model exhibits slightly less uncertainty in some periods, reflecting the increase in information extracted from the data. This model is also able to capture the day-of-week effect while maintaining good statistical coverage, with 96.9% of observed cases falling inside the 95% posterior predictive credible intervals. The temporally aggregated and naive models both overcover the observed data, with 100% of observations falling inside the 95% predictive credible intervals, although there are only 14 observations in the temporally aggregated example.

Four-week projections are produced by iterating the estimated state-space model forward and sampling from the observation model. In all three models, 100% of future cases fall inside the 95% posterior predictive credible intervals, suggesting that all models overestimate the level of uncertainty (producing credible intervals that are too wide), most likely a result of model

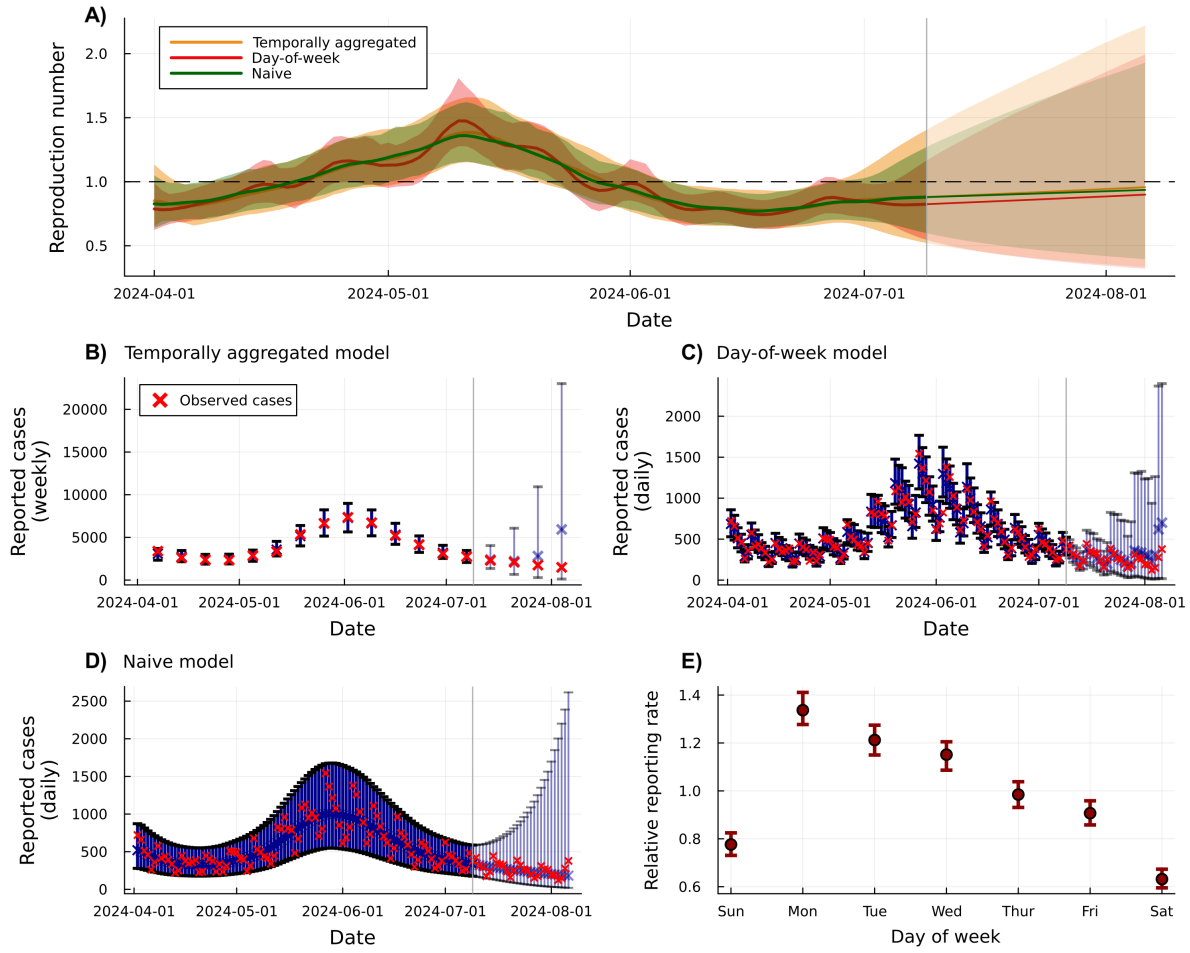

Figure 5.8: Results from fitting the three models of model 3 to data from the COVID-19 pandemic in New Zealand between 1 April 2024 and 9 July 2024. The dark red line and shaded region present the marginal posterior means and 95% credible intervals of R_t for the day-of-week model (red), the temporally aggregated model (orange), and naive model in green (panel A). Predictive weekly cases from the temporally aggregated model are shown in panel B, predictive daily cases from the day-of-week model are shown in panel C, and predictive daily cases from the naive model are shown in panel D. The day-of-week relative reporting rate estimates and 95% credible intervals are shown in panel E. Projections of R_t and reported cases are shown in panels (A-D), with lighter shading and the vertical grey bar marking the start of the projection.

misspecification. Allowing for more complex correlation structures in the dynamic model for R_t may reduce this uncertainty (for example, the assumption that R_t follows a random walk ignores potential mean-reverting behaviour) [105].

As we fit the naive model and the day-of-week model to the same data, we can compare the two using RMSE and CRPS. The RMSE of the naive model is 5.4 times greater than the day-of-week model on within-sample predictions and 255 times greater on the 4-week projections,

suggesting that the day-of-week model produces much more accurate central estimates. CRPS, which simultaneously accounts for calibration and precision, is 2.6 times greater for the naive model for within-sample predictions and 1.53 times greater on out-of-sample (four-week ahead) predictions. Despite its additional complexity and similar hidden-state estimates, the day-of-week model substantially outperforms the naive model.

5.7 Discussion

We have introduced a general framework for fitting discrete-time epidemic renewal models using SMC methods and demonstrated their application in estimating hidden states, parameters, and making projections. The primary strength of these methods lies in their flexibility: they can fit any state-space model, a form that encompasses many popular epidemiological models. This allows for rapid fitting, testing, and comparison of a wide range of models, facilitating the rapid development of solutions to emerging epidemiological challenges. All code and data to reproduce all results in this paper (and more) are provided online.

We demonstrate our methods on three examples. Each example consists of a distinct model structure, dataset, and aim, yet the same general framework is used to fit each model. The first example demonstrates the estimation of R_t using a simple renewal model, similar to EpiFilter [115], with the addition of joint estimation of the peak R_t value and the timing of this peak. The second example demonstrates the estimation of R_t while accounting for both reporting noise [310, 311] and imported cases [116, 312, 313], while simultaneously estimating the probability of elimination [118, 252], thus unifying the cited works. The final example demonstrates the estimation of R_t in the presence of reporting delays [239, 311, 314] and day-of-week effects [239, 315] or temporally aggregated data [45, 293, 294], as well as the ability to produce short-term projections [117], and the importance of model comparison and selection [36, 105], again unifying the cited (and many non-cited) works. Many other extensions are possible, particularly the ability to incorporate multiple data sources [24, 42].

Many modelling choices made in our examples are arbitrary: for example, a Poisson observation distribution could be used in the second example (instead of a negative binomial distribution), the day-of-week effect could be modelled as daily binomial reporting probabilities, and there are many ways to define elimination, to name just a few. Different researchers address the same

problems in different ways. The core strength of these methods is their ability to quickly and easily test these different models. We also provide principled methods for model evaluation and selection, including RMSE, CRPS, and coverage of posterior predictive credible intervals. These metrics are useful for comparing models fit to the same data, but less so when models are fit to different data sources, a potential area for future work.

Correctly accounting for uncertainty in model parameters has previously been shown to be critical for robust uncertainty quantification, and thus for robust policymaking [39, 235]. Our focus on marginalising out parameter uncertainty is a key strength of these methods, and is not included in many other epidemiological methods, including many SMC-based approaches [251, 252]. By coupling existing particle filter-only approaches with a PMMH algorithm like that presented here, an additional source of uncertainty (uncertainty in fixed parameters) can be accounted for, increasing the robustness of these methods.

Epidemic models range from highly mechanistic (such as compartmental SEIR-type models) to purely statistical (such as time series regression or exponential smoothers). The renewal model is semi-mechanistic in that it imposes a simple structure but does not model the entire underlying process. This leads to a model that is flexible, interpretable, and can produce well-calibrated short-term projections. However, like any mechanistic model, the renewal model and dynamic model on R_t jointly imply a specific autocorrelation structure in the data. If this structure is misspecified and the only goal is the projection of observable data, then the model may underperform compared to more flexible statistical models [105].

Our focus on simplicity and flexibility results in algorithms that are known to be less computationally efficient than more specialised alternatives. Areas for additional consideration include the use of a better-tuned proposal distribution (in Algorithm 1) [299], the use of more advanced resampling schemes (in Algorithm 1) [296], and the use of better-tuned PMMH proposal densities (in Algorithm 2) [289, 302]. Many solutions to these problems exist within the literature, often leveraging specific details about the chosen model, and we encourage the reader to seek out more efficient algorithms once a suitable model has been specified. In addition to epidemiology-specific literature [283, 284], we also recommend two texts from the broader SMC literature [120, 296] and a tutorial paper by Doucet and Johansen [121].

An effective class of alternatives to SMC-type methods for performing inference and generating projections with epidemic renewal models are probabilistic-programming-language-based Gaussian process approximations, such as EpiNow2 [139]. These methods offer a similar level of flexibility and, in some cases, can be faster than our approach. However, these methods are more complex mathematically, whereas ours are more accessible to those without a strong mathematical background and can be implemented in a few lines of code, making them easier to debug. Furthermore, our methods can easily be adapted for online updates, allowing for real-time analysis of epidemic data, whereas Gaussian-process-based models are typically offline methods. Finally, popular probabilistic programming languages, such as Stan, do not support discrete parameters. Many quantities of interest, such as infection incidence, are discrete, which our methods can handle natively, whereas other approaches require additional approximations.

Numerous methods for estimating R_t and conducting inference on epidemic data exist, often tailored to specific pathogens or datasets. More general models have also been created, although these can be challenging to implement or are confined to pre-built software packages. While these existing packages simplify the process, they can obscure the underlying structure of the model and associated assumptions. This primer aims to strike a balance: enabling researchers to quickly and easily construct their own models from scratch, ensuring they understand the assumptions and structure of the model, while facilitating rapid testing and comparison of different models. By providing a general framework for fitting discrete-time epidemic renewal models, we hope these methods contribute to the development of robust and useful epidemic models that can inform public health decision-making in future outbreaks across a range of pathogens.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

Title of Paper	A primer on inference and prediction with epidemic renewal models and sequential Monte Carlo
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Steyn, N., Parag, K. V., Thompson, R. N., & Donnelly, C. A. (2025). A primer on inference and prediction with epidemic renewal models and sequential Monte Carlo. *Statistics in Medicine*(44)18-19. https://doi.org/10.1002/sim.70204.

Student Confirmation

Student Name	Nicholas Steyn
Contribution to the Paper	Conceptualisation, methodology, software, visualisation, validation, writing (original draft), and writing (review and editing).

Signature NSteyn	Date 22/09/2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description above is accurate.

Supervisor name and title	Professor Christl A. Donnelly
Supervisor comments	

Signature Christl Ann Donnelly	Date 22/09/2025

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 6

Paper IV: A Bayesian model for repeated cross-sectional epidemic prevalence survey data

Status: Accepted at *PLOS Computational Biology*. A preprint is available on medRxiv here: https://doi.org/10.1101/2025.04.16.25325936. Minor revisions have been made since the preprint was posted.

Full author list: Nicholas Steyn, Marc Chadeau-Hyam, Paul Elliott, and Christl A. Donnelly.

Authorship contributions: N.S. conceived of the study, designed the methodology, wrote the software, created the visualisations, and wrote the original draft. M.C-H., P.E., and C.A.D. collected the data. All authors curated the data, validated the methods, and reviewed and edited the manuscript. C.A.D. supervised the project.

Supporting information: Supplementary material is provided in Appendix B. All source code and data (simulated and from the REACT-1 survey) used in this study are available at: https://github.com/nicsteyn2/EpidemicSurveySmoothing.

Abstract

Epidemic prevalence surveys monitor the spread of an infectious disease by regularly testing representative samples of a population for infection. State-of-the-art Bayesian approaches for analysing epidemic survey data were constructed independently and under pressure during the COVID-19 pandemic. In this paper, we compare two existing approaches (one leveraging Bayesian P-splines and the other approximate Gaussian processes) with a novel approach (leveraging a random walk and fit using sequential Monte Carlo) for smoothing and performing inference on epidemic survey data. We use our simpler approach to investigate the impact of survey design and underlying epidemic dynamics on the quality of estimates. We then incorporate these considerations into the existing approaches and compare all three on simulated data and on real-world data from the SARS-CoV-2 REACT-1 prevalence study in England. All three approaches, once appropriate considerations are made, produce similar estimates of infection prevalence; however, estimates of the growth rate and instantaneous reproduction number are more sensitive to underlying assumptions. Interactive notebooks applying all three approaches are also provided alongside recommendations on hyperparameter selection and other practical guidance, with some cases resulting in orders-of-magnitude faster runtime.

6.1 Introduction

The aims of infectious disease surveillance include describing disease burden, monitoring trends, and identifying outbreaks and novel pathogens through the collection, analysis, and interpretation of health data [143]. A range of passive systems, such as automated reporting from healthcare facilities, and active systems, such as field investigations, are used to gather these data. Recently, novel surveillance methods, such as wastewater testing [162], mobile phone data [316, 317], and social media monitoring [318], have been developed to complement traditional surveillance methods. While these methods provide an unprecedented volume of data, they are often subject to biases and limitations that make reliable inference and interpretation difficult [319].

Large-scale prevalence surveys are another tool for the surveillance of infectious diseases. These surveys typically use random sampling methods to produce estimates of the prevalence of infection in a given population. The quality of data collected also allows for the robust estimation of epidemiological quantities such as prevalence P_t, the growth rate r_t, and the instantaneous reproduction number R_t [320]. Established during the COVID-19 pandemic, the REal-time Assessment of Community Transmission (REACT-1) study in England [48] processed over 2.5 million self-administered throat and nose samples between May 2020 and March 2022. The ONS Coronavirus Infection Survey [150] monitored the spread of SARS-CoV-2 in the United Kingdom, also processing millions of samples over the course of the pandemic. These surveys have been instrumental in understanding, for example, the spread of SARS-CoV-2 [48], the dynamics of infection hospitalisation and infection fatality ratios [321, 322], and the impact of vaccination [151].

Implementing such large-scale surveys is expensive; for example, the ONS Coronavirus Infection Survey cost £988.5 million (until September 2023) [323]. It is therefore critical to maximise the information extracted from the data - not only to justify cost, but also to support timely decision-making and improve policy relevance. This is made challenging by the substantial noise inherent in point prevalence estimates. Even with an ideal survey design, large sample size, and only ignorable non-response, the coefficient of variation of common binomial proportion estimators can be large [324], particularly when infection prevalence is low.

Even when the goal is to perform "model-free" inference (making inferences that reflect only the data, rather than modelling assumptions), noise in these data often necessitates the use of smoothing methods to improve the quality of the outputs. Smoothing methods impose constraints on day-to-day variability, allowing data from multiple days to inform point estimates. Smoothing allows the researcher to make more confident statements, increasing information yield or reducing the sample size needed for a given confidence level. However, all smoothing methods make assumptions about the underlying process, whether explicitly or implicitly. Even seemingly simple smoothing methods can introduce substantial bias, turning "model-light" estimates (that rely only minimally on modelling assumptions) into "model-heavy" estimates (that are strongly shaped by modelling assumptions) that may not accurately reflect reality.

Here we introduce a novel Bayesian approach for smoothing and performing inference on epidemic survey data, referred to as SIMPLE (Survey Inference Method for Prevalence and other Latent variables in Epidemiology). This approach is based on hidden-state models and sequential Monte Carlo (SMC) methods [40, 120], and is designed to be flexible enough to incorporate key assumptions of existing approaches, while avoiding some of the computational and mathematical complexity of these approaches. We compare the SIMPLE approach to two existing approaches, one by Eales et al. [241] that leverages Bayesian P-splines and another by Abbott and Funk [320] that uses Gaussian processes approximations. We refer to these approaches as the Eales approach and the Abbott approach respectively. All three approaches are presented using common and general notation, allowing for direct comparison of their assumptions, performance, and results.

The results of this paper are structured in three parts. First, we use the SIMPLE approach to investigate the impact of survey design (such as the number of samples and individual response bias) and underlying epidemic dynamics (such as the variability in the growth rate) on the quality of estimates, highlighting key modelling decisions that should be made when analysing epidemic survey data. Second, we demonstrate how the approaches of Eales and Abbott can be adapted to account for these considerations, and compare all three (improved) approaches on simulated data to highlight similarities and differences in their performance. The use of simulated data provides a ground truth that allows us to calculate and compare quantities such as statistical coverage (the percentage of times that credible intervals of a given

level contain the true value). Third, we apply all three approaches to real-world data from the REACT-1 study, covering the COVID-19 pandemic between May 2020 and March 2022 in England, comparing the quality of estimates and computational requirements of each approach. We provide recommendations for future modelling and survey design based on our findings.

Our goal is to provide a framework for understanding the impact of modelling decisions on the quality of estimates, to consolidate and improve state-of-the-art methods, and to make recommendations for future modelling and survey design. We also provide documented source code for all three approaches, notebooks demonstrating their use, and recommendations for hyperparameter selection and other practical considerations.

6.2 Materials and methods

6.2.1 The SIMPLE approach

We introduce a suite of state-space models for simultaneously smoothing and performing inference on epidemic survey data. Each state-space model consists of an explicitly defined epidemic and observation model. The epidemic model encodes our assumptions about the unobserved dynamics of the epidemic, while the observation model describes how the observed data are generated from the underlying epidemic.

Two epidemic models are considered: one for estimating the daily exponential growth rate r_t in prevalence P_t and one for estimating the instantaneous reproduction number R_t, infection incidence I_t and prevalence P_t. Three observation models are considered: a basic model that assumes the number of positive swabs (diagnostic tests) follows a binomial distribution (as used by the original Eales approach), an extra-binomial model that accounts for overdispersion in the data, and a weighted model that accounts for survey weights.

We refer to the unknown time-varying quantities of interest, such as r_t and P_t, as hidden states. Other unknown parameters that are not time-varying, such as the level of overdispersion in the extra-binomial model, are referred to as static parameters.

Throughout this paper, we use the term *prevalence* to describe the proportion of the population that would test positive. This quantity may differ from the proportion of the population with

an active infection (due to imperfect test sensitivity/specificity) and from the proportion who are infectious (as test positivity can outlast infectiousness).

Growth rate epidemic model

This model assumes that the daily growth rate r_t in prevalence follows a Gaussian random walk, encoding an assumption that r_t varies smoothly over time. Prevalence P_t on day $t \in \{1, 2, \ldots, T\}$ then varies exponentially with rate r_t.

$$r_t = r_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma^2), \tag{6.1}$$

$$P_t = P_{t-1} e^{r_t}. \tag{6.2}$$

Coupled with initial distributions for r_0 and P_0, and a prior distribution for σ, equations Equation 6.1 and Equation 6.2 define the entire epidemic model. Parameter σ controls the smoothness of our estimates and is inferred from the data. As this model makes minimal assumptions about the underlying epidemic, we use it as the default epidemic model unless otherwise stated.

We use default initial distributions of $r_0 \sim \mathrm{Uniform}(-0.3, 0.3)$ and $P_0 \sim \mathrm{N}(\hat{p_1}, \hat{p_1}(1-\hat{p_1})/n_1)$, where $\hat{p_1} = n_1^+/n_1$ (the Wald interval), and default prior distribution $\sigma \sim \mathrm{Uniform}(0, 0.2)$, although inferences are insensitive to the choice of these prior distributions (Supplementary Section B.1).

Taking a function space view, modelling the growth rate r_t as a first-order random walk implies a log-prior probability proportional to $-\frac{1}{2\sigma^2}\sum(r_t - r_{t-1})^2$. This can be viewed as a discrete approximation of the integral penalty $\int \left(\frac{d}{dt} r_t\right)^2 dt$, thus implying that the equivalent continuous growth rate function lies in the $W^{1,2}$ Sobolev space (the space of functions in L^2 that have square-integrable weak first derivatives). Although these are locally smooth functions, they do not allow for abrupt changes in r_t which may arise from sudden changes in behaviour or policy [240].

Reproduction number epidemic model

Alternatively, we may want to estimate the instantaneous reproduction number R_t, a popular alternative measure to r_t for characterising the rate of epidemic spread [138]. R_t is the average number of secondary infections generated by a primary infection at time t if an individual were to undergo their entire infectious period at this time. As with r_t, we employ a Gaussian random walk to smooth estimates, now on $\log R_t$ to ensure positivity:

$$\log R_t = \log R_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, \sigma_R^2). \tag{6.3}$$

We then employ the renewal model [40, 114], which relates past infection incidence $I_{1:t-1}$ to current infection incidence I_t through the instantaneous reproduction number R_t and a generation time distribution (representing the time from infection of an infector and their infectee, described by probability mass function w_u):

$$I_t = R_t \sum_{u=1}^{t} I_{t-u} w_u. \tag{6.4}$$

Finally, we relate prevalence P_t to infection incidence I_t through a test-sensitivity function h_u that describes how likely an individual is to test positive u days after infection. We do not consider imperfect test specificity, although this could be included with the addition of a constant term to P_t:

$$P_t = \sum_{u=1}^{t} I_{t-u} h_u. \tag{6.5}$$

When applying the reproduction number epidemic model to data from the REACT-1 study, the generation time is assumed to follow a gamma distribution with mean 6.4 days and standard deviation 4.2 days [320], chosen to be consistent with the generation time distribution used in the published Abbott approach. For simplicity, we discretised this distribution by evaluating the gamma density function at integer time steps and normalising so $\sum w_u = 1$, although other discretisation methods can be more accurate [293, 325]. There is also evidence that the generation time of SARS-CoV-2 has shortened with more recent variants [326], and this should be kept in mind when interpreting our real-world results. In the absence of a REACT-1-specific test-sensitivity function, we use the pointwise central estimate of h_u from [327], reflecting the type of test used in REACT-1 (reverse transcription polymerase chain reaction (RT-PCR)),

but not other study-specific factors that may impact sensitivity. This is the same mean test-sensitivity function as used in the Abbott approach.

Estimating R_t requires many additional assumptions about the underlying epidemic, and thus has more potential to bias estimates of P_t. We focus on the growth rate epidemic model for much of this paper, although demonstrate the R_t estimator on real-world REACT-1 data.

Basic observation model

Given n_t swabs conducted on day t, the observed number of positive swabs n_t^+ is modelled as a binomial random variable with probability P_t:

$$n_t^+ \sim \text{Binomial}(n_t, P_t). \tag{6.6}$$

This is also the observation model used in the Eales approach as originally published.

Extra-binomial observation model

Prevalence data are often overdispersed relative to the binomial distribution, exhibiting what is known as extra-binomial variation [328]. If we assume that the "observable" prevalence at time t is beta-distributed with mean equal to prevalence P_t and variance $\rho P_t(1-P_t)$, our observation distribution becomes:

$$n_t^+ \sim \text{Beta-binomial}(n_t, \alpha_t, \beta_t), \tag{6.7}$$

where $\alpha_t = P_t\left(\frac{1}{\rho} - 1\right)$ and $\beta_t = (1 - P_t)\left(\frac{1}{\rho} - 1\right)$.

The additional parameter $\rho \in (0, 1)$ controls the modelled level of overdispersion in the data. Larger values of ρ indicate greater levels of overdispersion while letting $\rho \to 0$ recovers the binomial model. By default, we use a Uniform$(0, 0.01)$ prior distribution for this parameter. For context, the greatest upper bound of any 95% credible interval for ρ estimated from a real-world dataset is 0.0006.

This is our default observation model and is used in all subsequent analyses unless otherwise stated. It is common in other epidemiological settings to model overdispersed count data using the negative-binomial distribution [329], which is useful when the data have no natural upper limit. However, as the number of positive swabs are bounded above by the total number of swabs, we do not consider this here.

Weighted observation model

Survey weights are used to account for bias in survey data arising from unequal probabilities of selection and/or response. Letting $w_{t,i}$ be the (normalised) weight assigned to the i^{th} sample taken at time t, and $x_{t,i}$ be the corresponding result (where $x_{t,i} = 1$ is individual i on day t tests positive, and zero otherwise), the weighted swab positivity is:

$$\hat{p}_t = \sum_{i=1}^{n_t} x_{t,i} w_{t,i} \ . \tag{6.8}$$

If the weights are uncorrelated with swab positivity, then $Var(\hat{p}_t) = P_t(1 - P_t) \sum_{i=1}^{n_t} w_i^2$. However, if the weights are correlated with the individual probability of testing positive, the variance of $\hat{p}_t$ may be over or underestimated by this expression. The difference depends on the specific relationship between weights and outcome, which is generally unknown. Thus, we model $\hat{p}_t$ using a normal distribution with mean P_t and variance $cP_t(1 - P_t) \sum_{i=1}^{n_t} w_{t,i}^2$, where c is an estimated scalar parameter:

$$\hat{p}_t \sim \mathrm{N}\left(P_t, cP_t(1 - P_t) \sum_{i=1}^{n_t} w_{t,i}^2\right). \tag{6.9}$$

The observation distribution is no longer exact, in that we are approximating the unknown distribution of $\hat{p}_t$ with a normal distribution. This is similar to the approximation used in the Abbott approach. By default, we use a Uniform$(0.1, 10)$ prior distribution for c.

Sequential Monte Carlo (SMC)

We use an SMC algorithm (also known as a particle filter), the bootstrap filter, to estimate the posterior distributions of the hidden states (such as r_t, P_t, and R_t) at each time step t, given the observed data [40]. We also use this algorithm to estimate the posterior predictive distribution of n_t^+, which is similar to the frequentist's predictive distribution, and can be used to assess the quality of fit of the model when fitting to real-world data. Particle marginal Metropolis-Hastings (PMMH) is used to estimate the static parameters, the uncertainty of which is then marginalised over in the final inference.

We run three chains of the PMMH algorithm, each for 100 iterations at a time, until the maximum Gelman-Rubin diagnostic ($\hat{R}$) [303] is less than 1.05 and the minimum effective sample size (ESS) is greater than 100. These algorithms are detailed in full in Supplementary

Section B.2. We make the full source code (written natively in Julia [330]) available on GitHub, along with notebooks, to facilitate the application of these approaches to other datasets.

When fitting the model for the reproduction number, an additional "wind-in" period (default 10 days) is required to account for infection history prior to the first observation. After sampling from the standard prior distribution for P_0, assumed values of $I_{-9:0}$ are set to $P_0/\sum h_u$, ensuring the distribution of implied prevalence at time 0 is consistent with the chosen prior distribution. As 10 days may truncate the generation time distribution and test-sensitivity function, we rescale the generation time distribution to sum to 1 and the test-sensitivity function to sum to its original total over the combined length of past data and the wind-in period.

6.2.2 The Eales approach

The Eales approach [241] models logit-transformed P_t using Bayesian p-splines:

$$\text{logit}P_t = \sum_{i=1}^{N} b_i B_{i,n}(t), \quad \text{logit}(x) = \log\frac{x}{1-x}, \tag{6.10}$$

where $B_{i,n}(t)$ are basis functions and b_i are estimated spline coefficients. The coefficients give the value of the spline at the corresponding "knots" t_i while the basis functions allow for interpolation. By default, fourth-order basis functions $B_{i,4}(t)$ are used. We refer the reader to the original paper for the full construction [241].

The smoothness of these splines is controlled by the spacing of the spline knots (set by the user), and a second-order Gaussian random walk prior distribution on the spline coefficients, with standard deviation σ_{Eales} (estimated from the data):

$$b_i - b_{i-1} = b_{i-1} - b_{i-2} + u_i, \quad u_i \sim \text{N}(0, \sigma^2_{Eales}). \tag{6.11}$$

As $\text{logit}P_t - \text{logit}P_{t-1} \approx \log P_t - \log P_{t-1} = r_t$ for small P_t, this is approximately equivalent to modelling the growth rate using a Gaussian random walk, a similar smoothing assumption to our growth rate epidemic model. This approximation can be improved by noting that $\text{logit}P_t - \text{logit}P_{t-1} \approx \log P_t - \log P_{t-1} + (P_t - P_{t-1})$ and that the smoothness of P_t implies that $P_t - P_{t-1}$ is small. Taking a function space view, by directly penalising the second-order differences of the spline coefficients, the Eales approach places logit-prevalence in the $W^{2,2}$

Sobolev space. Since the (continuous) growth rate is approximately the first derivative of logit-prevalence, the implied growth rate function lies in the $W^{1,2}$ Sobolev space, the same space implied by the SIMPLE approach.

By default, spline knots are placed approximately every 5 days (exactly 5 days when the duration of the data is divisible by 5) to balance flexibility with computational efficiency. As we generally work in integer time, there is no benefit to knot spacing shorter than 1 day. If knots are placed on each day, we only need to evaluate the splines at the knot locations, and thus can work with b_i directly (i.e., we do not need to employ any splines). In this case, except for modelling the change in $\text{logit} P_t$ instead of $\log P_t$, the model is equivalent to the SIMPLE (extra-binomial) model. We examine this equivalence further in Supplementary Section B.3.

The original Eales approach modelled positive swabs n_t^+ with a binomial distribution, equivalent to our basic observation model. We update this to use a beta-binomial distribution with overdispersion parameter ρ, equivalent to our extra-binomial observation model, improving both convergence of the algorithm and the quality-of-fit of the model (Supplementary Sections B.3 and B.4). We do not provide a weighted implementation of the Eales approach, although this could be achieved by using the weighted observation model described above.

Growth rates (in prevalence) are back-calculated from the fitted splines by setting $r_t = \log(P_t/P_{t-1})$, where $P_{1:T}$ are sampled from the posterior distribution of the splines. A weakly informative inverse-gamma prior distribution ($\alpha = \beta = 0.0001$) is used for σ_{Eales}, a uniform prior distribution is used for ρ, and a uniform prior distribution is used for the first two spline coefficients.

Rather than explicitly model incidence, the Eales approach makes a series of simplifying assumptions when estimating the reproduction number. Specifically, it assumes that at each independent time step t, R_t has been fixed for the past τ days (a trailing-window approach akin to that used in EpiEstim [114]). If τ is chosen to be larger than both the maximum generation time and maximum duration of swab positivity, then the renewal equation (Equation 6.4) can be used to estimate R_t directly. In practice, $\tau = 14$ days is used to prevent over-smoothing of R_t, despite this likely being less than the maximum duration of test sensitivity

for SARS-CoV-2 (which, according to [327], remains above 10% even after 21 days after infection). The trailing-window approach is known to produce biased estimates of R_t [39], which can be partially accounted for by reporting estimates shifted by $\tau/2$ days [36].

Eales et al. [241] provide source code to fit this model using the R programming language [331] and Rstan. We adapt their code to fit the model using the actively developed interface to Stan, cmdstanr [332]. As with Rstan, inference is performed using Hamiltonian Monte Carlo (HMC) with the adaptive No-U-Turn Sampler (NUTS). We also adjust the hyperparameters of their algorithm (increasing the maximum tree-depth to 15 and decreasing the number of warm-up/sampling iterations to 200/300) to substantially improve convergence times, from at least 30 hours on the full REACT-1 dataset to less than 1 hour (Supplementary Section B.4).

As in the SIMPLE approach, convergence is measured by running three chains and checking that maximum $\hat{R} < 1.05$ and minimum $ESS > 100$. In this case, "hidden states" such as the growth rate are estimated alongside static parameters, so also feature in the convergence checks. This means the maximum $\hat{R}$ and minimum ESS diagnostics are not directly comparable to those of the SIMPLE approach, although they do reflect the computationally expensive aspects of each approach (estimating the static parameters in the SIMPLE approach is much more expensive than estimating the hidden states).

6.2.3 The Abbott approach

The Abbott approach [320] models the first-order difference (other order differences are possible but not considered here) of logit-transformed daily infection incidence using a Laplace eigenfunction approximation [122, 333] to a zero-mean Gaussian process:

$$\text{logit}(I_t) = \text{logit}(I_0) + \sum_{s=1}^{t} GP_s, \tag{6.12}$$

where I_0 is the initial incidence and GP_s is the value of the approximated Gaussian process at time s. At small incidence values I_t, this is approximately equivalent to modelling the growth rate in incidence r_t^{inc} as a Gaussian process:

$$r_t^{inc} = \log I_t - \log I_{t-1} \approx \text{logit} I_t - \text{logit} I_{t-1} = GP_t. \tag{6.13}$$

A squared exponential kernel, parameterised by variance σ^2_{Abbott} and lengthscale ℓ, is used for the Gaussian process:

$$k(GP_t, GP_s) = \sigma^2_{Abbott} \exp\left(-\frac{(t-s)^2}{2\ell^2}\right). \tag{6.14}$$

If the lengthscale ℓ is large relative to unit time steps, this Gaussian process can be locally approximated by an AR(1) process: $GP_{t+1} = \phi GP_t + \epsilon_t$ where $\phi = \exp(-1/2\ell^2)$ and $\epsilon_t \sim N(0, \sigma^2_{Abbott}(1-\phi^2))$. This approximation highlights a mild similarity to our basic model: the model used in the Abbott approach can be (very loosely) viewed as a Gaussian random walk, except on r_t^{inc} and with a drift towards zero growth. The strength of this drift is controlled by ℓ with larger values of ℓ indicating a slower drift towards zero growth. From a function space perspective, the sample paths of a Gaussian process with a squared exponential kernel are almost surely infinitely differentiable [334]. Because prevalence is a convolution of incidence, the implied prevalence in this approach inherits the same very high degree of smoothness. While outside the scope of this paper, alternative Gaussian process kernels could be used in the Abbott approach to allow for only local smoothness (the Matérn kernel is a popular choice with direct links to the function spaces of the SIMPLE and Eales approaches [334]).

Prevalence P_t is modelled as a convolution of past incidence with a test-sensitivity function that describes how likely an individual is to test positive u days after infection (Equation 6.5). During the COVID-19 pandemic, estimates of the test-sensitivity function produced by Hellewell et al. [327] were used. We use the same estimates in our implementation (including in simulated data where applicable), although these curves depend on the pathogen (including variant, in the case of SARS-CoV-2 [335]), testing procedures (e.g., swabbing technique [46] and storage during transport [151]), and the population being tested (e.g., age structure and vaccination status [336]). As a result, these curves are not universal, and should be separately estimated for each scenario. We compare multiple estimates of these curves, and the sensitivity of model results to these curves, in Supplementary Section B.5.

The original Abbott approach used a normal likelihood for observed swab positivity with mean P_t and an estimated variance, a necessary simplification given the format of their data. We employ the beta-binomial observation model instead, although the binomial or weighted observation models are also possible and can be implemented by changing only a few lines of

the Stan code. The original model also leveraged antibody data, accounting for vaccination status, in their model. We focus on swab positivity data so do not consider this here. Their antibody model could be included in any of the three approaches discussed in this paper. We leave this for future work.

Growth rates are back-calculated from the estimated prevalence by setting $r_t = \log(P_t/P_{t-1})$. A weakly informative inverse-gamma prior distribution is used for ℓ, a zero-mean normal distribution with standard deviation 0.1 is used as the prior distribution for σ_{Abbott}, a uniform prior distribution is used for ρ, a normal prior distribution is used for the initial logit-incidence (at the beginning of the 50-day wind-in period) with mean -4.6 and standard deviation 2, and a zero-mean normal prior distribution with standard deviation 0.25 is used for the initial growth rate in incidence (at the beginning of the 50-day wind-in period).

As the Abbott approach models infection incidence directly, the reproduction number can be estimated by directly applying the renewal model (Equation 6.4) to the estimated incidence curves.

Abbott and Funk provide source code in R and cmdstanr to fit their model. When fitting their model, we use 200 warm-up and 300 sampling iterations. As in the Eales approach, convergence is measured by running three chains and checking that maximum$\hat{R} < 1.05$ and minimum$ESS > 100$. This source code assumes data of a different format to ours (their inputs are daily means and credible intervals from a different Bayesian model), so we adapt their code to fit to raw data. We also remove code relating to their antibody model. The modified code is provided on GitHub.

6.2.4 Simulated data

Simulated datasets, approximately reflecting the dynamics of COVID-19 in England, are generated by sampling from the prior predictive distribution of a given model with predetermined parameter values. In general, T days (default 100) of an epidemic are simulated with a fixed initial prevalence P_0 (default 1%) and growth rate r_0 (default 0). As the Abbott approach models incidence and requires a wind-in period (default 50 days when simulating), we instead fix infection incidence at the start of the wind-in period at 0.1%. Other model-specific parameters are set to default values as outlined in Table 6.1 unless otherwise stated. These are chosen

to reflect estimates from the REACT-1 study, thus representing plausible values for SARS-CoV-2-like epidemics. To prevent unrealistic scenarios, we bound simulated prevalence in the range $(0.1\%, 30\%)$ by resampling if a simulation exceeds these bounds. Example simulations are shown in Figure 6.1.

Table 6.1: Default parameter values used when simulating data, unless otherwise stated. These values are chosen to reflect estimates from fitting each model to the REACT-1 study. Prevalence and incidence are as a proportion of the population.

Parameter	**Default value**	**Description**
General parameters		
T	100 days	Duration of the simulated epidemic
P_0	0.01	Initial prevalence
r_0	0	Initial growth rate (per day)
$n_t, \quad t \in \{1, \ldots, T\}$	5,000	Daily swabs
SIMPLE approach (growth rate epidemic model)		
σ	0.01	Standard deviation of random walk on r_t (per day)
ρ	2×10^{-4}	Overdispersion parameter (beta-binomial observation distribution)
ξ	0.5	Scale of log-normal distribution for simulated weights
Eales approach		
σ_{Eales}	0.1	Standard deviation of random walk on spline knots (per knot)
ρ	2×10^{-4}	Overdispersion parameter (beta-binomial observation distribution)
Abbott approach		
σ_{Abbott}	0.08	Gaussian process kernel amplitude
ℓ	10	Gaussian process kernel lengthscale
ρ	2×10^{-4}	Overdispersion parameter (beta-binomial observation distribution)
I_{-init}	0.001	Initial incidence at the start of the wind-in period
r_{-init}	0	Initial growth rate in incidence (per day) at the start of the wind-in period

Weighted survey data are generated by sampling artificial survey weights $w_{t,i}$ from a log-normal distribution (default log-mean 0 and scale $\xi = 0.5$) and normalising so $\sum_{i=1}^{n_t} w_{t,i} = 1$. To introduce bias, the probability that individual $i \in \{1, ..., n_t\}$ tests positive is set to:

$$P_{t,i} = P_t \frac{w_{t,i}}{\sum_{j=1}^{n_t} w_{t,j}^2}. \tag{6.15}$$

This results in individuals with higher survey weights (i.e., individuals that are underrepresented in the sample) being more likely to test positive, while the weighted average of $P_{t,i}$ is equal to P_t. The simulated dataset then consists of $\{x_{t,i}, w_{t,i}\}$ where each $x_{t,i}$ is a realisation of a Bernoulli($P_{t,i}$) random variable. This imposes a linear relationship between survey weight and individual swab positivity, so represents a high level of bias for any assumed scale ξ. Default $\xi = 0.5$ was chosen to reflect the observed distribution of per-round survey weights in the

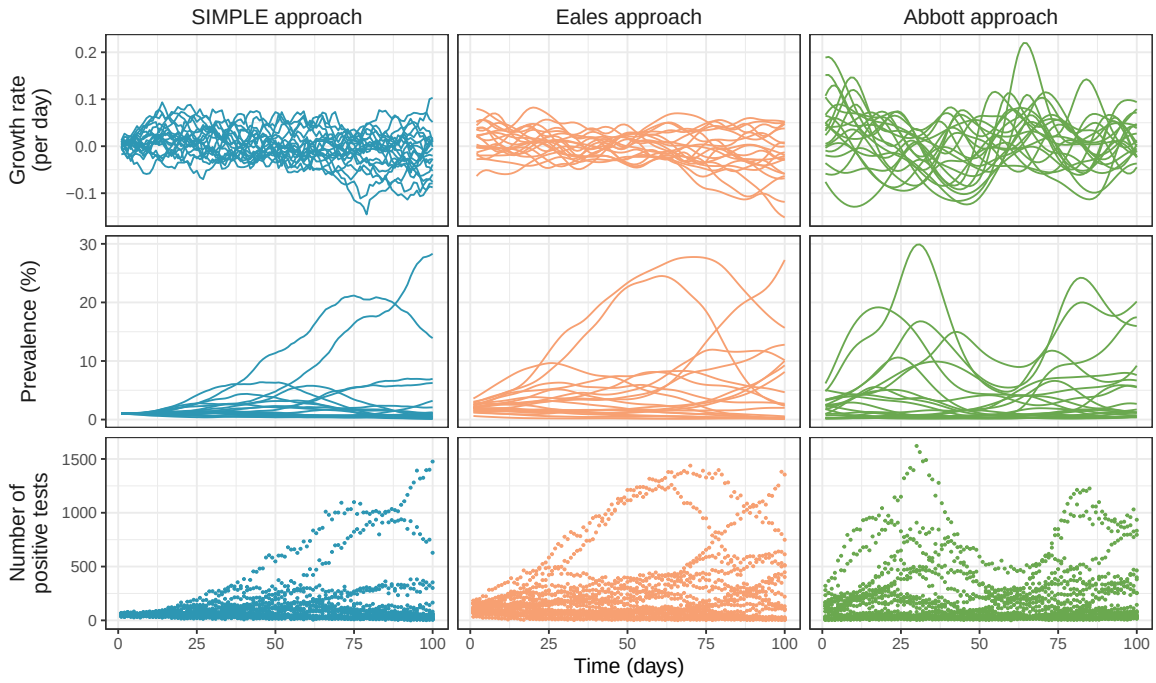

Figure 6.1: Simulated epidemic trajectories from the SIMPLE, Eales, and Abbott approaches. Default parameter values as described in Table 6.1 were used for these simulations. A total of 20 simulations are shown per approach, reflecting the range of possible outcomes. The top row shows the simulated growth rate r_t, the middle row shows the simulated prevalence P_t, and the bottom row shows the simulated number of positive swabs n_t^+.

REACT-1 study. An empirical analysis of these weights is provided in Supplementary Section B.6.

To consider the effect of survey design and epidemic dynamics, we fit each of the SIMPLE observation models (basic, extra-binomial, and weighted) to simulated datasets of duration $T = 100$ days from each observation model (with the growth rate epidemic model). Five values of the daily sample size n_t are considered (10, 100, 1,000, 10,000, 100,000) and two values of σ (0.008, 0.016). When simulating from the extra-binomial model we assume $\rho = 2 \times 10^{-4}$, and when simulating from the weighted model we assume $\xi = 0.5$. The choice of these values were guided by estimates from the REACT-1 study (see the REACT-1 results for ρ and Supplementary Section B.6 for ξ). Each simulation is repeated 10 times to average over stochasticity. To evaluate model performance, we consider the coverage and average width of the 95% credible intervals for estimated prevalence P_t. Equivalent results for the growth rate in swab positivity r_t are presented in Supplementary Section B.7.1.

6.2.5 The REACT-1 study

The REACT-1 study was an infection prevalence survey that tested for SARS-CoV-2 infection in England between May 2020 and March 2022 [48]. Conducted over 19 rounds, a total of 2.5 million self-administered throat and nose samples were processed using RT-PCR. Daily swab positivity and sample sizes for all 19 rounds of the REACT-1 study are reported in Figure 6.2. An average of 6,236 samples were taken on each of the 400 days that sampling occurred, or an average of 3,564 samples per day over the 700 days spanned by the study (from 1 May 2020 to 31 March 2022 inclusive).

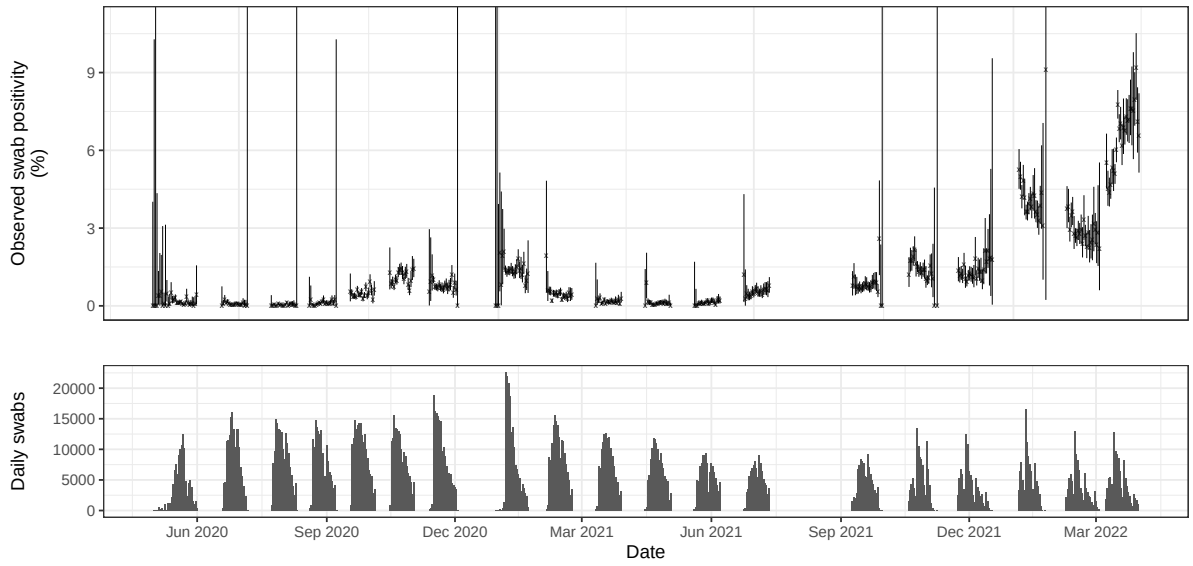

Figure 6.2: Daily SARS-CoV-2 swab positivity in England from the REACT-1 survey (upper) and corresponding daily sample sizes (lower). Daily 95% confidence intervals (vertical lines) for prevalence were calculated using the Agresti-Coull method [324] in the *binconf* function of the *Hmisc* package in R [337].

6.3 Results

6.3.1 Survey design and epidemic dynamics

In this section we use the SIMPLE approach to investigate the impact of survey design, particularly the number of daily samples and individual response bias, and underlying epidemic dynamics (the variability in the growth rate) on the quality of estimates. Specifically, we consider when each of the three observation models (basic, extra-binomial, and weighted) are appropriate, and how well they perform when fit to simulated data from each model.

All three observation models produce well-calibrated and similarly narrow credible intervals when fit to simulated data from the basic observation model (i.e., a survey with binomial-distributed observations), despite the extra-binomial and weighted observation models featuring an additional (and in this case, unnecessary) parameter (Figure 6.3). The width of these credible intervals decreases as n_t increases, with 1000 daily samples generally sufficient to produce credible intervals less than 1 percentage point in width, although this is highly simulation-dependent. Larger sample sizes may also allow for the estimation of prevalence by region and/or demographic, as seen in the REACT-1 study. When fit to simulated data with more variable growth rates (higher σ, dashed lines), the width of the credible intervals increases slightly, suggesting that a minor increase in daily sample size may be required when the growth rate is changing faster (to achieve the same level of precision). Finally, the weighted observation model produces poor coverage at low sample sizes, due to the breakdown of the normal approximation to the binomial distribution.

When fit to simulated data featuring extra-binomial variation, only the extra-binomial model consistently produces well-calibrated estimates of P_t (Figure 6.3). At larger values of n_t, when the normal approximation to the binomial distribution is valid, the weighted model also produces well-calibrated estimates of P_t, as the additional parameter c allows the model to account for the extra-binomial variation. The basic model, while well calibrated at smaller values of n_t, produces poorly calibrated estimates as n_t increases, an example of simple modelling assumptions leading to "model-heavy" inferences. This is due to the basic model assuming that the observation variance is inversely proportional to n_t, forcing the credible intervals on P_t to decrease in width as n_t increases, even if additional observation noise is present. The basic model initially accounts for this by artificially increasing the estimated value of σ, allowing variation in r_t to capture the additional variation in n_t^+ (at the cost of biased estimates of σ and r_t - see Supplementary Section B.7.2), although this breaks down as n_t gets very large. This has real-world implications as seen on the REACT-1 dataset, where the original Eales approach (using a binomial observation model) produces noticeably different estimates of r_t and P_t compared to the modified approach using an extra-binomial observation model (Supplementary Section B.4).

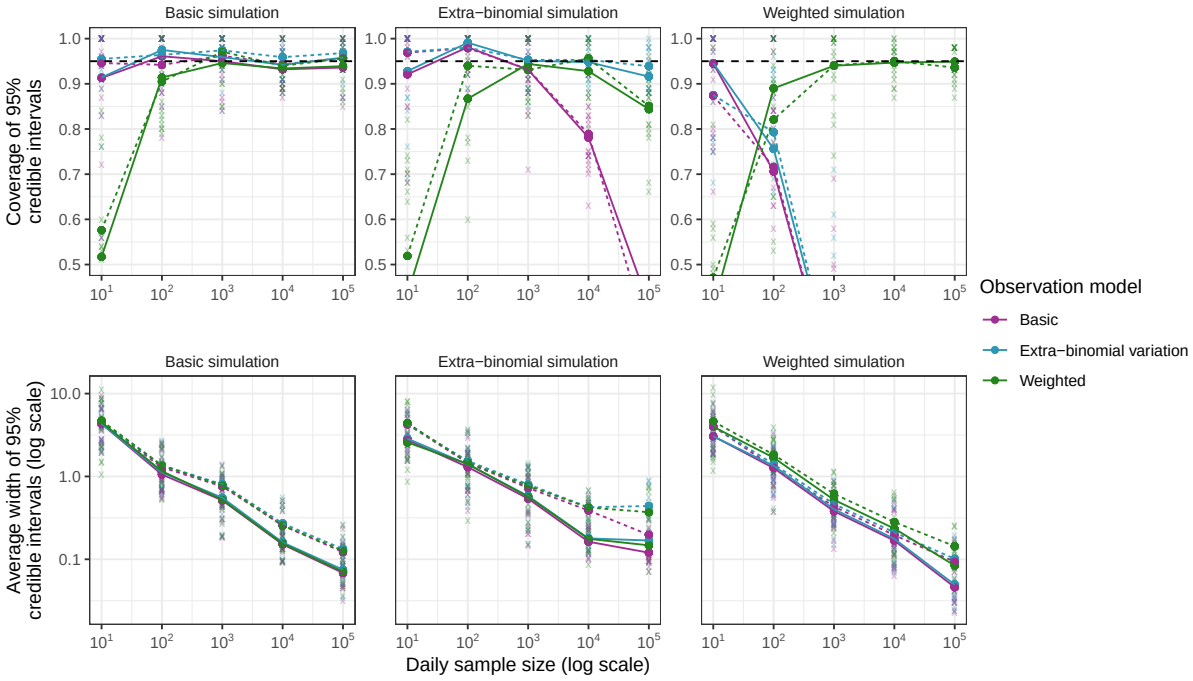

Figure 6.3: Coverage and average width of 95% credible intervals for prevalence P_t (%) from fitting all three observation models (purple: basic, blue: extra-binomial, green: weighted) to simulated data from each model (first column: basic, second column: extra-binomial, third column: weighted). Results from individual simulations are shown as semi-transparent crosses, with averages over 10 simulations shown as points connected by solid lines (for assumed $\sigma = 0.008$) and dashed lines (for assumed $\sigma = 0.016$). A range of assumed daily sample sizes n_t are considered (x-axis). The horizontal black dashed line indicates the target coverage of 0.95. The y-axis for coverage is truncated to $(0.5, 1.0)$, although the coverage in some cases falls outside this range: reaching a minimum average of 0.33 for the basic model fit to the extra-binomial simulations and a minimum average of 0 for the basic and extra-binomial models fit to the weighted simulations (all when $n_t = 10^5$).

Finally, both the basic and extra-binomial models perform poorly when fit to simulated weighted data, with only the weighted model being able to recover the true prevalence (Figure 6.3). Unlike in the other models, estimates of growth rates are less prone to weighting-associated bias (Supplementary Figure B.10). We emphasise caution when translating these results to real-world datasets: the simulated weighted model represents an extreme scenario where weights and individual swab positivity are perfectly correlated and the weights are known. In practice, these weights must be estimated through techniques such as random iterative method (RIM) weighting [338], which introduces additional uncertainty that is not accounted for by this model.

6.3.2 Comparing approaches

In this section, we compare the SIMPLE, Eales, and Abbott approaches on 10 simulated datasets from each model with default parameterisations (Table 6.1), assuming a beta-binomial observation process. All three models produce very similar posterior predictive distributions for swab positivity, and similar posterior distributions for prevalence, suggesting that, given the same observation model, the estimation of P_t and n_t^+ is robust to the differences in smoothing assumptions between models (Figure 6.4).

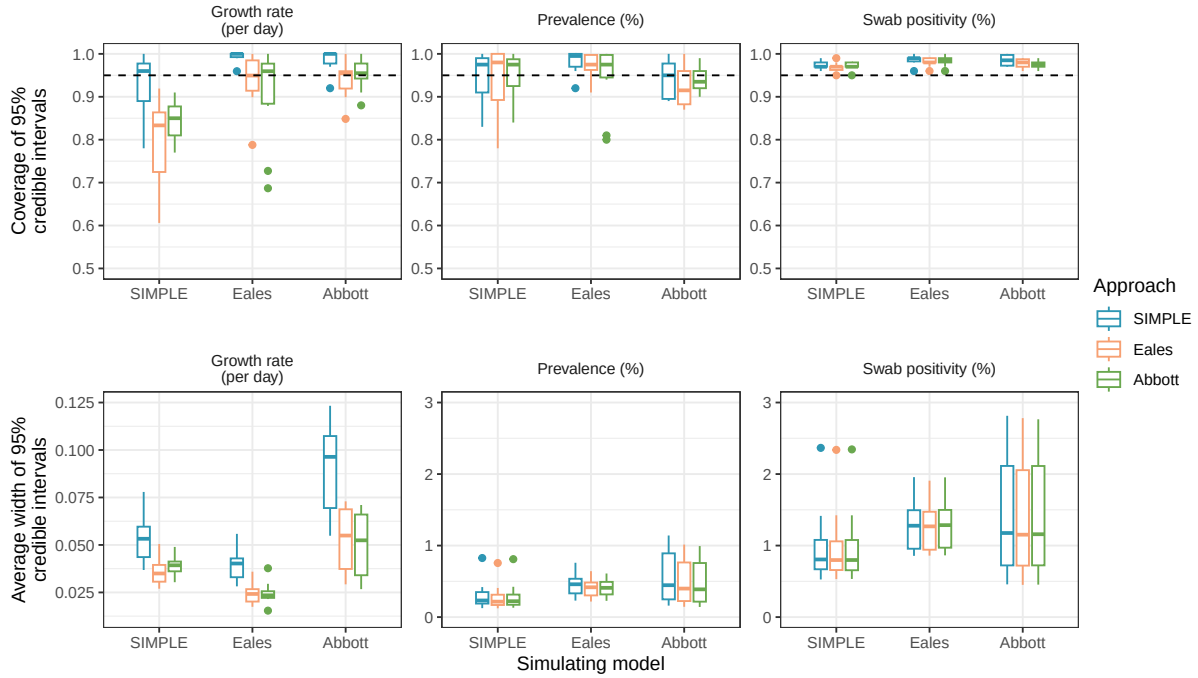

Figure 6.4: Coverage and average width of 95% credible intervals for the growth rate and prevalence, and coverage and average width of 95% predictive credible intervals for swab positivity. Each approach (SIMPLE, Eales, and Abbott) is fit to 10 simulated datasets from each model. The horizontal black dashed line indicates the target coverage of 0.95. Boxes present the interquartile range of the results with the median shown as a horizontal line. Whiskers extend to the most extreme data point within 1.5 times the interquartile range from the box. Outliers are shown as points.

The models in the Eales and Abbott approaches both enforce a fixed minimum amount of smoothness in the data. In the case of the Eales approach, this is via spline knots being placed less frequently than the observed data; while in the case of the Abbott approach, this results from modelling prevalence as a convolution of smooth incidence and the test-sensitivity function and assuming that the incidence function is infinitely differentiable. This minimum smoothness results in the Eales and Abbott approaches producing narrower credible intervals on r_t than

the SIMPLE approach, helpful when the true growth rate is smooth (the SIMPLE approach overcovers r_t in simulations from the Eales and Abbott models), but results in undercoverage when the true growth rate is more variable (Figure 6.4). These are further examples of simple modelling assumptions leading to "model-heavy" inferences.

The total time taken to fit the models to all 30 simulations was 18m 40s for the SIMPLE (extra-binomial) approach, 54m 58s for the Abbott approach, and 47m 44s for the Eales approach. A successful iteration of the Abbott approach takes less time than the Eales approach, however, the Abbott approach sometimes fails to converge, requiring refitting of the model. Further refinement of the code and/or better selection of prior distributions may improve convergence times for the Abbott approach. A more comprehensive comparison of the runtime of each approach is provided in Supplementary Section B.8.

6.3.3 The REACT-1 study

In this section we compare all three approaches on the REACT-1 dataset. First we focus on estimating the growth rate r_t and prevalence P_t, and then consider estimation of the reproduction number R_t. Assuming fixed values of the static parameters over the entire study period of 700 days may not be appropriate [339]. To assess temporal variation in these parameters, we fit the models separately to four time periods: 1 May 2020 to 3 December 2020 (REACT-1 study rounds 1-to-7), 30 December 2020 to 12 July 2021 (REACT-1 study rounds 8-to-13), 9 September 2021 to 17 December 2021 (REACT-1 study rounds 14-to-16), and 5 January 2022 to 31 March 2022 (REACT-1 study rounds 17-to-19). These periods align approximately with changes in the dominant SARS-CoV-2 variant in England (Wildtype, Alpha, Delta, and Omicron) and large gaps in sampling.

All approaches (when using beta-binomial observation distributions) produce similar estimates of the growth rate r_t, prevalence P_t, and predictive swab positivity n_t^+/n_t, when fit to the REACT-1 dataset (Figure 6.5), with an average width of 95% credible intervals for P_t of 0.26 to 0.27 percentage points. Minor differences, largely arising from different prior assumptions about the initial growth rate r_0, are observed in estimates at t near zero. All approaches also produce very similar estimates of the overdispersion parameter ρ, with posterior mean values of 1.9×10^{-4} to 2.0×10^{-4} (Table 6.2). Other parameters are not directly comparable. Finally,

all approaches produced posterior predictive distributions for n_t^+ that slightly overcovered the observed data (Table 6.2). These estimates differ from estimates reported by the REACT-1 study due to our use of a beta-binomial observation model, rather than the binomial observation model used in the original study [241] - we compare these for the Eales approach in Supplementary Section B.4.

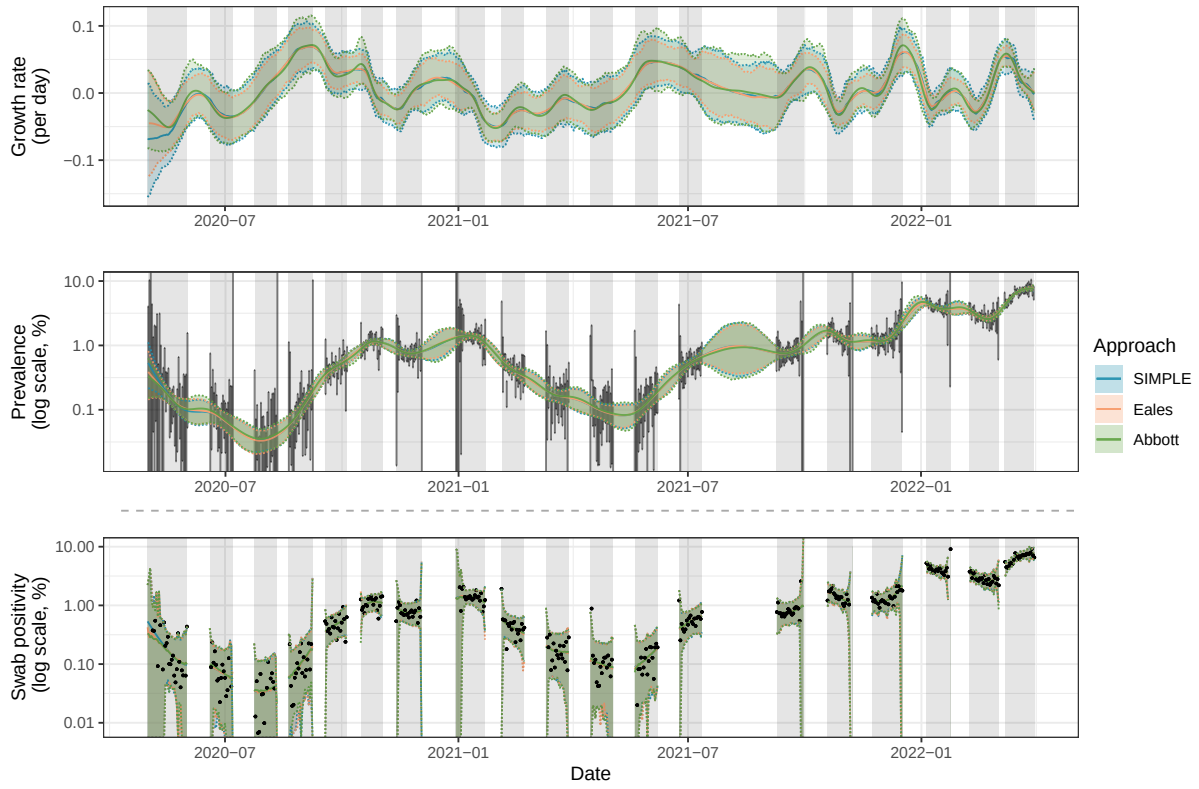

Figure 6.5: Estimates of the growth rate r_t, prevalence P_t, and predictive swab positivity n_t^+/n_t, for SARS-CoV-2 in England between 1 May 2020 and 31 March 2022 using data from the REACT-1 study. All three approaches are fit assuming a beta-binomial observation distribution. Solid coloured lines show the posterior means while shading and dashed lines show 95% credible intervals (of the posterior distribution for r_t and P_t, and of the posterior predictive distribution for n_t^+/n_t). Independent daily confidence intervals from the Agresti-Coull method [324] for P_t are shown in vertical grey lines. The data, daily observed swab positivity n_t^+/n_t, are shown in black points. Grey shading indicates the periods in which sampling was conducted. The predictive distribution for swab positivity depends on the number of swabs taken each day n_t, which tends to be lower in the early and late periods of each sampling round, hence the wider credible intervals at the boundaries of each study round.

Results from the SIMPLE approach (with the growth rate epidemic model) suggest that σ was higher in study rounds 14-to-16 and 17-to-19 (central estimates of $\sigma = 0.015$) than in study rounds 1-to-7 (central estimate of $\sigma = 0.011$) and rounds 8-to-13 (central estimate of $\sigma = 0.0083$), indicating greater variability in the growth rate in later study rounds. The results

Table 6.2: Results from fitting the three approaches to data from the REACT-1 prevalence study. Runtimes were measured once for each dataset considered and can vary considerably. Convergence diagnostics of maximum $\hat{R}$ and minimum ESS are reported, although these are not directly comparable between the SIMPLE and Eales/Abbott approaches. Measures of fit are reported as coverage of the posterior predictive distribution and average width of 95% credible intervals on P_t (in terms of percentage points). Parameter estimates are shown as posterior means with 95% credible intervals in parentheses. Note that σ_{Eales} depends on the knot spacing, which varies slightly between study rounds, so these estimates are not directly comparable even within the same model.

Study round	**All rounds**	**1-to-7**	**8-to-13**	**14-to-16**	**17-to-19**
Observations	400	147	119	67	67
Duration (days)	700	217	195	100	86
SIMPLE approach					
Stopping criteria: max $\hat{R} < 1.05$ and min ESS > 100					
Runtime	4m 12s	42s	46s	25s	14s
Max $\hat{R}$/Min ESS	1.01/113	1.02/110	1.02/100	1.04/110	1.04/129
Coverage of 95% Cr.I. n_t^+	97.2%	97.3%	97.5%	98.5%	98.5%
Avg. width of 95% Cr.I. P_t	0.27	0.16	0.17	0.37	0.84
Parameter σ	0.010 (0.0071, 0.011)	0.011 (0.0055, 0.022)	0.0083 (0.0041, 0.016)	0.015 (0.0068, 0.028)	0.015 (0.0075, 0.028)
Parameter ρ ($\times 10^{-4}$)	1.9 (1.3, 2.7)	2.6 (1.6, 4.5)	1.3 (0.61, 2.2)	1.1 (0.15, 3.1)	2.8 (1.2, 5.0)
SIMPLE approach (reproduction number epidemic model)					
Stopping criteria: max $\hat{R} < 1.05$ and min ESS > 100					
Runtime	23m 38s	3m 9s	1m 26s	59s	1m 23s
Max $\hat{R}$/Min ESS	1.02/101	1.04/106	1.02/101	1.03/109	1.04/127
Coverage of 95% Cr.I. n_t^+	97.2%	96.6%	97.5%	98.5%	98.5%
Avg. width of 95% Cr.I. P_t	0.26	0.14	0.16	0.36	0.79
Parameter σ_R	0.069 (0.050, 0.097)	0.063 (0.034, 0.12)	0.056 (0.023, 0.11)	0.12 (0.046, 0.23)	0.11 (0.050, 0.22)
Parameter ρ ($\times 10^{-4}$)	2.0 (1.4, 2.6)	2.6 (1.8, 3.8)	1.3 (0.55, 2.2)	1.2 (0.12, 2.8)	3.4 (1.6, 6.0)
Eales approach*					
Stopping criteria: 500 samples (200 warm-up and 300 sampling)					
Runtime	38m 54s*	2m 58s	1m 14s	43s	35s
Max $\hat{R}$/Min ESS	1.04/152*	1.02/380	1.02/259	1.01/245	1.01/260
Coverage of 95% Cr.I. n_t^+	96.5%	96.6%	97.5%	97.0%	98.5%
Avg. width of 95% Cr.I. P_t	0.26	0.14	0.15	0.34	0.74
Parameter σ_{Eales}	0.11 (0.084, 0.16)	0.097 (0.051, 0.17)	0.074 (0.033, 0.15)	0.15 (0.060, 0.30)	0.15 (0.079, 0.28)
Parameter ρ ($\times 10^{-4}$)	1.9 (1.4, 2.5)	2.7 (1.7, 4.1)	1.3 (0.59, 2.3)	0.98 (0.075, 2.6)	2.9 (1.2, 5.4)
Abbott approach					
Stopping criteria: 500 samples (200 warm-up and 300 sampling)					
Runtime	6m 43s	47s	49s	40s	29s
Max $\hat{R}$/Min ESS	1.03/223	1.02/163	1.02/124	1.02/144	1.02/164
Coverage of 95% Cr.I. n_t^+	97.5%	98.0%	97.5%	95.5%	97.0%
Avg. width of 95% Cr.I. P_t	0.27	0.15	0.17	0.33	0.75
Parameter σ_{Abbott}	0.074 (0.033, 0.21)	0.069 (0.026, 0.19)	0.077 (0.024, 0.22)	0.10 (0.031, 0.23)	0.081 (0.026, 0.22)
Parameter ℓ	8.9 (0.38, 22)	22 (0.58, 79)	19 (0.29, 100)	3.9 (0.18, 20)	7.3 (0.17, 27)
Parameter ρ ($\times 10^{-4}$)	1.9 (1.3, 2.5)	2.7 (1.7, 4.1)	1.3 (0.54, 2.3)	1.0 (0.16, 2.5)	3.1 (1.5, 5.8)

*In order to obtain convergence of the Eales approach on the full REACT-1 dataset, the "maximum tree-depth" HMC hyperparameter was increased from 15 to 16, resulting in an increase in runtime.

also suggest that ρ was higher in study rounds 1-to-7 and 17-to-19 (central estimates of 2.6 and 2.8, respectively) than in study rounds 8-to-13 and 14-to-16 (central estimates of 1.3 and 1.1), suggesting greater observation noise at the study's start and end. These trends are consistent with results from the SIMPLE approach with the reproduction number epidemic model, and the Eales and Abbott approaches (Table 6.2). While differences in parameter estimates are not statistically significant, estimates of the growth rates r_t do show some sensitivity to whether the models are fit separately to each time period or all together (Supplementary Section B.9.1).

The SIMPLE approach (with the growth rate epidemic model) and Abbott approach exhibit comparable runtimes, taking 4m 12s and 6m 43s on the full dataset (all study rounds), or 14s and 29s on the smallest dataset (study rounds 17-to-19), respectively. This is verified in Supplementary Section B.8, where a more extensive comparison of runtimes is provided. The Abbott approach can sometimes fail to converge, requiring refitting of the model, thus increasing the total time taken to fit the model. The Eales approach is slower than both of these approaches, particularly on larger datasets, taking 38m 54s to fit to the full dataset and 35s to fit to the smallest dataset. This slowdown is partially due to the need to increase the maximum tree-depth HMC hyperparameter from 15 to 16 for convergence on larger datasets. Finally, the SIMPLE approach with the reproduction number epidemic model is the slowest on smaller datasets, but remains faster than the Eales approach on larger ones (taking 23m 8s on the full dataset and 1m 23s on the smallest dataset). The runtimes reported in Table 6.2 are for a single run of each approach. A more extensive analysis of runtimes is included in Supplementary Section B.8.

Despite each approach producing similar estimates of r_t and P_t and a similar predictive distribution for n_t^+ (Figure 6.5), a number of arbitrary decisions were made when fitting these models. For example, we present estimates of r_t and P_t from fitting to all 19 study rounds simultaneously with constant static parameters, rather than grouping the data into shorter time periods. We also assume the extra-binomial observation model is appropriate, and we do not consider survey weights (as daily-applicable weights were not available). We test these specific modelling decisions in Supplementary Sections B.9.1, B.9.2, and B.9.3 respectively. Furthermore, we find evidence for variant-specific parameter values in Supplementary Section B.10. Finally, we compare estimates from all approaches with official consensus estimates of the

growth rate of COVID-19 in England produced by the UK Health Security Agency (UKHSA) [340] in Supplementary Section B.11.

The reproduction number

While all approaches produce very similar estimates of P_t, and similar estimates of r_t when fit to the REACT-1 data, estimates of R_t differ more substantially (Figure 6.6). Additional assumptions are required when estimating R_t and this is a quantity known to be sensitive to such assumptions [341]. All models in this section are fit using the extra-binomial observation model.

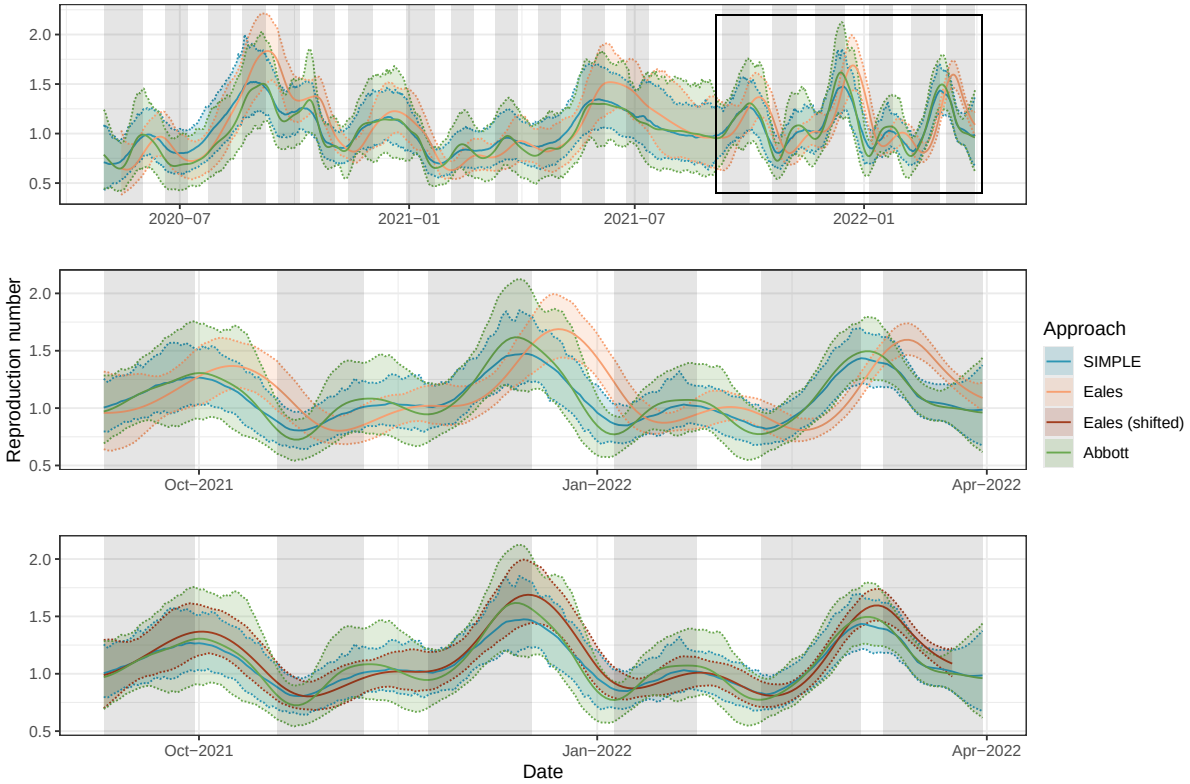

Figure 6.6: Estimates of the instantaneous reproduction number R_t for SARS-CoV-2 in England between 1 May 2020 and 31 March 2022 using data from the REACT-1 study. All approaches are fit assuming a beta-binomial observation distribution. Solid coloured lines show central estimates while shading and dashed lines show 95% credible intervals. Grey shading indicates the periods in which sampling was conducted. The second panel shows the same estimates for a shorter period (9 September 2021 to 31 March 2022), emphasising the differences between the three approaches. The third panel shows the same estimates, with estimates from the Eales approach shifted by $\tau/2 = 7$ days to partially account for bias induced by the trailing-window smoothing assumption.

The SIMPLE and Abbott approaches produce similar central estimates of R_t, with the Ab-

bott approach producing wider credible intervals on average, reflecting the incorporation of uncertainty in the test-sensitivity function (where the SIMPLE approach uses the mean of the test-sensitivity function). Adapting the SIMPLE approach to allow for uncertainty in this function is possible: by treating the test-sensitivity function as a parameter within PMMH, individual PCR positivity curves can be sampled from estimates in the literature and incorporated into the model. By storing accepted samples of the test-sensitivity function, it is possible to integrate out uncertainty about this function alongside other parameters. We leave the implementation of this to future work.

The Eales approach produces estimates of R_t that are delayed and oversmoothed in comparison to the SIMPLE and Abbott approaches (most apparent in Figure 6.6-B), a direct result of the trailing-window approach to estimating R_t. The delay can be partially accounted for by shifting estimates by $\tau/2$ days [36] (Figure 6.6-C), but the oversmoothing is inherent to the approach. Furthermore, as a trailing window of length $\tau = 14$ days does not include the entire generation time distribution or test-sensitivity duration, additional biases are introduced. Note that, while the Eales approach was used to estimate R_t in the published results from the REACT-1 study, our estimates may not align with these due to our use of a beta-binomial observation model instead of the original binomial model.

Despite these differences, there is no clear best approach. Most notably, estimates from the SIMPLE and Abbott approaches depend on the assumed test-sensitivity function, which can vary significantly between settings. This is explored further in Supplementary Section B.5, where we compare three different curves estimated for SARS-CoV-2 RT-PCR tests, and demonstrate how the choice of curve impacts estimates of R_t. We also compare estimates of R_t from these approaches with official consensus estimates produced by the UKHSA [340] in Supplementary Section B.11.

As the reproduction number epidemic model in the SIMPLE approach is no longer Markovian, we must use fixed-lag resampling during the parameter inference stage [40], leading to an increase in computation time (23m 38s for the full dataset or 1m 23s on the smallest dataset). The faster runtime of the SIMPLE approach with the growth rate epidemic model could be obtained by fitting a Markovian model (instead of the renewal model), such as a compartmental susceptible-infectious-recovered-type model, which has the added benefit of not requiring a

test-sensitivity function. However, this type of model places additional assumptions on the underlying epidemic dynamics, which can be difficult to validate. We leave this for future work.

6.4 Discussion

Smoothing prevalence data allows multiple days to inform point estimates, reducing noise and increasing the confidence in estimates. However, all smoothing methods make assumptions, whether explicitly or implicitly, about the underlying process, and results can be sensitive to these assumptions.

A key concern highlighted in this paper is the presence of overdispersion in the observed data. As shown in Section 6.3.1, not properly accounting for this can lead to biased prevalence estimates and credible intervals with poor coverage. Growth rate estimates also become more variable, as the model attempts to explain extra-binomial noise through artificial variability in r_t. In addition to the simulated results presented in this paper, we also find that this impacts real-world estimates of P_t and r_t in the REACT-1 study. An explicit comparison is given in Supplementary Section B.4, emphasising that results can vary considerably, particularly in earlier study rounds. Once overdispersion is appropriately handled, all approaches produce accurate prevalence estimates, even when fit to data generated by other models. This is because prevalence estimates are largely driven by the observed data, not by the smoothing assumptions.

Even when the data are not generated by a beta-binomial sampling process, the overdispersion parameter ρ serves as a general error term that can capture unmodelled heterogeneities and sources of variance. The similarity of estimates of ρ between the approaches (when fit to real-world data from the REACT-1 study) suggests that this parameter is estimable from observed data independently of the assumed smoothing mechanism. Growth rate estimates (and to a lesser extent, prevalence estimates) in the REACT-1 study were sensitive to the inclusion of this parameter. As the allowance for overdispersion, even when simulated data were generated from a binomial model, had little impact on the accuracy of estimates, we recommend including this parameter in all models. If survey weights are available and the daily sample size is sufficiently large (typically $n_t > 100$, although low-prevalence scenarios may require more), then the weighted model is able to account for both overdispersion and

sampling bias, and thus should be the first choice. Unfortunately, without access to survey weights applicable on a daily basis, we were unable to apply this model to the REACT-1 study.

In addition to smoothing prevalence data, the approaches presented in this paper allow for the estimation of the growth rate in prevalence r_t. Unlike estimates of prevalence, assumed smoothing dynamics have a substantial impact on estimates of r_t. Only the SIMPLE approach produced 95% credible intervals with consistently good coverage of this quantity, at the cost of substantially wider intervals. However, as prevalence is naturally a convolution of past incidence, the additional smoothness implied by the SIMPLE approach with the reproduction number epidemic model, or the Abbott approach, may provide an appropriate way to reduce estimated uncertainty about r_t.

Given a predetermined model structure, uncertainty about r_t and P_t arises from two sources: (1) process and observation noise in the epidemic and observation model, and (2) uncertainty about the static parameters governing these models. While uncertainty associated with the former can only practically be reduced by collecting more data, uncertainty about static parameters can be decreased by using more informative prior distributions. At smaller values of n_t, a wide uniform prior distribution can result in the posterior distribution assigning probability mass to implausibly large values of σ and ρ, leading to overestimation of the width of credible intervals for r_t and P_t. Using more informative prior distributions that place less prior mass on implausibly large values can help to reduce the width of the credible intervals for r_t, P_t, and n_t^+, particularly at smaller values of n_t. This is demonstrated in Supplementary Section B.1.2.

Estimating R_t from prevalence data requires assumptions about the relationship between incidence and swab positivity. This can be achieved by assuming R_t is constant over a sufficiently long time period (Eales approach), by incorporating a test-sensitivity function (SIMPLE and Abbott approaches), or by fitting a fully mechanistic model (not considered here). Each approach introduces different sources of potential bias. The trailing-window assumption in the Eales approach guarantees some degree of misspecification [39], which can only be partially mitigated by shifting estimates by $\tau/2$ days [36]. Test-sensitivity functions offer an alternative but also vary substantially: several distinct estimates of this function for SARS-CoV-2 RT-PCR tests were produced during the COVID-19 pandemic [46, 327, 342, 343], with meaningful differences in both duration and peak sensitivity. Ideally this function would be estimated for

each epidemiological setting being modelled, which would require repeat swabs from a subset of survey participants, although collecting sufficient data may be infeasible in low-prevalence scenarios.

An alternative way to relate incidence to prevalence is to model individual-level cycle threshold (Ct) values [336]. Instead of relying on an assumed test-sensitivity function, this approach fits a model of Ct values as a function of time since infection, replacing the observation model for aggregated data with one for individual Ct data. Models for individual Ct value trajectories have previously been estimated from cross-sectional data, without needing repeated sampling from individuals. This method makes fuller use of the information contained in Ct values, rather than reducing test results to binary outcomes, but requires access to individual-level data, is computationally intensive, and depends on correctly specifying the Ct model.

Our SIMPLE approach requires no external software (e.g., Stan), has a faster runtime than the Eales approach, and avoids the Gaussian process approximations used in the Abbott approach. This computational efficiency allows for the rapid testing of a range of models, as demonstrated in Supplementary Section B.9.1, where we use the SIMPLE approach to fit a range of models to the REACT-1 dataset. We also demonstrate an adaptation that allows the SIMPLE approach to estimate variant-specific growth rates while leveraging all collected data (including unsequenced samples) in Supplementary Section B.10. Finally, the sequential nature of SMC allows for easy modification for online inference, where replacing PMMH with an SMC^2 algorithm [344] would allow for real-time inference as new data are collected without the need to refit the model from scratch.

There are two main limitations to the SIMPLE approach. First, the resampling step in the PMMH algorithm is computationally expensive. For Markovian models, past-state resampling can be disabled during PMMH, enabling fast inference. For non-Markovian models, past-state resampling is necessary, slowing down the PMMH algorithm, as seen in the SIMPLE model for the reproduction number. In these cases, the Abbott approach has a faster runtime. Second, PMMH relies on stochastic likelihood estimates, whose variance increases with data length and model complexity. This results in $O(T^2)$ time complexity for parameter inference, which was not a bottleneck for the datasets used here ($T \leq 700$) but will become important for longer

time series. While we use relatively simple models in this paper, more complex models would benefit from more advanced algorithms [289].

While we make several improvements to the Eales and Abbott approaches - including using beta-binomial observation distributions and modifying HMC hyperparameters - further optimisation is likely possible which could improve their relative performance. In particular, the Abbott approach sometimes fails to converge. Reparameterising the model, or using better-specified initial values, could reduce or prevent the occurrence of this. In the Abbott approach, we also recommend considering alternative, less smooth, Gaussian process kernels, as the squared exponential kernel is known to produce overly smooth estimates [345]. The very smooth kernel used in the Abbott approach can be partially compensated for by estimating a more variable growth rate r_t, which is why simulated data from the Abbott model (Figure 6.1) exhibits r_t values of greater magnitude than the SIMPLE and Eales approaches, on average.

Estimate precision can be improved by leveraging additional data sources. The original Abbott approach, for example, included an observation model for antibody data [320]. If individual-level data are available, the approach of Pouwels et al. [150], later refined in [346], can be used to model individual probabilities of testing positive. This approach performs poststratification of individual-level estimates by demographic-geographical response types and has several advantages: multilevel regression and poststratification (MRP) has been shown to outperform classical survey weighting [347], and partial pooling improves estimation of demographic effects (whereas the approaches presented in this paper must produce demographic-specific estimates by fitting separate models). Future work could explore how to combine the approaches and lessons presented in this paper with MRP-type approaches to improve estimates of P_t and r_t.

Our findings indicate that while all three approaches provide robust estimates of prevalence and observed positive swabs, minor differences in their smoothing assumptions can impact growth rate estimates. Applying these approaches to data from the REACT-1 study, we observed that all three produced similar estimates of prevalence and growth rates, with the SIMPLE and Abbott approaches demonstrating greater computational efficiency. By presenting these approaches in common and general notation, alongside well-documented code online, this paper offers a suite of validated tools that can be readily adapted for future epidemic surveys.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

Title of Paper	A Bayesian model for repeated cross-sectional epidemic prevalence survey data
Publication Status	☐ Published ☒ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Steyn, N., Chadeau-Hyam, M., Elliott, P., & Donnelly, C. A. (2025). A Bayesian model for repeated cross-sectional epidemic prevalence survey data. Accepted at *PLOS Computational Biology*.

Student Confirmation

Student Name	Nicholas Steyn
Contribution to the Paper	Conceptualisation, methodology, software, visualisation, data curation, validation, writing (original draft), and writing (review and editing).

Signature NSteyn	Date 22/09/2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description above is accurate.

Supervisor name and title	Professor Christl A. Donnelly
Supervisor comments	

Signature Christl Ann Donnelly	Date 22/09/2025

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 7

Paper V: Jointly estimating epidemiological dynamics of COVID-19 from case and wastewater data in Aotearoa New Zealand

Status: Published in *Communications Medicine* [42].

Full author list: Leighton M. Watson, Michael J. Plank, Bridget A. Armstrong, Joanne R. Chapman, Joanne Hewitt, Helen Morris, Alvaro Orsi, Michael Bunce, Christl A. Donnelly, and Nicholas Steyn.

Authorship contributions: L.M.W., M.J.P., M.B., C.A.D., and N.S. conceived of the study. L.M.W., M.J.P, C.A.D., and N.S. designed the methodology. L.M.W., M.J.P, and N.S. wrote the software, validated the models, created the visualisations, and supervised the study. L.M.W., M.J.P., B.A.A., J.R.C., J.H., H.M., and N.S. conducted the investigation. B.A.A., J.R.C., J.H., and H.M. curated the data and resources. L.M.W., M.J.P., and N.S. wrote the original draft. All authors reviewed and edited the manuscript. A.O. and M.B. administered the project and acquired funding.

Acknowledgements: The authors acknowledge the role of the New Zealand Ministry of Health in supplying data in support of this work. The authors thank the wastewater treatment plant staff members who collected the wastewater samples and the ESR laboratory staff who processed and tested the samples used in this study.

Supporting information: Supplementary material is provided in Appendix C. All code and data used in this study are available at https://github.com/nicsteyn2/NZWastewaterModelling. Daily reported case data for Aotearoa New Zealand are available from the Ministry of Health at https://github.com/minhealthnz/nz-covid-data [348] and seven-day average wastewater data are available from ESR at https://github.com/ESR-NZ/covid_in_wastewater [349]. The data used in this paper are also archived at https://doi.org/10.5281/zenodo.11081779 [350].

Abstract

Background: Timely and informed public health responses to infectious diseases such as COVID-19 necessitate reliable information about infection dynamics. The case ascertainment rate (CAR), the proportion of infections that are reported as cases, is typically much less than one and varies with testing practices and behaviours, making reported cases unreliable as the sole source of data. The concentration of viral RNA in wastewater samples provides an alternate measure of infection prevalence that is not affected by clinical testing, healthcare-seeking behaviour or access to care.

Methods: We construct a state-space model with observed data of levels of SARS-CoV-2 in wastewater and reported case incidence and estimate the hidden states of the effective reproduction number, R_t, and CAR using sequential Monte Carlo methods.

Results: We analyse data from 1 January 2022 to 31 March 2023 from Aotearoa New Zealand. Our model estimates that R_t peaks at 2.76 (95% CrI 2.20, 3.83) around 18 February 2022 and the CAR peaks around 12 March 2022. We calculate that New Zealand's second Omicron wave in July 2022 is similar in size to the first, despite fewer reported cases. We estimate that the CAR in the BA.5 Omicron wave in July 2022 is approximately 50% lower than in the BA.1/BA.2 Omicron wave in March 2022.

Conclusions: Estimating R_t, CAR, and cumulative number of infections provides useful information for planning public health responses and understanding the state of immunity in the population. This model is a useful disease surveillance tool, improving situational awareness of infectious disease dynamics in real-time.

7.1 Introduction

Understanding and predicting the trajectory of infectious diseases is important in planning an effective public health response. Reported case data depend heavily on testing modalities and practices which typically change over time, resulting in considerable uncertainty in the case ascertainment rate (CAR; the fraction of infections that are officially reported). During the COVID-19 pandemic, many countries relied primarily on symptom-based testing programmes to inform situational awareness and public health responses. In Aotearoa New Zealand, the CAR for COVID-19 has been influenced by factors such as access to testing, a shift from healthcare worker-administered polymerase chain reaction (PCR) tests to self-administered rapid antigen tests (RATs), reduction in rates of symptomatic and severe disease due to rising population immunity, relaxation of testing requirements and recommendations, and/or lack of perceived need to test or 'pandemic fatigue' [322, 351, 352]. As a result, over time, officially reported cases of COVID-19 have become a less reliable measure of levels of SARS-CoV-2 infection.

Data on hospital admissions and deaths are more consistent and are less affected by testing practices and behavioural change than reported cases but are subject to additional delays [205] that limit their usefulness for understanding disease dynamics. Infection prevalence surveys [353] that aim to regularly test a representative sample of the population are the gold standard for tracking the spread of an infectious disease, but these surveys are resource intensive, making them harder to justify as countries move out of the acute phase of the pandemic. The UK was the only country to implement regular representative national SARS-CoV-2 prevalence surveys [150, 354] and there are no current plans for similar surveys in New Zealand.

Wastewater surveillance, where levels of SARS-CoV-2 RNA in wastewater samples are measured, can provide additional data on the prevalence of the virus that are unaffected by individual testing and self-reporting behaviours. Wastewater surveillance (also known as wastewater-based epidemiology or WBE) also has the potential to contribute to an integrated global network for disease surveillance [155, 156, 162]. These data, however, can be highly variable and subject to other biases, such as rainwater dilution, sampling methodologies, and changing

locations of selected sampling sites. To realise this potential, appropriate models and analytical tools are needed to deliver epidemiological insights from raw data.

Two previous studies have presented novel methodology for the real-time estimation of the effective reproduction number using wastewater data [355, 356], while others have leveraged or extended these methods [357, 358, 359, 360]. One study used reported cases to estimate the reproduction number and then fitted a model to estimate this quantity from wastewater data [361]. Another study used wastewater data to fit a mathematical model of multiple viral strains [362] from which estimates of the reproduction number can be derived. Other studies have analysed wastewater data but did not use it to estimate the reproduction number [363, 364]. Only [355] presented a model for simultaneously considering clinical and wastewater data, however they assume a fixed ascertainment rate. No previous work has combined wastewater-based epidemiology with reported cases to infer changes in case ascertainment over time.

Semi-mechanistic models based on the renewal equation are a popular method for epidemic forecasting and estimation of the instantaneous reproduction number [114, 116, 139]. Such methods are robust to constant under-ascertainment of cases, but may be biased by rapid changes in CAR and cannot provide any information about the total number of infections. In this paper, we extend the renewal equation framework [114, 116, 139] for reproduction number estimation to incorporate wastewater time-series data. The model treats the instantaneous reproduction number and CAR as hidden states and reported cases and quantity of viral RNA in wastewater as observed states. We use a sequential Monte Carlo approach to infer the hidden states. We apply the model to national data from Aotearoa New Zealand on reported COVID-19 cases and the average number of SARS-CoV-2 genome copies per person per day measured in municipal wastewater samples between January 2022 and March 2023. Because the relationship between infections and wastewater concentration is only determined in the model up to an overall scaling constant, it cannot be used to infer the absolute CAR but can be used to estimate relative changes in case ascertainment over time. The model is designed to be regularly updated as new data become available, producing real-time estimates of the effective reproduction number and relative change in CAR. The model has been used to support situational awareness via regular reports to the New Zealand Ministry of Health from November 2022 to date.

From March 2020 until December 2021 New Zealand used strict border controls and intermittent non-pharmaceutical interventions to suppress and eliminate transmission of SARS-CoV-2. By the beginning of 2022, there had been a cumulative total of around 3 confirmed cases of COVID-19 per 1,000 people and around 90% of the population over 12 years old had received at least two doses of the Pfizer-BioNTech vaccine. From October 2021, interventions were progressively eased and in January 2022 the B.1.1.529 (Omicron) variant began to spread in the community, causing the first large wave of infection. Since then community transmission has been sustained, with multiple further waves of infection being driven by various Omicron subvariants. Between 1 January 2022 and 31 March 2023, there was a cumulative total of around 440 confirmed cases per 1,000 people, most of which were from self-administered RATs. During this period, SARS-CoV-2 concentration was regularly measured at various wastewater treatment plants, providing an additional data source on changes in community prevalence over time.

We model the epidemic dynamics and the observed case and wastewater data at the national level, aggregating over New Zealand's population of 5.1 million and ignoring regional variations. This is similar to other studies that have aggregated regional case and/or wastewater data to produce national-level estimates in countries with a comparable population size [365, 366, 367]. Our methodology could, in principle, be applied at a finer geographical scale, although this would come at the cost of higher levels of noise.

7.2 Methods

7.2.1 Data

National daily reported cases of COVID-19 were obtained from the New Zealand Ministry of Health [348]. Until February 2022, these cases were diagnosed solely by healthcare-administered PCR testing. From February 2022, in response to the rapid increase in reported cases, RATs were widely distributed. Since then, the vast majority of reported cases have been from self-administered RATs, with results reported via an online portal. Hence, data on the number of tests conducted are not available. Reported cases are shown in Figure 7.1. As these data exhibit a clear day-of-the-week effect, we remove the weekly trend before fitting the model (Supplementary Section C.1.1).

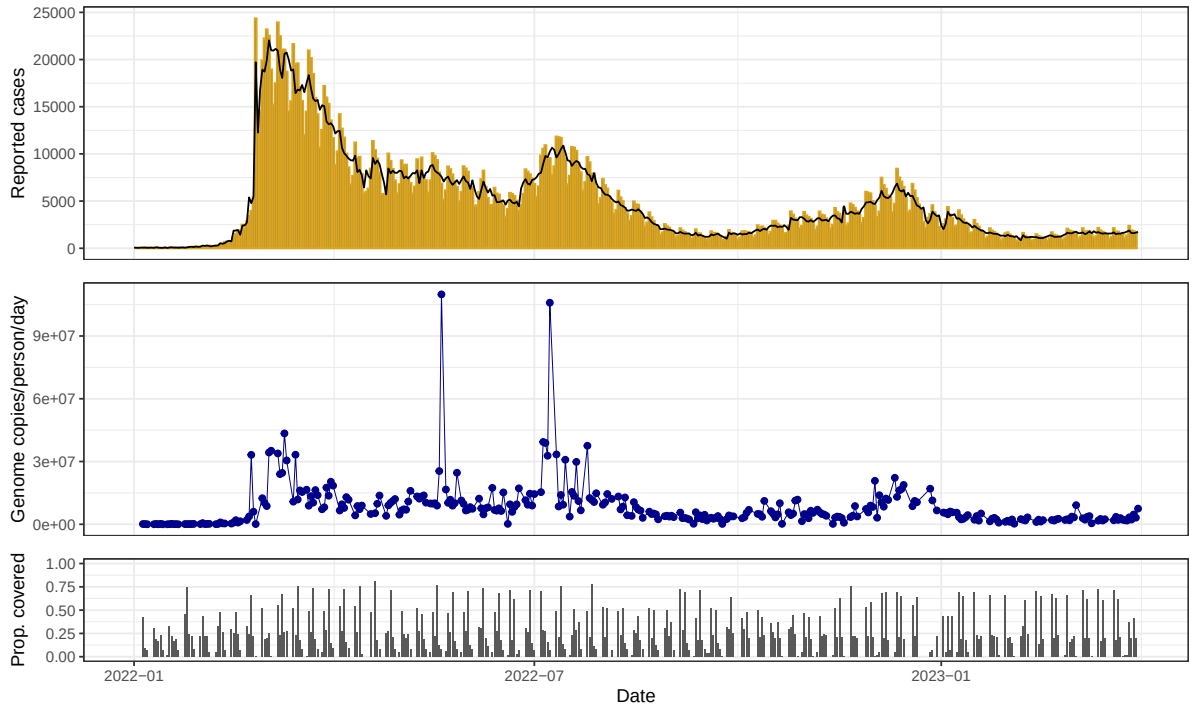

Figure 7.1: **Reported daily cases and wastewater data between 1 January 2022 and 31 March 2023 in Aotearoa New Zealand.** (a) Reported daily cases of COVID-19. The black line shows the adjusted case series with the multiplicative day-of-the-week effect removed (Supplementary Section C.1.1). (b) SARS-CoV-2 genome copies per person per day in sampled wastewater. The two outliers in wastewater data arise from estimates of a high wastewater flow-rate in Wellington following high rainfall. Since rainfall is a source of noise in wastewater sampling, we retain these samples in our analysis. (c) Proportion of the total population covered by sampled wastewater catchments. Reported case data were obtained from the New Zealand Ministry of Health [348] and wastewater data were obtained from ESR [349].

SARS-CoV-2 concentration data from wastewater samples tested by the Institute for Environmental Science and Research (ESR) were used for this study [349]. Wastewater samples were collected every week at municipal wastewater treatment plants located throughout the country, serving communities with populations ranging from 400 to over 500,000 people. Typically 70-90% of the national population connected to reticulated wastewater was covered by wastewater sampling in any given week (60-124 sites, usually sampled twice per week). Each site-level measurement was normalised to provide an estimate of the number of genome copies per person per day for that site (Supplementary Section C.1.1). Typically multiple sites were sampled per day and, for each day that had at least one sample, we calculated the catchment-population-weighted average of the genome copies per person (see Figure 7.1). Because we do not attempt to model regional variations, we assumed this provided a series of representative

observations of the average concentration of genomic material in the national wastewater.

7.2.2 Hidden state model

We construct a state-space model (Figure 7.2) consisting of time-varying hidden states (the instantaneous reproduction number R_t, daily case ascertainment rate CAR_t, and daily infection incidence I_t) and time-varying observed states (daily reported cases of COVID-19 C_t and daily wastewater observations W_t) [368]. We use subscript $s:t$ to refer to all values between day s and t inclusive.

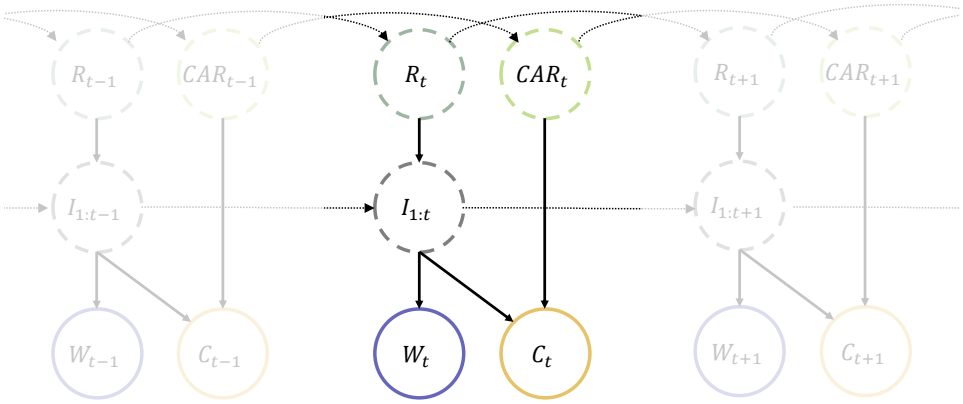

Figure 7.2: **Diagram of the state-space model**. Schematic of model showing the dependency between hidden-states (dashed circles) and the observed data (solid circles). R_t is the instantaneous reproduction number on day t, CAR_t is the case ascertainment rate on day t, I_t is the number of new infections on day t, C_t is the number of reported cases on day t, and W_t is the observed wastewater, measured as the total genome copies per person per day for the sites that were sampled on day t. $I_{1:t}$ denotes the set of states $\{I_1, I_2, \ldots, I_t\}$. In practice the current infections I_t, reported cases C_t and wastewater W_t depend only on recent values of I_t as specified by the generation interval distribution, the infection-to-reporting distribution, and infection-to-shedding distribution respectively (see Methods).

We assume the hidden states R_t and CAR_t follow independent Gaussian random walks, encoding the fact we expect them to vary continuously over time. We also assume that the hidden state I_t follows a Poisson renewal process, a simple epidemic model commonly used

when estimating R_t [114]. Thus, our state-space transitions are governed by:

$$(R_t|R_{t-1}) \sim N_{(0,\infty)}(R_{t-1}, \sigma_R R_{t-1}), \tag{7.1}$$

$$(CAR_t|CAR_{t-1}) \sim N_{(0,1)}(CAR_{t-1}, \sigma_{CAR}), \tag{7.2}$$

$$(I_t|R_t, I_{1:t-1}) \sim Poisson\left(R_t \sum_{u=1}^{t-1} g_u I_{t-u}\right). \tag{7.3}$$

Parameters σ_R and σ_{CAR} determine how quickly R_t and CAR_t vary. The standard deviation of the transition distribution for $R_t \to R_{t+1}$ is given by $\sigma_R R_t$, which means that R_t varies more rapidly at larger values. The distribution for R_t was truncated on $(0, \infty)$ and for CAR_t on $(0, 1)$. Finally, g_u is the pre-determined generation time distribution, describing the proportion of transmission events that occur u days after infection (Supplementary Section C.1.2).

We assume that the expected number of reported cases μ_t^c at time t is equal to CAR_t multiplied by the convolution of past infections with the infection-to-reporting distribution L_u:

$$\mu_t^c = CAR_t \sum_{u=1}^{t} I_{t-u} L_u. \tag{7.4}$$

Similarly, we assume that the expected number of genome copies μ_t^w detected per person at time t is equal to the convolution of past infections with the infection-to-shedding distribution ω_u, multiplied by a fixed parameter α representing the average total detectable genome copies shed into the wastewater by an infectious individual, divided by total national population size N:

$$\mu_t^w = \frac{\alpha}{N} \sum_{u=1}^{t} I_{t-u} \omega_u. \tag{7.5}$$

We model reported cases using a negative binomial distribution:

$$(C_t|CAR_t, I_{1:t}) \sim NegBin\left(r = k_c, p = \frac{k_c}{k_c + \mu_t^c}\right), \tag{7.6}$$

which has mean μ_t^c and variance $\mu_t^c\left(1 + \frac{\mu_t^c}{k_c}\right)$. A negative binomial distribution is used to account for noise in the observations beyond that predicted by a binomial distribution. This is a common choice in other methods of reproduction number estimation [139, 369].

The observed wastewater data W_t is the total genome copies per person from the wastewater sites sampled on day t. We model this using a shape-scale gamma distribution:

$$(W_t|I_{1:t}) \sim \begin{cases} \Gamma\left(k_w \text{pop}_t, \frac{\mu_t^w}{k_w \text{pop}_t}\right) & \text{if } \text{pop}_t > 0 \\ \mathbb{I}(W_t = 0) & \text{if } \text{pop}_t = 0, \end{cases} \tag{7.7}$$

which has mean μ_t^w and variance $\frac{(\mu_t^w)^2}{k_w \text{pop}_t}$. This assumes that the observed daily data are independent draws from the national distribution, which may not hold if there are regional differences between the subsets of sites that are sampled on different days. In practice, any such differences will be absorbed into the variance of the daily observation distribution via fitting of the dispersion parameter k_w. Since we marginalise out the effect of this parameter when presenting results, the increased uncertainty associated with regional variability is propagated through to the credible intervals. The variable pop_t refers to the total population in the catchment areas of the sampled wastewater sites on day t. Setting the variance of the observation distribution to be inversely proportional to pop_t allows the model to account for increased variability around the national mean on days when fewer or smaller sites were sampled. $\mathbb{I}$ is the indicator function, so on days when no sites were sampled, the probability of observing no wastewater samples is set to 1, and the model fits to case data alone.

Consistent with previous models [370, 371], this formulation assumes that the expected population shedding rate is proportional to the number of infected individuals, with observations drawn from a distribution around this mean. We used a gamma distribution, which is a reasonably flexible choice for a non-negative continuous random variable. However, other distributions could be considered, such as a Weibull or log-normal.

In the absence of additional information we are unable to estimate α, which is proportional to the average total genome copies shed by an infected individual over the course of their infection. This means we are unable to estimate the absolute value of CAR_t. Instead, we run the model with a range of different values for α, and estimate the change in CAR_t relative to its initial value. This additionally requires the assumption that α is constant over time, which is unlikely to be true in general and is a key limitation of our model (see the discussion section).

In practice, the range of values of α that we used (Table 7.1) was chosen by calibrating model output for the number of infections with external sources of information. Firstly, we compared model output to the number of cases in a cohort of around 20,000 border workers who were tested weekly between January and July 2022 [372]. Secondly, around 40% of all 20-25-year-olds (an age group unlikely to have a higher CAR than older adults) reported a case of COVID-19 in the 6 months from 1 February to 31 July 2022 [348]. This suggests that the overall

CAR for this period was likely to be at least 0.4, which translates to an approximate upper bound of 4 million for the total cumulative number of infections up to 31 July 2022. Neither of these observations definitively determines the number of infections as they are subject to approximation, bias and uncertainty, but they nevertheless serve to bracket the likely range of values for the parameter α.

Table 7.1: **Parameter values used in the model**. The infection-to-reporting and infection-to-shedding distributions are calculated as convolutions of the incubation period distribution [373] and the onset-to-reporting and onset-to-shedding distribution [157] respectively (Supplementary Section C.1.2).

Parameter	Symbol	Value
Coefficient of variation of R_t transitions	σ_R	Fitted
Std dev. of CAR_t transitions	σ_{CAR}	Fitted
Reported cases tuning parameter	k_c	Fitted
Wastewater tuning parameter	k_w	Fitted
Generation time distribution [374, 375]	g_u	Mean = 3.3 days, s.d. = 1.3 days
Infection-to-reporting distribution	L_u	Mean = 5.8 days, s.d. = 2.6 days
Infection-to-shedding distribution	ω_u	Mean = 5.2 days, s.d. = 2.9 days
Average total genome copies per infection	α	3×10^9 $[2 \times 10^9,\ 4 \times 10^9]$
Fixed-lag resampling window	h	30 days

The infection-to-reporting and infection-to-shedding distributions are calculated as the convolution of the incubation period distribution with the onset-to-reporting and onset-to-shedding distribution respectively. The incubation period is modelled as a Weibull distribution with mean 2.9 days and standard deviation 2.0 days [373]. The onset-to-reporting distribution is estimated empirically from New Zealand case data extracted on 16 September 2022, representing over 1.2 million cases, and has mean 1.8 days and standard deviation 1.8 days. The onset-to-shedding distribution comes from [157] and has mean 0.7 days and standard deviation 2.6 days. The resulting infection-to-reporting distribution has mean 5.8 days and standard deviation 2.6, and the resulting infection-to-shedding distribution has mean 5.2 days and standard deviation 2.9 days (Supplementary Figure C.1).

The model is solved using a bootstrap filter [295] with fixed-lag resampling. This produces estimates for the marginal posterior distribution of the hidden states at each time step. The random walk step variance parameters (σ_R and σ_{CAR}) and observation variance parameters (k_c and k_w) are estimated using a particle marginal Metropolis Hastings Markov chain Monte Carlo method. We use uninformative uniform prior distributions for these parameters, with

the exception of σ_{CAR}, where we use an informative prior distribution to ensure an appropriate level of smoothness in our estimates of CAR_t. Different parameter values are fitted in three-month blocks to allow for some variation over time. See Supplementary Section C.1.2 for further details of the numerical methods used.

7.3 Results

7.3.1 Reproduction number, relative case ascertainment, and infection incidence

The estimated value of the reproduction number R_t (Figure 7.3-a) increases from around 1 at the beginning of 2022 to a peak of 2.46 (95% CrI 2.04, 3.20) on 18 February 2022 (95% CrI 10 Feb, 23 Feb), corresponding to the sharp increase in cases seen during the first Omicron wave, which was a mixture of the BA.1 and BA.2 variants [376]. The estimated value of R_t drops below 1 on 1 March 2022 (95% CrI 25 Feb, 5 Mar) and infection incidence peaked on 28 February 2022 (95% CrI 23 Feb, 7 Mar), suggesting this is when the wave peaked.

The estimated CAR (Figure 7.3-b) increases rapidly between mid-February and mid-March 2022. RATs became widely available for the first time in the last week of February 2022. This likely led to a substantial increase in case ascertainment as the testing system, which had previously relied solely on laboratory-processed PCR tests, had become overwhelmed [352]. The estimated CAR approximately halves between April and July 2022, when a second wave of infection caused by the BA.5 Omicron subvariant [372, 376] occurred. This second wave was visible in both reported cases and wastewater sampling, with estimated peak infections occurring on 7 July 2022 (95% CrI 3 Jul, 12 Jul). The estimated CAR increases somewhat between mid 2022 and early 2023, with a noticeable dip in December 2022, possibly reflecting reduced testing during the Christmas and summer school holiday period (from mid-December to late-January/early-February). Alternatively, the estimated increase in CAR from mid-2022 could be explained by a decrease in the average genome copies shed by an infected individual α, although without further information we are unable to discern changes in α. Overall, the model provides a reasonably good fit to the observed data on cases and wastewater (Figures 7.3-c and 7.3-d).

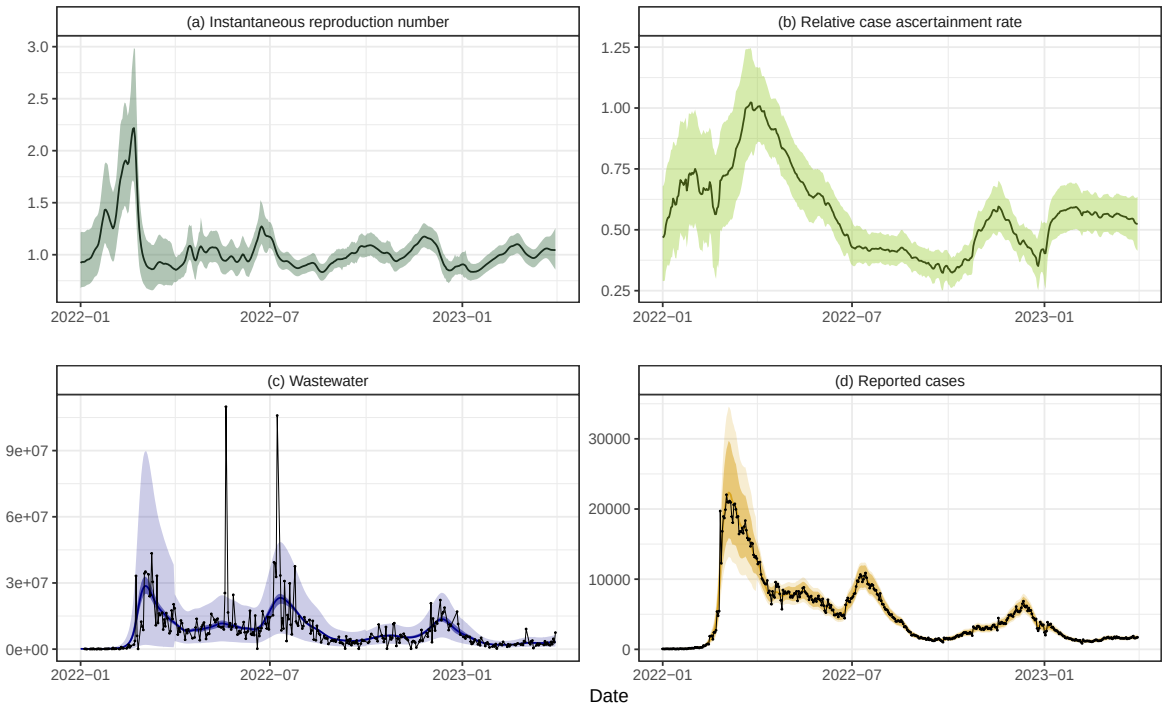

Figure 7.3: **Results for Aotearoa New Zealand data from 1 January 2022 to 31 March 2023.** (a) Instantaneous reproduction number R_t, (b) relative case ascertainment rate (compared to the central estimate on 1 April 2022), (c) wastewater data W_t measured in genome copies per person per day and (d) reported cases C_t. Results assume the average total shedding per infection does not vary over time ($\alpha = 3 \times 10^9$). Solid lines present central estimates. Shaded regions show 95% credible intervals on the value of the hidden states (subplots a and b), and 95% credible intervals on the expected reported cases and wastewater data (darker shaded regions in subplots c and d) and 95% credible intervals on the prediction distribution for wastewater data and reported cases (lighter shaded regions in subplots c and d). Black dots show the observed data.

Figures 7.4-a and 7.4-b show the estimated daily incidence and cumulative infections for three values of α, corresponding to estimated CAR values on 1 April 2022 of 0.42 (95% CrI 0.35, 0.50), 0.61 (95% CrI. 0.51, 0.71), and 0.80 (95% CrI. 0.67, 0.93), for $\alpha = 2 \times 10^9$, 3×10^9, and 4×10^9 respectively. For comparison, the graphs also show the number of cases per capita in a cohort of approximately 20,000 border workers who were tested weekly between January and July 2022 [372], scaled according to population size. These data were used to help inform the range of values of α selected.

Whilst peak reported cases (adjusted for the day-of-the-week effect) in the second wave were only 49% of the peak in the first wave (10,879 vs 22,038 respectively), under the assumption of constant α, the central estimate from the model suggests that true infections peaked at approximately 78% of the peak of the initial wave (Figure 7.4a). Figures 7.4-c to 7.4-e show the estimated absolute and relative CAR and R_t. These panels show that, while we are uncertain about the absolute level of infections and CAR, the relative CAR and reproduction number estimates are robust to reasonable choices for (constant) α.

Fitting the model to case data alone instead of cases and wastewater (Supplementary Figure C.7) produced qualitatively similar estimates of R_t, but with greater temporal fluctuations. Fitting the model to wastewater data alone led to substantially wider credible intervals, although the overall trend was similar. Estimates of the relative CAR are only possible when fitting to case and wastewater data simultaneously.

7.3.2 Parameter estimates

The estimated standard deviation σ_R of the random walk on R_t was greatest in the first time period (1 Jan – 31 Mar 2022) – see Table 7.2. This is unsurprising as it coincided with the rapid increase and then decrease in incidence associated with the first Omicron wave. σ_R decreased in the second period (1 Apr – 30 Jun 2022) and then remained relatively constant throughout the remaining periods (1 Jul 2022 – 31 Mar 2023). The estimated standard deviation σ_{CAR} of the random walk on CAR_t was also estimated to be greatest in the first time period, although this is primarily because we applied a prior distribution with a higher mean in this period (Supplementary Section C.1.2).

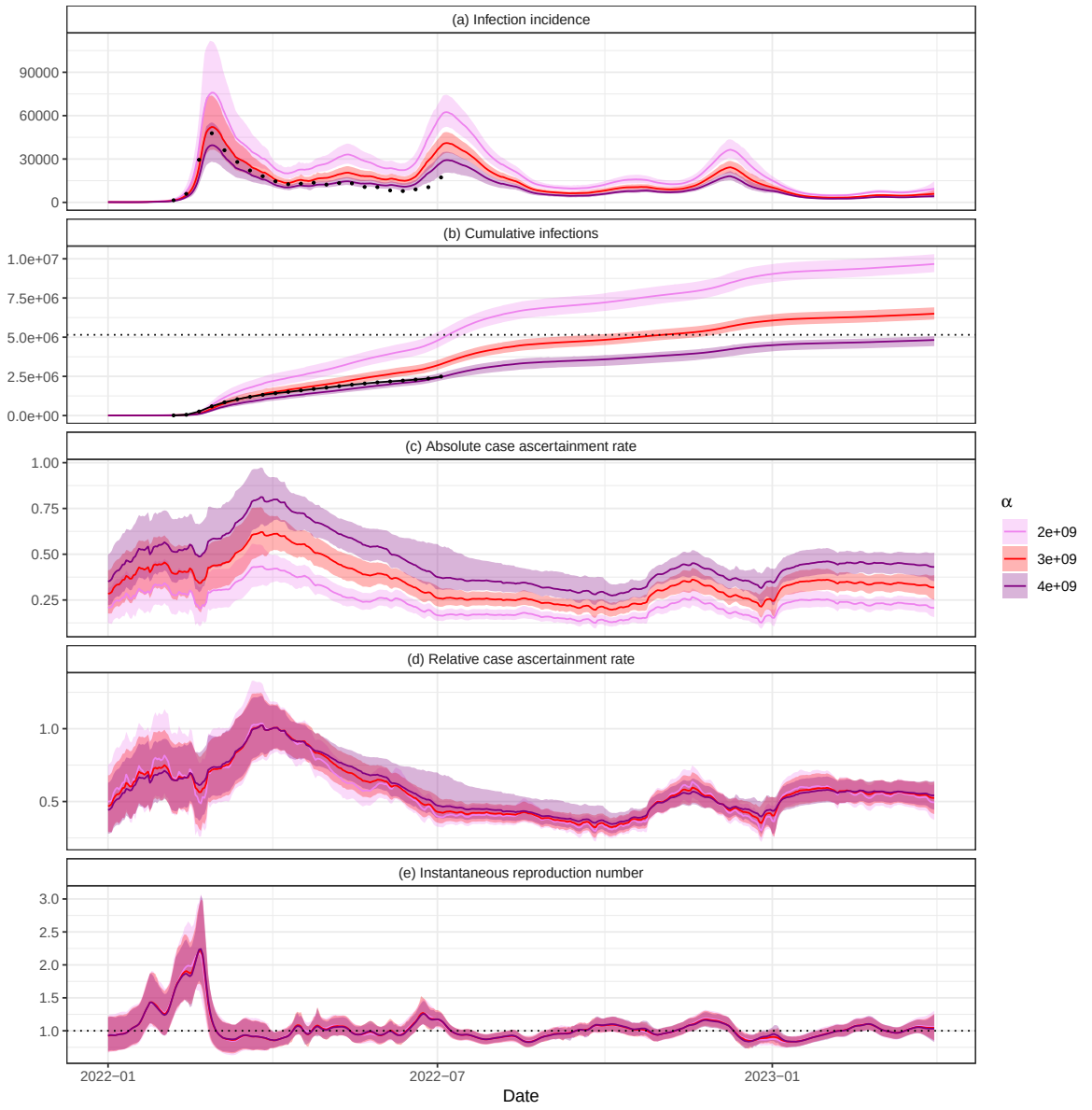

Figure 7.4: **Sensitivity of results to choice of α.** Estimated (a) daily infections I_t, (b) cumulative infections $\sum_{s=0}^{t} I_s$, (c) case ascertainment rate CAR_t, (d) relative case ascertainment rate (compared to the central estimate on 1 April 2022), and (e) instantaneous reproduction number, R_t. Results are presented for three values of α: 2×10^9, 3×10^9, and 4×10^9. Solid lines show central estimates and coloured regions are the 95% CrIs. Estimates and credible intervals on cumulative infections are calculated by taking cumulative sums of the estimates and credible intervals in panel (a). Black dots in panels (a) and (b) show the number of per capita cases in a cohort of regularly-tested border workers, scaled according to population size. The horizontal dashed black line in panel (b) shows the New Zealand population at the end of 2022 (5.15 million people) [178]. While changing α results in different estimates of infections and absolute CAR, the relative CAR and reproduction number estimates are robust to different values, provided α remains relatively constant.

Table 7.2: **Central estimates and 95% credible intervals for estimated model parameters in each time period**. Dates in the 'Period' column are the start date for the three-month period. All outputs presented to 2 s.f. Higher values of σ_R and σ_{CAR} suggest R_t and CAR_t vary faster. Higher values of k_c and k_w indicate a lower variance in the corresponding observation distribution. Note a different prior distribution was used for σ_{CAR} in the first period (Supplementary Section C.1.2), which may also impact estimates of other parameters in this period.

Period starting	σ_R	σ_{CAR}	k_c	$k_w(\times 10^{-6})$
1 Jan 2022	0.12 (0.069, 0.21)	0.03 (0.017, 0.043)	31 (20, 49)	1.5 (1.1, 2.0)
1 Apr 2022	0.069 (0.041, 0.12)	0.0099 (0.0053, 0.014)	170 (100, 250)	4.8 (3.2, 6.8)
1 Jul 2022	0.037 (0.02, 0.066)	0.0063 (0.0018, 0.01)	330 (220, 400)	4.8 (3.3, 6.5)
1 Oct 2022	0.038 (0.02, 0.068)	0.011 (0.0073, 0.014)	170 (110, 270)	7.2 (4.7, 10.0)
1 Jan 2023	0.038 (0.018, 0.073)	0.0093 (0.0041, 0.015)	150 (84, 330)	6.8 (4.4, 10.0)

The estimated variance parameters, k_c and k_w, for cases and wastewater observations, were lowest in the first time period (1 Jan 2022 – 31 Mar 2022). This implies there is more variability in the data that is not explained by the model in this time period, possibly as a consequence of the sharper variations in incidence compared to the later time periods. A less consistent weekly pattern in reported cases during the first time period, and higher levels of noise in wastewater observations at the low concentrations seen at the beginning of 2022, could also be contributing factors.

7.4 Discussion

Wastewater-based epidemiology has been used globally for COVID-19 surveillance and has been shown to be a useful public health tool for policy and public health responses [377]. We present a semi-mechanistic model that combines reported cases with wastewater data to estimate the time-varying reproduction number and CAR. This work demonstrates the value of wastewater-based epidemiology and how the additional data that it provides can be combined with traditional monitoring (e.g., reported cases) to learn more about the state of an epidemic, disease dynamics, and the true number of infections in the community. This provides useful information to inform the public health response.

To make reliable estimates of the state of the epidemic from reported cases, it is essential to understand how case ascertainment changes with time. For example, are there fewer cases because there are fewer infections or because fewer people are reporting? We apply our model to national data from Aotearoa New Zealand and derive insights into changes in case ascertainment

that would not be possible using case data alone. Reported cases during the second wave in July 2022 were substantially lower than in the first wave in February and March 2022. However, the model infers that there was a substantial drop in case ascertainment between these waves, and the true number of infections was likely more similar in each wave. The reduced CAR during the second and subsequent waves may have been due to a higher number of reinfections with individuals displaying fewer symptoms or due to "pandemic fatigue" and reduced compliance with public health measures, including testing. This type of insight would not be possible without regular wastewater surveillance data and without a robust analytical framework in which to integrate these data with traditional epidemiological data streams.

We apply our model to the first period of widespread community transmission of SARS-CoV-2 in New Zealand. During this time, rapid antigen tests were freely available to everyone, there was a requirement to report positive results, and a mandatory isolation period for cases with financial support via employers. Partly as a result of these factors, the CAR, while lower than in the previous elimination phase, was still reasonably high. The mandatory isolation period was removed in September 2023, which led to a substantial drop in case ascertainment. For the datasets we consider, similar (albeit noisier) estimates for the reproduction number could be obtained from case data alone. However, in a context where case ascertainment is low and/or unrepresentative, wastewater data are likely to add even greater value compared to using reported cases. In contrast, in a low-prevalence context (e.g., pre-Omicron in New Zealand), applicability of the method would be constrained by the amount of noise in the wastewater data. In this situation, wastewater surveillance may better used for presence/absence monitoring, for example as an early warning system for the presence of infection in specific catchments, as opposed to quantitative estimation [378].

Strengths of our model include the fact that it has relatively minimal data requirements, requiring only time series for reported cases and wastewater concentrations. The model can be fitted to datasets in which different sites are sampled on different days and some days have no observed data. This means that it could be readily applied in other jurisdictions with wastewater surveillance programs, either for SARS-CoV-2 or other pathogens such as influenza viruses [377, 379]. It is a relatively simple model with minimal mechanistic assumptions and

parsimonious parameterisation. This avoids the need for assumptions about time-varying contact patterns, transmission rates, and the level of prior immunity that are required by more complex mechanistic models. The model we present here was operationalised by ESR in late 2022 and results for R_t and relative CAR are regularly provided to the Ministry of Health to inform situational awareness and decision-making.

There are several limitations to this model and the results. We assume that the average number of genome copies shed by an infected individual (represented by the parameter α) was constant between January 2022 and March 2023 and did not depend on the infecting variant or history of prior infection or vaccination. It is possible that some of the inferred changes in CAR may be partly explained by these factors. For example, some of the inferred increase in case ascertainment between October and December 2022 may have been due to decreasing α, caused by a combination of new immune evasive subvariants displacing the previously dominant BA.5 variant [380] and/or an increase in the proportion of reinfections or asymptomatic infections [348]. Although estimates of viral shedding rates per infected individual are available [370, 371], the value of α may also depend on physical characteristics of the wastewater collection system, sample collection method, and the method used to quantify concentration of SARS-CoV-2 RNA in samples. Therefore, α is likely to vary between jurisdictions and will require recalibration using local data.

As we are unable to estimate the true value of α, we are unable to estimate the absolute CAR. Nonetheless, relative CAR is a useful metric and, given an estimated range of values for α, we are able to provide plausible bounds on the total number of infections (Figure 7.4).

Wastewater surveillance does not provide any information on how infections are distributed among population groups (e.g., age groups, ethnicity) and biases in self-administered testing mean that case counts are not representative either. This information is important for assessing the clinical burden of disease and addressing health inequities [12]. Thus, other approaches are needed to determine the distribution of disease burden, such as representative sampling [150, 381], cohort studies [382] or sentinel surveillance [383, 384]. Although wastewater surveillance could, in principle, be used to investigate differences in prevalence and case ascertainment between sites and/or regions, this would require adaptations to our method that are beyond the scope of this study.

As our model is flexible, future work could integrate hospitalisations (such as in [385]) and deaths data. In principle, this could allow the effects of varying CAR and varying rate of shedding per infection to be separated. However, this would additionally require the effects of age, immunity, ethnicity, and other variables on clinical severity to be accounted for.

Although national-level approaches to situational awareness and reproduction number estimation are common [341, 366, 386], particularly in countries such as New Zealand with a relatively small population size, this ignores regional variations. Results should therefore be interpreted as national averages, which could mask demographic and spatial heterogeneity. Our model could be implemented at a regional level so that local epidemic dynamics can be compared, although this would be subject to increasing levels of noise in the wastewater data at finer spatial scales. This paper has focused on modelling for inference: understanding epidemic dynamics that have already occurred. However, the state-space transition model coupled with the estimated parameters provides a natural method for forecasting [139, 387]. Forecasts generated using this state-space transition model naturally incorporate increasing uncertainty about the future reproduction number and CAR.

While this model focused on COVID-19, there is a wealth of genetic information within municipal wastewater that could also benefit from modelling. The detection and concentration of viral, bacterial and anti-microbial resistance genes within wastewater have the ability to inform public health decision-making in a number of ways, especially as methodology is refined allowing more rapid turnaround times. As many jurisdictions seek to retain the wastewater capabilities they built during the pandemic phase of COVID-19 (and to diversify microbial targets), there is an 'opportunity springboard' to build tools that can predict the trajectories and spread of pathogens. Modelling has a key role to play in this journey.

Chapter 8

Paper VI: Pandemic-risk-related behaviour change in England from June 2020 to March 2022: REACT-1 study among over 2 million people

Status: Submitted for publication. A preprint is available on medRxiv here: https://doi.org/10.1101/2025.03.03.25323250. Minor revisions have been made since the preprint was posted.

Full author list: Nicholas Steyn, Marc Chadeau-Hyam, Matthew Whitaker, Christina Atchison, Deborah Ashby, Graham S. Cooke, Helen Ward, Paul Elliott, and Christl A. Donnelly.

Authorship contributions: N.S., M.C-H., and C.A.D. conceived of the study and wrote the original draft. N.S., M.C-H., M.W., and C.A.D. designed the methodology. N.S., M.C-H., M.W., C.A.D., and P.E. performed the investigation, created the visualisations, validated the methodology, and reviewed and edited the manuscript. N.S. and M.W. wrote the software. All authors contributed to data curation. M.C-H., P.E., and C.A.D. performed project administration.

Supporting information: Supplementary material is provided in Appendix D. Aggregated and anonymised data are publicly available online: https://github.com/nicsteyn2/REACTBehaviouralData.

Abstract

Objective: To determine how people in England changed their infection risk-related behaviours during the COVID-19 pandemic and in response to control measures, 19 June 2020 to 31 March 2022.

Design: Over 18 (of 19) rounds, randomly selected participants across England completed a questionnaire about risk-related behaviours, socio-demographics, and symptoms.

Participants: Between 85,018 and 154,060 randomly selected participants per round, aged 5+ years, totalling 2,177,657 responses with relevant data.

Main outcome measures: Primary outcomes were self-reported shielding and/or taking specific precautions, not leaving home in the prior week, not being in close proximity with anyone outside their household the day before, and wearing face coverings outside the home. Secondary community-level measures of mobility and public health policy stringency were compared to the primary outcomes to provide population-level context to the observed findings.

Results: Infection risk-related behaviours varied considerably over the nearly two years under study. Protective behaviours peaked in January 2021, during England's winter wave, before widespread vaccination. At that time, the estimated proportion of self-reported shielding and/or taking specific precautions reached 21.6% (95% confidence interval 21.4%, 21.8%), of self-reported not leaving home in the week prior to completing the questionnaire reached 7.99% (7.85%, 8.13%) and of self-reported not having contact with anyone outside their household on the day before answering the questionnaire reached 89.2% (89.1%, 89.4%). As self-reported vaccination rates increased and prevalence of infection decreased, protective behaviours decreased, although patterns varied by demographics. Protective behaviours were strongly correlated with community-level mobility data and the stringency of public health measures.

Conclusions: Individual-based data showed sizeable proportions of people undertook protective behaviours during the pandemic especially during the second lockdown in January 2021 although there was evidence of "pandemic fatigue" in the study's later stages. Self-reported behaviours were closely aligned with community mobility data and the stringency of government policies, indicating policy-driven behavioural changes.

8.1 Introduction

The COVID-19 pandemic in England was a complex and quickly evolving public health crisis consisting of multiple waves of infection. The REal-time Assessment of Community Transmission-1 (REACT-1) study provided real-time estimates of the prevalence of SARS-CoV-2 swab positivity in England from 1 May 2020 to 31 March 2022 [48]. Conducted over 19 rounds, more than 2.5 million nasal swab samples were processed using reverse transcriptase polymerase chain reaction (RT-PCR), and participants completed an extensive questionnaire about their behaviours related to infection risk, socio-demographic characteristics, and symptoms.

Various non-pharmaceutical interventions (NPIs) were utilised throughout the pandemic, including stay-at-home orders ("lockdowns") on multiple occasions, notably in March 2020 (prior to the start of REACT-1) [388], and then in the winter of 2020/21 (around the time of study rounds 7-to-9) [389]. These interventions were mostly lifted in the middle of 2021, although some intermediate measures, such as recommended working from home, were re-introduced in December 2021 to slow the spread of the Omicron variant [390].

Changes in individual behaviours, influenced by perceived risks, governmental guidelines, societal norms, and "pandemic fatigue", could directly impact the trajectory of the epidemic and the effectiveness of policy decisions. Understanding how these behaviours changed over the course of the pandemic is critical to quantify the impact of NPIs, to gauge their effectiveness at reducing SARS-CoV-2 transmission, and to understand how compliance varied across demographic groups and over time [391].

Community-level mobility data, such as the Google community mobility reports [153], have been widely used in epidemiological studies to evaluate the effectiveness of social-distancing measures [392] or to predict the rate of virus transmission [167]. These data offer valuable insights but are only aggregated proxies for individual behaviours and for social interaction.

The Oxford COVID-19 Government Response Tracker (OxCGRT) [154] collected information on a variety of policy measures that governments used during the pandemic such as school closures and workplace restrictions. In particular, their stringency index has been used for comparing the strictness of policy responses between countries, and has featured directly in

epidemiological models [393]. Their containment and health index closely resembles the stringency index, with additional consideration given to contact tracing, testing, face covering policy, vaccinations, and protection of elderly people.

Focusing on self-reported behaviours relating to shielding and/or taking specific precautions, not leaving the home, not having physical contact/close proximity with people outside the household, and wearing face coverings outside the home, we estimated the self-reported prevalence of these behaviours from 19 June 2020 to 31 March 2022 in England and by socio-demographic group. We also investigated the consistency of individually reported and community-level behavioural data by relating these estimates to (i) (Google) mobility data and (ii) the OxCGRT stringency index and OxCGRT containment and health index.

8.2 Materials and methods

8.2.1 REACT-1 data

The REACT-1 study conducted 19 rounds of sampling from 1 May 2020 to 31 March 2022 with 85,018 to 154,060 participants per round. Participants completed a comprehensive survey reporting their risk-related behaviours, socio-demographic characteristics, health and health history, history of COVID-19, and risk tolerances. The survey was completed by the respondent (if over the age of 18), by the respondent's parent/guardian (for ages 5 to 12), and either by the parent/guardian or by the respondent themselves (for ages 13 to 17). Round-specific random iterative method (RIM) weights [338] were calculated at the time of survey administration to enable prevalence estimates representative of the population of England. These were calibrated for participants' age, sex, area-level deciles of deprivation [394], lower-tier local authority counts and ethnicity [48]. A Public Advisory Panel provided input into the design, conduct, and dissemination of the REACT research programme.

Of the questions on behaviours, we focused on one related to shielding ("Are you shielding and/or taking specific precautions because you are concerned that you/your child will become severely ill with COVID-19?"), one to leaving the home ("Did you/your child leave home for any reason in the last 7 days?" and later "In the last 7 days, for what reasons have you left home? Select all that apply."), one to social distancing ("Not including members of your household, how many different people did you have contact with yesterday? By contact, we mean: any

direct skin-to-skin physical contact (e.g., kiss/embrace/handshake), being less than 2 metres from another person for over 5 minutes."), and one to wearing face coverings outside home ("Do you/does your child mainly wear any kind of face covering or mask when you/they are outside your/their home, because of COVID-19?"). The specific question wording occasionally changed between study rounds (Supplementary Section D.1.1, Supplementary Tables D.1 to D.3).

During study round 1, the relevant questions were not included in the questionnaire; we therefore excluded round 1 participants from this analysis ($N = 120,620$). Of the REACT-1 participants in study rounds 2-to-19 (19 June 2020 - 31 March 2022) ($N = 3,239,620$), we excluded, prior to analysis, participants with invalid survey weights (N=1,019) as well as those that registered but did not complete the questionnaire (N=1,060,944). Some participants did not answer every question, provided an invalid response, or were not asked a specific question (the latter primarily being due to age). When responses to individual questions were missing, the response was discarded (set as "NA"). Hence, reported proportions are relative to the total number of respondents to each question. A round-by-round breakdown of non-response rates by question is provided in Supplementary Section D.1.2 (Supplementary Tables D.5 and D.6).

8.2.2 Google community mobility data

Community mobility data were obtained from Google for the period 15 February 2020 to 15 October 2022 [153]. These publicly available data were aggregated from device users who had turned on the location history setting (by default it is off) and present the day-to-day percentage change (relative to a day-of-the-week-specific baseline calculated from the five-week period from 3 January 2020 to 6 February 2020) in the total number of visitors to places of retail and recreation, grocery and pharmacy, parks, transit stations, and workplaces as well as the daily percentage change in time spent at residential locations.

8.2.3 Oxford COVID-19 Government Response Tracker data

The OxCGRT stringency index and containment and health index were obtained for the period 1 January 2020 to 31 December 2022. These data aggregate indicators measuring the implementation of school closures, workplace closures, public event cancellations, restrictions

on gathering sizes, closure of public transport, stay-at-home requirements, restrictions on internal movement, restrictions on international travel, and public health information campaigns. The containment and health index additionally considers testing, contact tracing, face covering requirements, and vaccination policy.

8.2.4 Statistical methods

Results are presented at the national level, as well as for four age-groups in years (5-17, 18-34, 35-64, 65 and over) and four household sizes (lives alone, 2 people, 3-4 people, 5+ people). Additionally, we considered geographical area, area-based deciles of deprivation [394], ethnicity, and sex [48] in Supplementary Section D.2.

The *svyciprop* function with the logit method, from the *survey* package in R, was used to estimate weighted round-level binomial proportions and corresponding confidence intervals [132] from the REACT-1 survey data. Daily prevalence estimates were unweighted proportions because day-specific RIM weights were not available, and the corresponding confidence intervals were calculated using the *binconf* function from the *Hmisc* package [337] in R.

We analysed if and to what extent responses to the behavioural questions varied by age-group, sex, geographical area, deprivation, ethnicity, and household size, via survey-weighted logistic regression models. These models were fit to data from each study round using the *svyglm* function with the binomial family from the *survey* package in R [132]. Results from these models are presented as odds ratios (ORs), providing the relative odds of reporting a given behaviour compared to a specified reference group.

To evaluate the consistency of the Google community mobility data and OxCGRT indices with self-reported behaviours we fitted random forest regression models, using daily aggregated REACT-1 data to predict each mobility series and index separately. When the mobility series were used as the outcome variable, we used the daily proportion of people leaving their homes for various reasons (and not leaving) as predictors (the variables listed in Supplementary Table D.3). When the OxCGRT indices were used as the outcome variable, we used the daily proportion of people reporting shielding and/or taking specific precautions, not leaving home in the prior week, not having physical contact/close proximity with people outside their household the day before, and wearing face coverings outside the home as predictors (4 total predictors).

To reduce the impact of day-of-the-week variability in the mobility data, we smoothed these series using a trailing 7-day moving average prior to fitting the models. To limit measurement error in our predictors we only considered days when more than 200 survey responses were recorded.

We used the *randomForest* package in R [395] with default hyperparameters (500 trees each fit with approximately 63.2% of the available observations, selected at random) to fit the random forest models. Model performance was assessed using the proportion of the within-bag variance explained (the proportion of variance in the response that is explained by the predictors when testing on data used to fit the models) and out-of-bag variance explained (the proportion of variance in the response that is explained when testing on data not used to fit the models). The within-bag variance explained is analogous to R-squared in a linear regression context.

Three separate models for each mobility series were constructed to reflect changes in question formulation between study rounds. We defined three main periods (i) from 19 June 2020 to 11 November 2020 (study rounds 2-to-6, N=94 days with $>$ 200 survey question responses, 11 predictors), (ii) 4 January 2021 to 28 September 2021 (study rounds 8-to-14, N=135 days with $>$ 200 survey question responses, 8 predictors), and (iii) 19 October 2021 to 31 March 2022 (study rounds 15-to-19, N=104 days with $>$ 200 survey question responses, 14 predictors). Study round 7 is not included in this analysis as the question regarding reasons for leaving the home in the past 7-days was not asked. One model was fit for each OxCGRT index, covering the period 16 October 2020 to 31 March 2022 (study rounds 6 and 8-to-19, N=259 days with $>$ 200 survey question responses, 4 predictors).

8.2.5 Data portal

We have developed an online portal to facilitate visualisation and download of aggregated behavioural data while protecting the confidentiality of survey participants. This can be accessed here: `https://m-whit-ic.shinyapps.io/react-social-shiny/`. Further information about this portal is provided in Supplementary Section D.6.

8.3 Results

8.3.1 Behaviours over time

We observed a peak in reported protective behaviours in January 2021 (study round 8), at the time of the third national lockdown in England (Figure 8.1). At this time, self-reported shielding and/or taking specific precautions reached 20.8% (95% CI 20.6%, 21.1%), self-reported not leaving home in the 7 days prior to completing the questionnaire reached 9.7% (95% CI 9.5%, 9.9%), and self-reported not having physical contact/close proximity with anyone outside the household on the day before answering the questionnaire reached 93.3% (95% CI 92.8%, 93.7%).

Following this peak, reported protective behaviours decreased over the summer months in 2021, with a noticeable uptick again in January 2022 (Figure 8.1, study round 17), during the first Omicron wave and the government's "Plan-B" interventions and messaging [396]. This uptick did not result in social-distancing behaviours reaching the same level as winter 2020/21, except for the proportion of people reporting shielding and/or taking specific precautions. To some extent, the increase between December 2021 and January 2022 (study rounds 16 and 17) of this behaviour can likely be attributed to rewording of the question (specifically, the removal of the word "shielding", a term with a much narrower definition than "specific precautions"). These wording changes are described in Supplementary Section D.1.1.

After the uptick in December 2021/January 2022, protective behaviours decreased rapidly over February and March 2022 (study rounds 18 and 19) [49], despite infection prevalence reaching the highest measured by the study in March 2022 (weighted swab positivity in round 19 was 6.37%). The proportion self-reporting wearing face coverings reached the lowest recorded level in the study of 63.2% (95% CI 62.9%, 63.6%), down from 91.6% (95% CI 91.4%, 91.8%) in January 2022 (study round 17) and 92.4% (95% CI 92.2%, 92.6%) in January 2021 (REACT-1 study round 8). Similarly, we observed reductions in the proportion self-reporting no contacts outside the home 41.2% (95% CI 40.9%, 41.6%) in March 2022, down from 51.3% (95% CI 50.9%, 51.7%) in January 2022 and 88.6% (95% CI 88.3%, 88.8%) in January 2021.

Figure 8.1 shows that those aged 65 years and older were more likely to report protective behaviours to limit the risk of SARS-CoV-2 infection (except for reporting not leaving home)

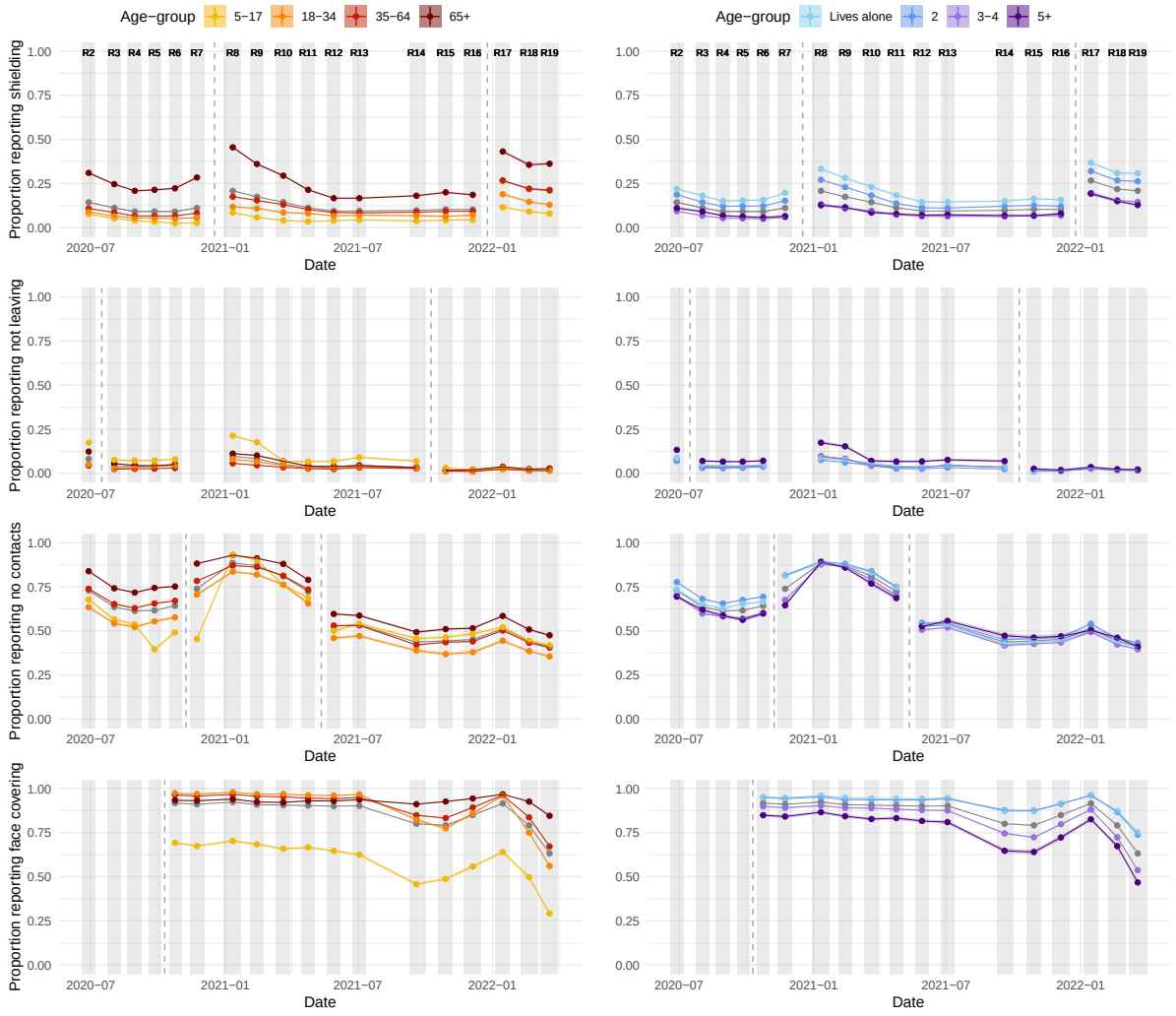

Figure 8.1: The proportion of people in England self-reporting shielding and/or taking specific precautions because they consider themselves "to be at risk of severe illness for COVID-19" (first row), not leaving the home on the day preceding answering the questionnaire (second row), having no contacts outside the home in the preceding seven days (third row), and wearing a face covering outside the home (fourth row) by age-group (left column) and household size (right column). Changes in question wording are demarcated with a vertical dashed grey line - caution should be exercised in comparing values on either side. Shaded regions and vertical coloured lines show 95% confidence intervals about the estimated proportion, although they are frequently too small to discern. Individual points are joined by lines and shading for clarity, but this should not be interpreted as an interpolation. National averages are shown in grey.

than younger participants. A similar but less marked difference was observed for those reporting living alone when asked whether they had left the home. Responses to these behavioural questions disaggregated by other demographics are considered in Supplementary Section D.2 (Supplementary Figures D.1 to D.4): those living in neighbourhoods with greater socioeconomic deprivation were more likely to report protective behaviours, although less likely to report

wearing face coverings. Results were less consistent between questions and over time for region, sex, and ethnicity.

Survey-weighted logistic regression models estimating the independent effect of each demographic variable, show that, compared to 18-34-year-olds, those aged 65+ were more likely to report having no contacts outside the home, with ORs between 1.64 (95% CI 1.55, 1.73) and 2.99 (95% CI 2.84, 3.14, $p < 0.05$) (Figure 8.2). The impact of 5-17-year-olds returning to school between September and November 2020 (study rounds 5-to-7) is also clearly visible (Figure 8.2); compared to 18-34-year-olds, the ORs were between 0.34 (95% CI 0.32, 0.35) and 0.64 (95% CI 0.61, 0.67) for having no contacts outside the home over this period. These ORs were largely unaffected by further adjustment for being a self-reported suspected or confirmed COVID-19 case in the preceding two weeks (Supplementary Section D.3).

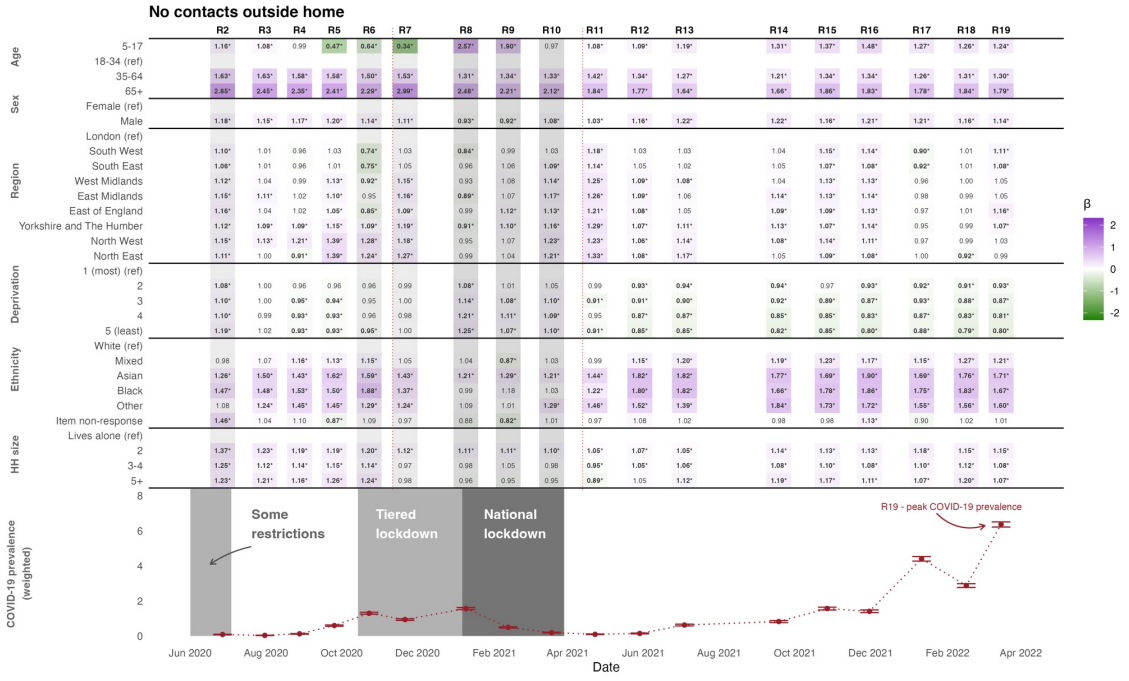

Figure 8.2: Odds ratios for whether an individual self-reports having no contacts outside the home in the preceding seven days. Purple shading indicates odds ratios greater than 1 (compared to the reference group) while green shading indicates odds ratios less than 1. Darker shading indicates estimates further from 1. Reference groups are indicated with (ref) in the row-labels. Vertical dotted red lines denote changes in question wording. Odds ratios that are statistically significantly different to 1 at $p < 0.05$ are denoted with an asterisk and set in bold face. The bottom panel shows weighted prevalence of SARS-CoV-2 swab positivity in each round of REACT-1.

Figure 8.2 also suggests that men were more likely than women to report having no contacts

outside the home, except in January and February 2021, while those living with other people were more likely to report having no contacts outside the home, except between November 2020 and April 2021. Further results regarding the average total number of contacts individuals had outside their home are provided in Supplementary Section D.2.2 (Supplementary Figure D.5). Results from logistic regression models on additional outcomes, including specific reasons for leaving the home, are included in Supplementary Section D.3 (Supplementary Figures D.17 to D.27). The same results after controlling for infection status are reported in Supplementary Figures D.28 to D.39.

We also consider responses to these behavioural questions conditional on self-reported risk, vaccination status, and infection history (Figure 8.3). Those that considered themselves to be at risk of severe illness were more likely to report shielding, as were those that reported having been vaccinated, and those who had not reported previous infection. When the question was first asked, there was high self-reported use of face coverings, although over time those that did not consider themselves at severe risk and those that reported no vaccination became relatively less likely to report wearing a face covering. These results are elaborated on in Supplementary Section D.2.4 (Supplementary Figures D.12 to D.15).

8.3.2 Comparisons with mobility data

Despite the limited duration of each period considered (19 June 2020 - 11 November 2020, 4 January 2021 - 28 September 2021, and 19 October 2021 - 31 March 2022), substantial variation in mobility and behaviours was observed within each period. The random forest models of the mobility series capture key trends while missing some finer day-to-day variations (Figure 8.4). The out-of-bag proportion of variance explained (PVE) (Supplementary Table D.8) was high for all six mobility indices (ranging from 86% to 97%) for data from study rounds 8-to-14 and was lower for rounds 15-to-19 (ranging from 68% to 86%). Additional results show that shopping was consistently the main reason for leaving the home, while the proportion of individuals reporting errands and work as a reason for leaving the home increased from January 2021 onwards (Supplementary Section D.2.5, Supplementary Figure D.16). Correlation coefficients between the self-reported behaviours and mobility series are reported in Supplementary Section D.4 (Supplementary Table D.7).

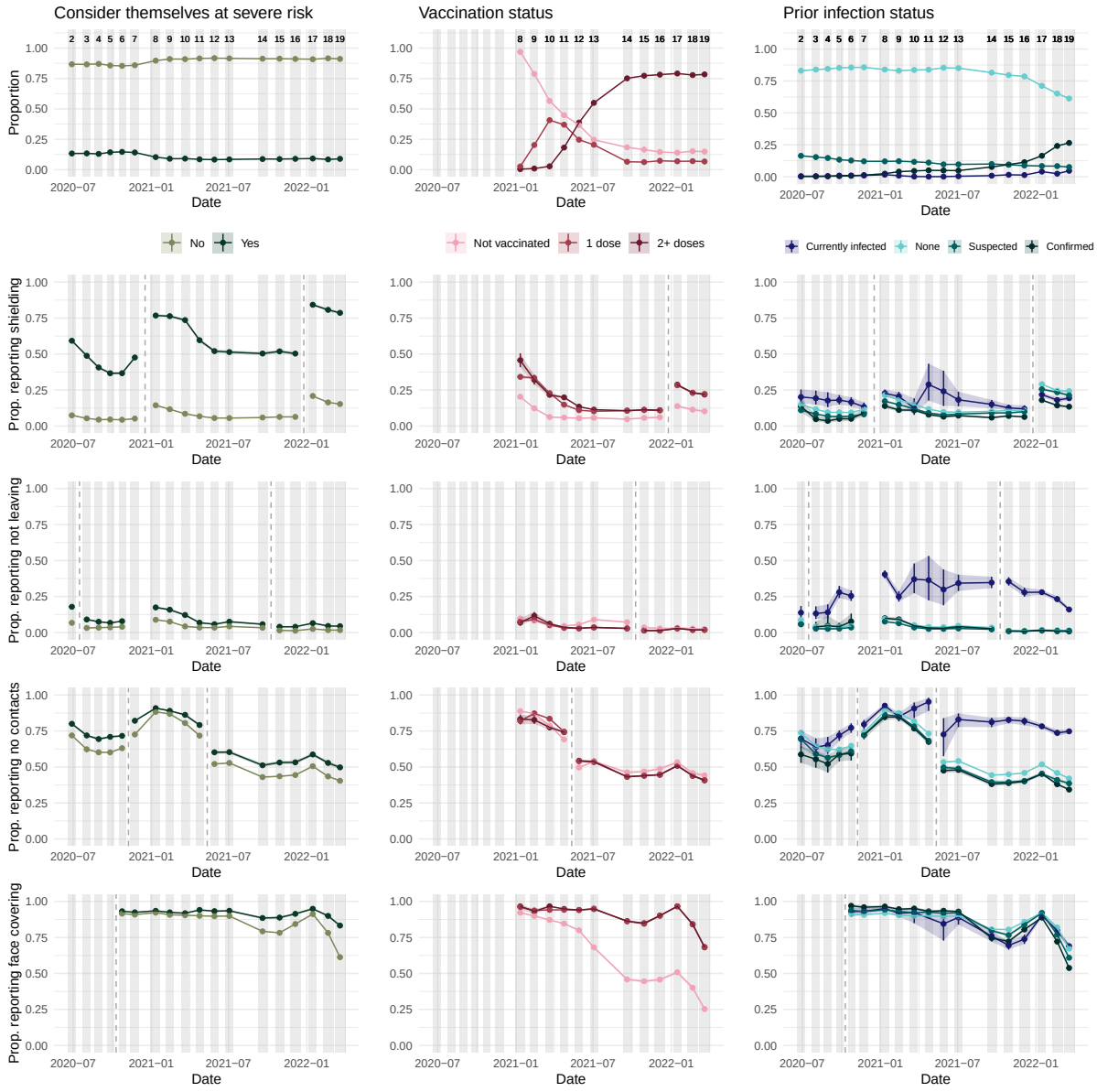

Figure 8.3: The proportion of people in England self-reporting considering themselves at severe risk, having no/one/two-plus doses of a vaccine, and having no/a suspected/a confirmed prior infection or current infection (top row). Rows 2-5 then present responses to the four focus questions, disaggregated by these responses. Vertical dashed lines denote changes in question wording. Study round numbers are provided at the top of row 1 and shaded regions indicate periods when the study was actively collecting data.

8.3.3 Comparisons with policy response indices

Results from the random forest models for the two policy response indices demonstrate a strong association between self-reported behavioural data and the stringency indices. The out-of-bag PVE was 0.98 for both models demonstrating a strong association between self-reported behaviours and aggregated measures of the stringency of public health policy.

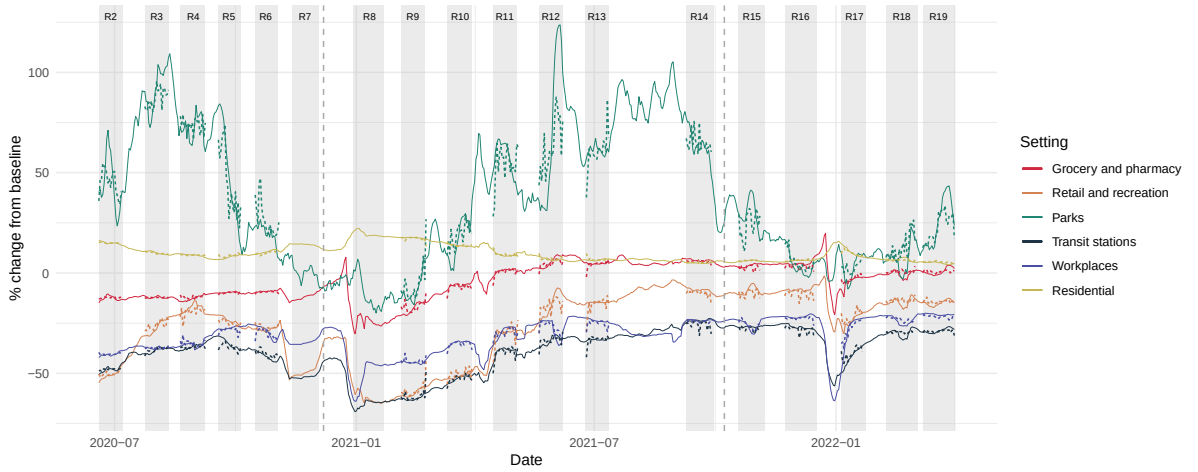

Figure 8.4: The Google mobility data (solid lines) between June 2020 and March 2022 and the random forest model predictions based on behaviours reported in REACT-1 questionnaires (dotted lines). Modelled values shown are out-of-bag predictions, so the model has not been trained on the data-point being predicted. The vertical dashed lines demarcate individual model fits, selected to account for substantial changes in question wording. We do not fit a model to data from study round 7 as the relevant questions were not asked.

The fitted index values based on behavioural data also identified the timing of when the change from baseline in the stringency index switched from being above to below that for the containment and health index, which occurred near the start of study round 11 (Figure 8.5).

For comparison, we also fit a logistic regression model to the policy indices using the Google mobility data as the predictors (Supplementary Section D.5, Supplementary Figure D.40). Responses to the four risk-related behavioural questions from the REACT-1 study explained more variance in the policy index than did the six Google mobility variates (Supplementary Table D.8), with an out-of-bag PVE (using the Google data) of 0.91 for the stringency index and 0.90 for the containment and health index (compared to the out-of-bag PVE of 0.99 for the model leveraging REACT-1 data), despite the Google mobility series using more covariates.

8.4 Discussion

The COVID-19 pandemic was a major shock to health systems, requiring near-unprecedented responses from governments and individuals to protect individual and public health. Government response in England and many other countries included imposition of lockdowns and other containment measures, with legal restrictions on population mobility and behaviours. We

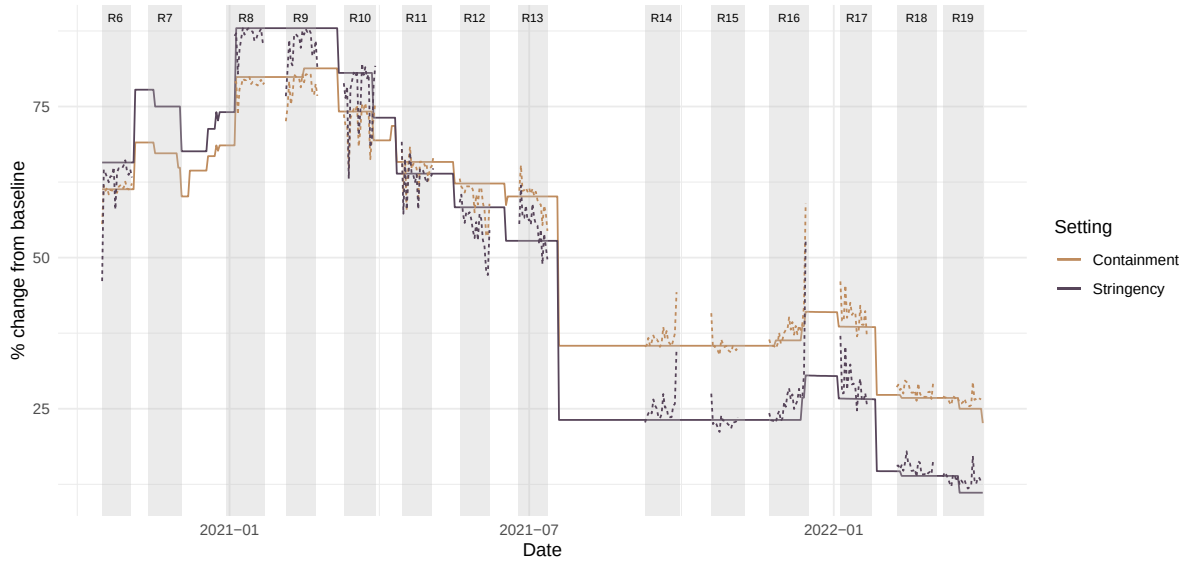

Figure 8.5: The OxCGRT stringency index and containment and health index (solid lines) between October 2020 and March 2022 and the model predictions based on behaviours reported in REACT-1 questionnaires (dotted lines). We do not fit the model to data from study round 7 as the relevant question was not asked in this round.

show strong evidence that individuals substantially changed their behaviours over time, with behaviours to reduce risk of infection closely reflecting these policy shifts during the pandemic.

Within the period considered (19 June 2020 - 31 March 2022), protective behaviours were most pronounced in January 2021, during the winter wave of COVID-19 in England, when very few people had been vaccinated. Over the following months, vaccination rates increased while the prevalence of infection decreased, and the proportion of people that reported protective behaviours decreased.

Infection prevalence sharply increased over December 2021 and January 2022 due to the Omicron BA.1 variant [390] accompanied by an uptick in protective behaviours, although not to the previous levels seen in January 2021 during the second national lockdown. This was then followed by a second wave during February and March 2022 due to the Omicron BA.2 variant, resulting in the highest recorded SARS-CoV-2 prevalence in REACT [49]. However, at the same time there was a fall in protective behaviours, indicating possible "pandemic fatigue". This is particularly noticeable in the wearing of face coverings, which decreased rapidly over February and March 2022 to an all-time low, after being at relatively constant levels since face covering wearing was first recommended in July 2020.

We found important differences between demographic groups. For example, those living in larger households were generally less likely to report shielding, but were more likely to report not leaving the home. There were also large differences by age, and at times, by sex, ethnicity, and region. Understanding potential differences in behaviours by demographic groups is critical for planning and implementing NPIs [397].

Our results show that reported behaviours and reasons for leaving the home were strongly correlated with the OxCGRT stringency index and the OxCGRT containment and health index, and the publicly available Google mobility data, demonstrating the potential value of both the Google mobility data and the OxCGRT indices as population-level proxies for social-distancing behaviours. These are complementary to individual-based self-reported variables enabling the investigation of potential determinants of behavioural changes throughout the pandemic. However, there is no guarantee these relationships would generalise outside the time periods considered, particularly as signs of pandemic fatigue in later rounds could decrease the correlation between self-reported behaviours and the OxCGRT indices. We also emphasise that our analyses do not indicate whether policy changes are leading or lagging indicators of behavioural change, a key issue in preparing for any future pandemic.

8.4.1 Limitations

Interpretation of our data is dependent on the accuracy of the self-reported behaviours. Previous studies have indicated that behaviour-related signals may be drowned out by bias and noise in self-report data [398], and adherence to protective behaviours may be substantially overestimated in self-reported data both historically [399], and during the COVID-19 pandemic [400]. While the strong associations between our data and aggregated community mobility data and policy indices lend some contextual validity to our findings, we cannot exclude biases in reporting, for example, across demographic groups. We made changes to the questionnaire over the nearly two years of the study, in some cases to reflect the evolving situation and terminology such as “shielding”, “support bubbles”, and the tier system of restrictions, which meant that participant responses over time were not always directly comparable. Additionally, the design of the study involving repeated cross-sectional sampling per round [48] meant that the observed trends reflect changes across representative samples of the population, and in this analysis we did not investigate how an individual’s behaviour changed over time.

8.4.2 Conclusion

Despite speculation that populations such as England would not be willing to change their behaviours in response to an infectious disease threat [401], our study, along with others [402, 403] show that this is patently not the case. We show that the behavioural response closely followed the implementation of government restrictions, but these large behavioural changes may not be replicated in future events, particularly given the strong evidence of "pandemic fatigue" at the end of the study in March 2022. COVID-19 shaped and strengthened many people's views about the effectiveness of such measures [404], and with multiple pandemic-potential infectious diseases being monitored (such as H5N1 [405] and Mpox [406]), understanding people's behavioural decisions, notably though self-reported data, remains a critical component of future pandemic preparedness.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor (only required where there isn't already a statement of contribution within the paper itself).

Title of Paper	Pandemic-risk-related behaviour change in England from June 2020 to March 2022: REACT-1 study among over 2 million people
Publication Status	☐ Published ☐ Accepted for Publication ☒ Submitted for Publication ☐ Unpublished and unsubmitted work written in a manuscript style
Publication Details	Steyn, N., Chadeau-Hyam, M., Whitaker, M., et al. (2025). Pandemic-risk-related behaviour change in England from June 2020 to March 2022: REACT-1 study among over 2 million people. Submitted for publication.

Student Confirmation

Student Name	Nicholas Steyn
Contribution to the Paper	Conceptualisation, methodology, investigation, software, visualisation, data curation, validation, writing (original draft), and writing (review and editing).

Signature NSteyn	Date 22/09/2025

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description above is accurate.

Supervisor name and title	Professor Christl A. Donnelly
Supervisor comments	

Signature Christl Ann Donnelly	Date 22/09/2025

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 9

Revisiting Chapters 4, 5, and 6 through the lens of Chapter 3

This thesis is structured approximately in reverse chronological order. While uncertainty is a focus throughout, lessons from later chapters inform earlier chapters. Here, we briefly revisit aspects of later chapters through the uncertainty framework introduced in Chapter 3. For consistency with Chapter 4, we use the names "EpiEstim" and "EpiFilter" to refer to the R_t estimators described in [114] and [115], respectively.

9.1 Uncertainty in the Poisson renewal model (Chapter 3)

EpiEstim [114] and EpiFilter [115] estimate R_t using the Poisson renewal model, defined as:

$$C_t \sim \text{Poisson}\left(R_t \Lambda_t\right), \quad \text{where } \Lambda_t = \sum_{u=1}^{t} C_{t-u}\omega_u, \tag{9.1}$$

C_t is the number of reported cases at time t, R_t is the reproduction number, and ω is the probability mass function of the serial interval. Both models assume the infection pressure Λ_t at each time step is known.

Consider estimating R_t at time t. The Bayesian approach places a prior distribution on R_t and updates it with the likelihood (Equation 9.1). Using a conjugate Gamma prior distribution and conditioning on past cases $C_{1:t-1}$ (equivalent to EpiEstim with no smoothing), the posterior

distribution for R_t is:

$$p(R_t|C_{1:t}) \propto p(C_t|R_t, C_{1:t-1}) \cdot p(R_t|C_{1:t-1}) \tag{9.2}$$

$$\propto \underbrace{(R_t\Lambda_t)^{C_t}e^{-R_t\Lambda_t}}_{\text{Likelihood (renewal model)}} \cdot \underbrace{R_t^{\alpha-1}e^{-\beta R_t}}_{\text{Prior distribution}} \tag{9.3}$$

$$\sim \text{Gamma}\left(\alpha + C_t, \beta + \Lambda_t\right). \tag{9.4}$$

This model assumes no uncertainty flows from the infection pressure Λ_t because both the historical case trajectory $C_{1:t-1}$ and serial interval ω are treated as known. This contradicts the intuition that uncertainty should be accounted for in past (as well as current) cases, an intuition that likely arises because, in reality, infections generate further infections, which are then reported as cases. In this model, cases directly generate cases. Once observed, the C_t are treated as a ground truth and are not a source of uncertainty. This is a fundamental assumption of the model rather than a reflection of reality.

The models employed by EpiEstim and EpiFilter differ solely in their approach to smoothing[1]. By default, both feature fixed irreducible uncertainty and no reducible uncertainty. In Chapter 4, we introduced estimators for their smoothing parameters (k and η), extending the model's joint posterior distributions to include these parameters. Estimating these parameters substantially improves the quality of fit, due to (a) improved parameter selection, and (b) the fact that marginalising out nuisance parameters better accounts for overall uncertainty.

The framework in Chapter 3 helps conceptualise why this estimation and marginalisation helps: (i) it allows the model to select an appropriate level of *irreducible* uncertainty by estimating the value of k/η, and (ii) it introduces a new source of *reducible* uncertainty by propagating uncertainty about these parameters. The decision-theoretic definition of uncertainty in Chapter 3 enables explicit decomposition of these sources.

9.1.1 Example: EpiFilter with η estimation

Consider the *smoothing* (forward- and backward-looking) version of EpiFilter which assumes that R_t follows a random walk, and targets marginal posterior distributions $p(R_t|C_{1:T})$. We examine the following questions:

[1] A range of adaptations to these models have been developed. We refer here to the original models as published in [114] and [115], respectively.

1. How much uncertainty about η is irreducible?
2. Is uncertainty about R_t reduced by learning about η?
3. Is marginalising out η necessary, or does a point estimate suffice?

We take a time-series view of the missing data, treating y_{miss} as the (potentially infinite) values of $C_{T+1:\infty}$ that could be observed. Knowing these data would improve estimation of η, and therefore of R_t. We consider the same dataset used in Chapter 4: reported cases of COVID-19 in Aotearoa New Zealand between 17 August 2021 and 14 February 2022.

To sample the missing data, we first fit the model to the observed cases $C_{1:T}$. We then sample $(R_T^{(i)}, \eta^{(i)}) \sim p(R_T, \eta | C_{1:T})$ from the joint posterior distribution at time T, and iteratively simulate $R_{t+1}^{(i)} \sim p(R_{t+1} | R_t^{(i)}, \eta^{(i)})$ and $C_{t+1}^{(i)} \sim \text{Poisson}(R_{t+1}^{(i)} \Lambda_{t+1}^{(i)})$ for $t = T+1, \ldots, T^*$, appending the sampled cases to the observed cases. To ensure the plausibility of the simulated data, we restrict $\max C_t \leq 10^6$ (one fifth of the population of New Zealand), resampling the entire trajectory if violated. We refer to this as Scenario A. Examples of simulated trajectories are shown in Figure 9.1-A.

Most trajectories in Scenario A result in rapid extinction, an implication of the chosen model. To see this, note that as R_t follows a random walk and cases follow the Poisson renewal model, the modelled case trajectory either goes extinct or explodes. By forcing $\max C_t \leq 10^6$, we rule out the explosive trajectories (which are also the most informative about η), leaving only those that go extinct.

To examine how the uncertainty reduction changes with longer-running epidemics, we also consider the scenario where $C_{t+1}^{(i)}$ are sampled from a Poisson distribution truncated to values ≥ 10. The same maximum-case restriction is imposed. We refer to this as Scenario B. Examples of simulated trajectories are shown in Figure 9.1-B.

The grid-based approximations employed by EpiFilter can be viewed as enforcing discrete prior distributions on R_t and η, so we directly calculate measures of uncertainty from the grid-based posterior distributions. Focusing on probabilistic modelling, we use the log-loss function and therefore entropy as our measure of uncertainty. As we cannot simulate an infinite quantity of

data, we truncate the missing data at $T^* = T + 500$ days for Scenario A and $T^* = T + 1000$ days for Scenario B. In Scenario A, all trajectories go extinct by T^*.

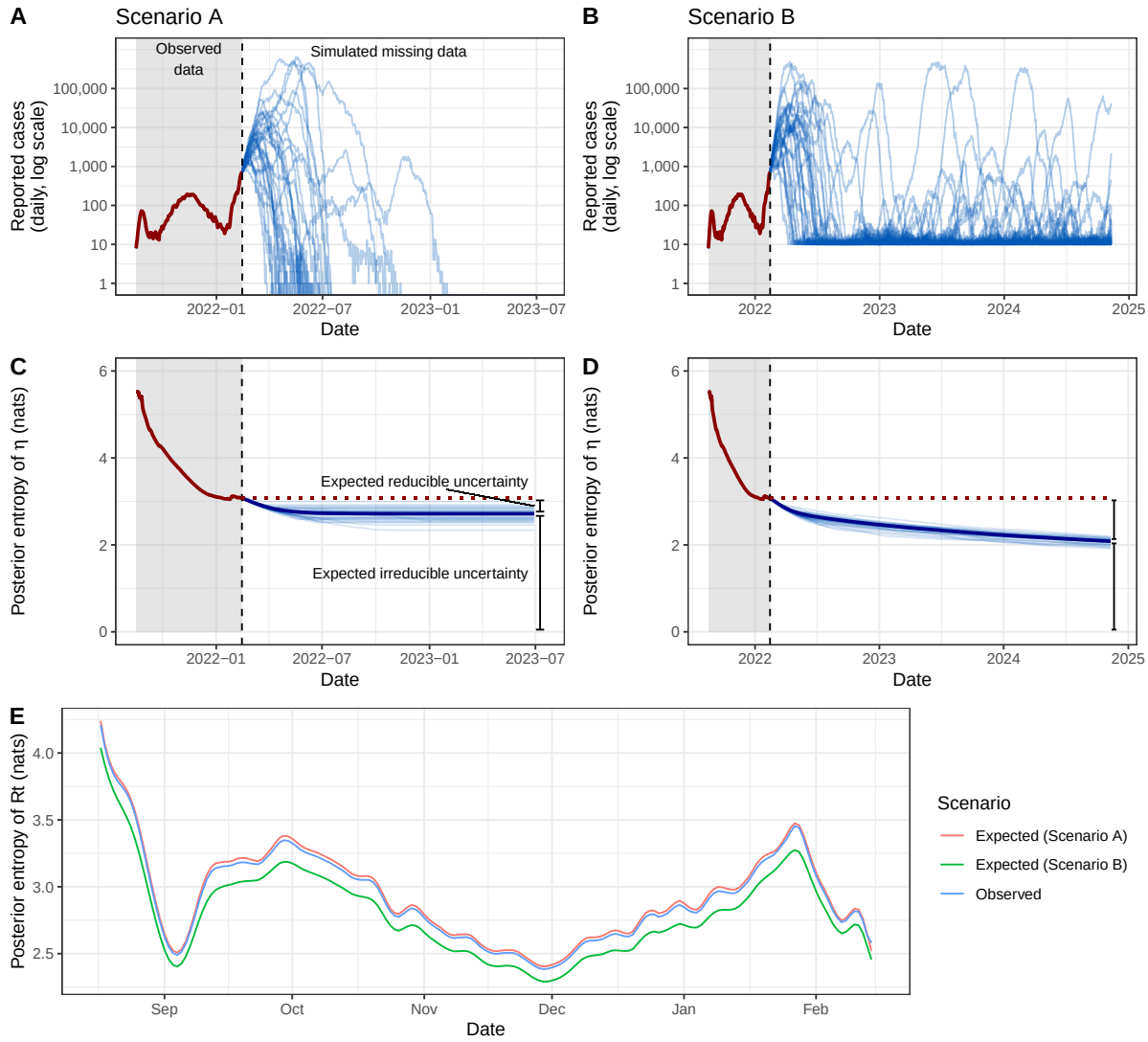

Figure 9.1: Panels A and B show the observed case trajectory (red) and 30 simulated case trajectories (blue) for Scenario A and Scenario B respectively. Panels C (Scenario A) and D (Scenario B) show the posterior entropy of the posterior distribution of η fit to data up to time t. Thin blue lines show the posterior entropy from fits to individual simulations, while the thick blue line shows the average posterior entropy. The dotted red line projects the posterior entropy of η at the final observed time step. The gap between the dotted red line and thick blue line at T^* is the expected uncertainty reduction about η. Panel E shows the daily posterior entropy of R_t for Scenario A (red), Scenario B (green), and the observed posterior entropy (blue).

Question 1: How much uncertainty about η is irreducible?

The posterior entropy of η over time is shown in Figure 9.1-C (for Scenario A) and Figure 9.1-D (for Scenario B). In theory, if the epidemic continued indefinitely, all model-based uncertainty

about η would be reducible. However, because trajectories in Scenario A go extinct quickly, the information about η is limited. The expected uncertainty reduction about η is small (0.36 nats, where a nat is the natural-logarithm unit of information) relative to total uncertainty (3.07 nats). The expected uncertainty reduction is greater in Scenario B (0.99 nats), as the epidemic is forced to continue for longer, although a very large increase in T^* would be required to reduce uncertainty about η to near zero. Even the *observed* uncertainty about η flattens out over time, suggesting that after a certain point (in this example we observe 182 days of reported cases), there is little benefit to collecting more data, if the goal is to learn about η.

Question 2: Can we reduce uncertainty about R_t by learning about η?

As η determines the smoothness of R_t, the absolute value of η partly determines the irreducible uncertainty in R_t (small η implies that R_t changes slowly, so less irreducible uncertainty is introduced at each time step). Indeed, this is the primary motivation behind the estimation procedure developed in Chapter 4. In addition to optimising the value of η, uncertainty about η itself propagates to R_t uncertainty via marginalisation. All else held equal, decreasing uncertainty about η should reduce uncertainty about R_t. We present the observed daily posterior entropy of R_t in Figure 9.1-E, alongside the expected daily posterior entropy under Scenarios A and B.

The $\max C_t \leq 10^6$ restriction (in both scenarios) and truncation of $C_t \geq 10$ (in Scenario B), mean that the predictive models are not coherent with respect to the fitted model, so the expected uncertainty reduction is not guaranteed to be positive. Under a coherent model, we should observe $E[\eta|C_{1:T}] = E_{C_{T+1:\infty}}[E[\eta|C_{1:\infty}]]$. That is, our central estimate of η should not change by including simulated missing data (intuitively, our simulated data should not introduce any bias in our estimate of η). As seen below, this is not the case here.

In Scenario A, the expected uncertainty reduction is slightly *negative*, that is, uncertainty about R_t (for $t < T$) increases as the missing data are collected. This occurs because the $\max C_t \leq 10^6$ restriction rules out the case trajectories that are most informative about η (those with larger values of C_t) and also results in simulated trajectories eliminating earlier on average, thus providing less information about η than expected. This prevents the model from "ruling out" higher values of η, thus resulting in greater uncertainty about R_t. Scenario A provides a clear

example of how a non-coherent predictive model, where implicit assumptions about the missing data do not correspond to the realised missing data, can lead to counterintuitive results.

In Scenario B, the expected uncertainty reduction is *positive*. At least part of this reduction can be attributed to the truncation of the predictive distribution for C_t ensuring that simulated trajectories continue indefinitely (or, at least until T^*), thus decreasing uncertainty about η relative to Scenario A. However, by forcing C_t to be at least 10, we also ensure that R_t appears to change less abruptly. The average posterior mean of η after fitting to the simulated data in Scenario B is 0.098, compared to 0.122 when fitting to only the observed data, which also contributes to the reduction in uncertainty about R_t.

Question 3: Is marginalising out η necessary, or does a point estimate suffice?

Uncertainty about η has less of an impact on our uncertainty about R_t than the central estimate of η does. We can formalise this using mutual information, stated in our context as:

$$\underbrace{H(R_t)}_{\text{Entropy}} = \underbrace{H(R_t|\eta)}_{\text{Conditional entropy}} + \underbrace{I(R_t;\eta)}_{\text{Mutual information}} .$$

The conditional entropy is estimated by sampling $\eta' \sim p(\eta|C_{1:T})$, fitting the model to the observed $C_{1:T}$ conditional on η', calculating the posterior entropy of R_t and repeating this procedure and averaging the entropies. The mutual information (a Kullback-Leibler divergence) is then the difference of the total and conditional entropies, and describes the excess "surprise" from modelling (R_t, η) using their independent marginal distributions, rather than their joint distribution.

Total entropy $H(R_t)$ ranges between 2.38 nats and 4.21 nats, with an average of 2.92 nats. The estimated conditional entropy $H(R_t|\eta)$ ranges between 2.38 nats and 4.19 nats, with an average of 2.91 nats. That is, the vast majority of R_t entropy comes from the conditional entropy (a function of the central value of η), with very little coming from the mutual information.

This provides strong evidence that, in this example, using a point estimate of η would suffice when quantifying the level of uncertainty about R_t. This has practical implications, as estimating the value of η (or k in EpiEstim) is easier than marginalising it out. Estimation of the optimal k in the Microsoft Excel-based implementation of EpiEstim is feasible, while marginalisation is more complex, making our finding particularly relevant in this setting. Investigating

whether this finding generalises to other datasets and models is a potential avenue for future work.

9.2 Using SMC methods to quantify uncertainty in a renewal model with latent infections (Chapter 5)

We have thus far only considered the simple renewal model for reported cases. If we want to fully account for uncertainty associated with the reported case trajectory, we should explicitly model latent infections and the observation process. There are many possible sources of reducible uncertainty in this context. For this example, we focus on uncertainty arising from case underreporting, although other sources of observation noise (e.g., reporting delays, false positive/negative tests, and day-of-week effects [36, 205, 293]) could also be considered.

We consider the following model:

$$\begin{aligned} \log R_t &\sim \text{Normal}(\log R_{t-1}, \sigma), \\ I_t &\sim \text{Poisson}\left(R_t \sum_{u=1}^{t-1} I_{t-u}\omega_u\right), \\ C_t &\sim \text{Binomial}(I_t, \rho), \end{aligned}$$

where I_t is the latent infection incidence which we only fully observe when $\rho = 1$. For demonstration purposes, we assume that $\sigma = 0.1$ and $\rho = 0.2$ are known parameters. We fit this model using SMC methods (Chapter 5) to the same New Zealand case data as in the previous section (COVID-19 cases reported between 17 August 2021 and 14 February 2022).

To estimate the expected uncertainty reduction associated with increasing the case ascertainment ratio ρ, we fit the model to the observed data, sample $C_{1:T}^{(i,\rho^*)}$ from the posterior predictive distribution (Chapter 5, Section 5), refit the model to these data at the new value of $\rho = \rho^*$, calculate the posterior variance of R_t, and repeat this procedure 100 times. We consider values of $\rho^* = 0.4, 0.6, 0.8,$ and 1.0. By measuring uncertainty with the posterior variance, we are implicitly targeting a point estimate of R_t (the posterior mean) and have a quadratic loss function.

Figure 9.2-A shows the observed posterior mean and 95% credible intervals of R_t over time for the observed data (red) and expected observed data under perfect reporting ($\rho^* = 1$, blue).

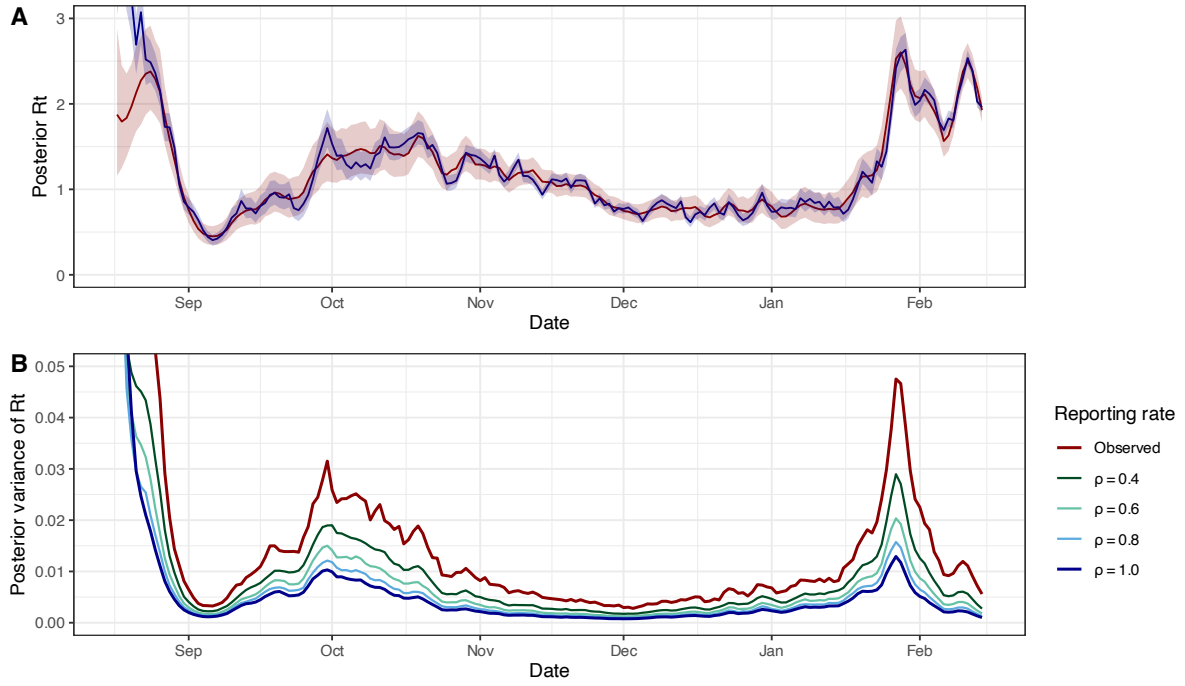

Figure 9.2: Panel A shows the posterior mean and 95% credible intervals of R_t over time, fit to observed case data (red) and fit to expected observed case data under perfect reporting (blue). Panel B shows the posterior variance of R_t over time for the observed data (red) and expected posterior variance under increasing values of ρ (green to blue). The dark blue line represents the irreducible uncertainty about R_t in this model, while the gap between the red and blue lines represents the expected reducible uncertainty.

The specific simulated dataset used for the expected observed data was chosen by finding the simulated dataset that minimised the squared difference in daily posterior variance between the simulation and the average posterior variance of all simulations. That is, the blue curve represents a "typical" posterior mean and 95% credible interval for R_t under perfect reporting. Coherence between the fitted and predictive models ensures that the credible intervals of the expected posterior distribution are generally contained within those of the observed posterior distribution, ignoring pathological behaviour during the initial time steps (where there are insufficient past cases to reliably estimate Λ_t).

The expected uncertainty reduction associated with increasing ρ is also shown in Figure 9.2-B. In this example, a considerable reduction in variance can be expected to be obtained by improving case reporting. To determine whether it is worth investing in improved case reporting, the loss function could be expanded to include an economic cost of uncertainty in R_t. By comparing this with the economic cost of improving case reporting, the underlying model can

be directly connected to decision-making.

This modelling approach has the added advantage of allowing for uncertainty in the infection pressure Λ_t over time. Λ_t is an interesting quantity in its own right, as it describes the total infectiousness of the population at time t. For example, while there are many definitions of the elimination of an epidemic (e.g., no reported cases for two mean generation intervals) [118, 252, 270], a natural definition is $\Lambda_t = 0$. Knowing when elimination has been achieved is crucial in public health decision-making, particularly for considering the relaxation of expensive control measures. Future work could consider how uncertainty about Λ_t can be controlled and reduced, and therefore how to best design surveillance systems to reliably detect elimination, ideally minimising the duration of costly interventions.

9.3 Data collection strategies for epidemic prevalence surveys (Chapter 6)

For this example, we apply our framework to the SIMPLE model of epidemic prevalence surveys introduced in Chapter 6. We first fit the model (of the growth rate with a beta-binomial observation distribution) to publicly available data from the REal-time Assessment of Community Transmission-1 (REACT-1) study between 30 December 2020 and 21 January 2021, stratified by the nine regions of England. We then assume a hypothetical scenario in which 360 daily tests are available over the following seven days, and consider how to best allocate these tests. In addition to an equal allocation (40 tests per region per day), we consider combinations of lower tests (20 tests per region per day) and higher tests (50, 56, 65, 80, 110, and 200 tests per region per day, depending on how many regions receive the lower and higher allocations).

As each region is modelled independently, the total entropy is the sum of the entropy from each region:

$$H\left[p(z_{1,t}, \ldots, z_{9,t}|y_{obs})\right] = \sum_{i=1}^{9} H\left[p(z_{i,t}|y_{obs})\right], \tag{9.5}$$

where $z_{i,t}$ represents either the prevalence (P_t) or growth rate in prevalence (r_t) of SARS-CoV-2 in region i on day t after the conclusion of the observed data.

The missing data are the results of the hypothetical future tests, $y_{miss} = n^+_{i,t}$. We sample these by first sampling initial values from the fitted model $(r_{0,i}, P_{0,i}, \theta_i) \sim p(r_0, P_0, \theta|y_{obs})$ (where

time step 0 is the final time step of the observed data) and then simulating the model forward, iteratively sampling $r_t \sim p(r_t|r_{t-1}, \theta_i)$ and setting $P_t = P_{t-1}\exp(r_t)$. By fitting the model to $\{y_{obs}, y_{miss}\}$, calculating the associated entropy, and repeating the procedure 50 times, we obtain a Monte Carlo estimate of the expected uncertainty reduction associated with a specific test allocation.

At the final time step of the observed data, we find that the North East has the greatest uncertainty about both r_t and P_t (Figure 9.3). While the South West has the second-greatest uncertainty about r_t, it has the lowest uncertainty about P_t. Similarly, while London has the second-greatest uncertainty about P_t, it has the second-lowest uncertainty about r_t. This inverse relationship between uncertainty about r_t and P_t holds in general. Regions with greater prevalence will return more positive tests, allowing for more accurate estimates of r_t, but less accurate estimates of P_t in absolute terms (simply because the variance of the beta-binomial distribution increases with P_t for $P_t < 0.5$). The North East breaks this trend as it has a higher estimated level of overdispersion and returned fewer tests than the other regions, so features greater uncertainty about both P_t and r_t.

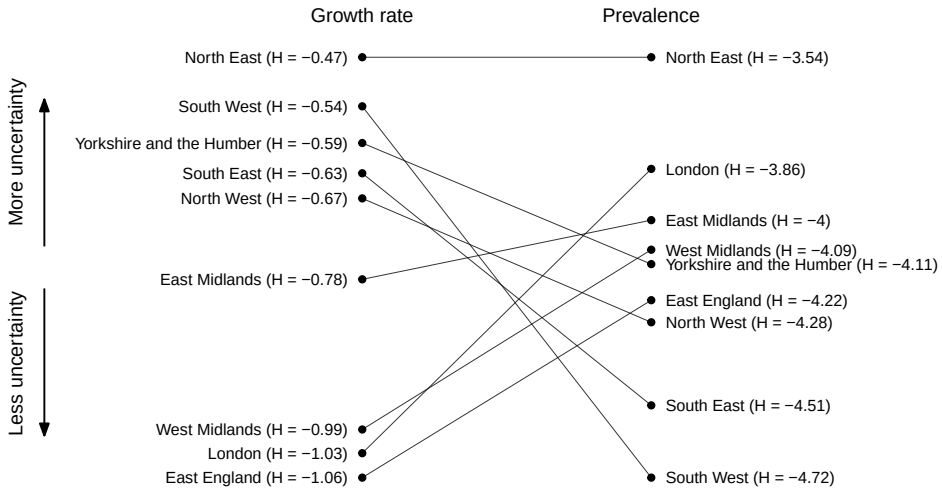

Figure 9.3: Uncertainty (posterior entropy) about the growth rate (left) and prevalence (right) of SARS-CoV-2 by region in England on 21 January 2021, from fitting the SIMPLE model to REACT-1 data between 30 December 2020 and 21 January 2021. NE denotes the North East region and YH denotes Yorkshire and The Humber.

Short-term expected uncertainty about P_t is minimised by prioritising testing in the North East and Yorkshire and The Humber (Figure 9.4), regions with larger uncertainty about the growth rate. This strategy is optimal as greater uncertainty about r_t today translates into greater

uncertainty about P_t tomorrow. Longer-term expected uncertainty about P_t is minimised by an equal allocation of tests.

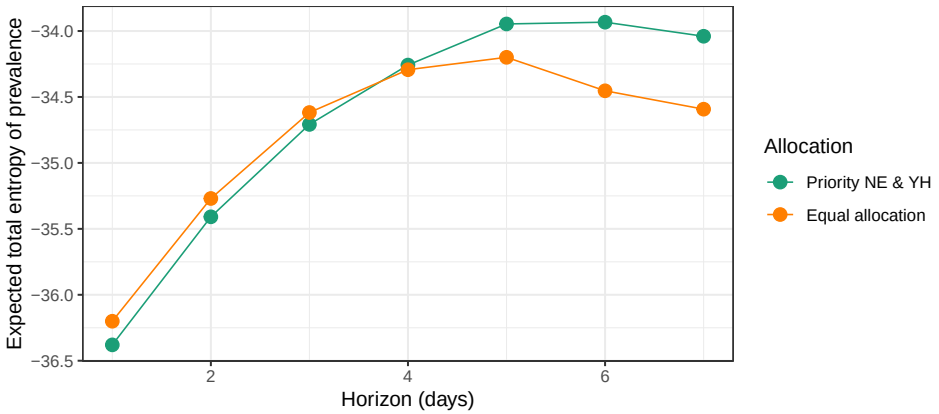

Figure 9.4: Total expected uncertainty about prevalence of SARS-CoV-2 in England under different regional test allocation strategies. The total expected uncertainty is shown for 1, 2, ..., 7 time steps ahead, assuming the allocation is the same on each day. When two/five regions have priority, they receive 110/56 tests per day each with 20 tests per day for the other regions.

9.4 Conclusion

This chapter presented examples of how the uncertainty framework from Chapter 3 can be applied to different scenarios. These applications improved understanding of model behaviour, justified the use of simpler modelling techniques (using a point estimate of k instead of marginalisation), quantified the benefits of improved data collection (through increasing the case ascertainment ratio), and informed practical data collection strategies (test allocation in an epidemic prevalence survey). Overall, these examples demonstrate how a unified treatment of uncertainty can sharpen inference, clarify trade-offs, and guide study design.

Chapter 10

Discussion

As Chapters 3-to-8 feature standalone papers, each chapter contains its own discussion section. We therefore provide a single-paragraph summary of each chapter here, and then focus on a discussion of select overarching themes. We conclude with a reflection on the limitations of this thesis as a whole and directions for future work.

10.1 Summary of key contributions

Chapter 3 develops a conceptual framework for quantifying uncertainty in epidemiological models. By taking a first-principles decision-theoretic approach, an otherwise-ethereal notion of uncertainty is tied directly to real-world decision-making. We use this to separate model-based uncertainty into irreducible and reducible components, giving a mathematically precise decomposition of the total uncertainty in a model. Most importantly, this framework provides a principled basis on which more advanced methods for uncertainty quantification can be developed.

Chapter 4 focuses on improving the uncertainty quantification in two popular methods for estimating the instantaneous reproduction number (R_t): EpiEstim and EpiFilter. We show that estimating and marginalising the smoothing parameters in each model substantially improves calibration of uncertainty intervals, while also improving point-estimate accuracy. This provides a simple way to improve the robustness of inferences for public health decision-making, without requiring substantial changes to existing workflows, as many practitioners are already familiar with these methods. Julia implementations of both methods are provided here: https://nicsteyn2.github.io/RobustRtEstimators/.

Chapter 5 develops methods for fitting hidden-state-based renewal models using sequential Monte Carlo (SMC) and particle marginal Metropolis-Hastings (PMMH). This creates a toolkit for researchers to quickly and easily construct a wide range of models from scratch, ensuring they fully understand the underlying assumptions of their models, while facilitating rapid testing and comparison of these models. By coupling SMC with PMMH, we also enable the joint estimation of both static parameters and latent states, allowing for comprehensive uncertainty quantification. Extensive online material is provided to facilitate uptake of these methods, which can be found here: https://nicsteyn2.github.io/SMCforRt/.

Chapter 6 presents a novel approach for smoothing and performing inference on epidemic prevalence survey data, comparing this with two existing approaches. We highlight how minor changes in model structure, such as allowing for observation overdispersion, can have substantial implications for inference. This chapter also presents the two existing approaches (which were developed for specific applications) in a general setting using common notation, allowing for direct comparison of their assumptions and implications, and improving the accessibility of all three approaches.

Chapter 7 constructs a joint model for SARS-CoV-2 wastewater and case data in Aotearoa New Zealand and fits it using the SMC/PMMH framework from Chapter 5. This work highlights the value that wastewater data can bring to epidemiological modelling, particularly in compensating for biases in traditional passive surveillance data. By simultaneously modelling both data streams, we can also estimate the relative case ascertainment ratio over time, providing a valuable tool for public health decision-making answering the question: are reported cases decreasing because of reduced transmission, or reduced reporting?

Chapter 8 uses data from the large-scale REal-time Assessment of Community Transmission-1 (REACT-1) study to estimate the prevalence of self-reported behaviours in England during the COVID-19 pandemic. This provides the largest-sample estimates of risk-related behaviours during the pandemic in England. We also use these outputs to explore relationships with population-level mobility datasets, justifying their use as proxies for behaviour in epidemiological models. Finally, we make extensive anonymised data available to facilitate further research, which can be found here: https://github.com/nicsteyn2/REACTBehaviouralData.

Finally, **Chapter 9** revisits the models from Chapters 4, 5, and 6 through the lens of Chapter 3. We provide further insight into why the estimation and marginalisation of smoothing parameters improves uncertainty quantification in Chapter 4, demonstrate how the SMC/PMMH framework from Chapter 5 can be used to extend uncertainty quantification in the renewal model to latent infections, and explore survey sampling strategies to reduce uncertainty in estimates made using epidemic prevalence survey data in models from Chapter 6.

10.2 Discussion topics

10.2.1 Uncertainty quantification in epidemiological modelling

Uncertainty is a recurring theme throughout this thesis. This is most explicit in Chapter 3, in which we provide a definition of uncertainty and discuss specific ways of conceptualising it in epidemiological modelling. However, uncertainty is also a central consideration in other chapters.

While many modern epidemiological methods quantify uncertainty within the Bayesian paradigm, the foundational training of public health practitioners has traditionally emphasised frequentist tools [407]. The notion of calibration, that a $100(1-\alpha)\%$ uncertainty interval should contain the true value $100(1-\alpha)\%$ of the time, is fundamentally frequentist and also a natural expectation for many users of epidemiological models. The subjective Bayesian framework, however, does not ensure calibration, since probabilities represent *beliefs* rather than long-run frequency. Despite this, many practitioners intuitively expect Bayesian credible intervals, such as those produced by EpiEstim and EpiFilter, to be reasonably well calibrated.

Despite Bayesian methods not targeting calibration, the marginalisation approach we use in Chapter 4 improves the calibration of Bayesian credible intervals. By allowing the model to be sufficiently expressive (by adapting the smoothness of R_t), it can approximate the true data-generating process, even if the model (the prior distribution and likelihood) is misspecified. Bernstein-von Mises (BvM) theorems then imply that the Bayesian credible intervals will asymptotically coincide with frequentist confidence intervals as the quantity of data increases [408, 409]. However, we note that whether a relevant BvM theorem holds for the R_t estimates themselves, or only to estimates of the smoothing parameter, remains an open question.

This idea also applies to the models for epidemic prevalence survey data considered in Chapter 6, where we again consider the calibration of Bayesian credible intervals after parameter marginalisation.

Chapter 5 provides methods for comprehensive uncertainty quantification in hidden-state-based renewal models. Many alternative published methods for fitting these models, such as linearised Kalman filters [119] or Laplace approximations [235] can misrepresent uncertainty through comparatively crude approximations of the targeted posterior distribution. The true advantage of SMC methods in this context is that they directly produce samples from the posterior distribution of interest. The computational cost of traditional Markov chain Monte Carlo methods here is often prohibitive, particularly as hidden-state models often feature highly correlated latent states (this is why the Eales approach in Chapter 6 can be slow). The advantage of SMC methods in this setting is they explicitly leverage the temporal structure of the model to more efficiently sample from the posterior distribution.

Chapter 8 takes a distinctly different view of uncertainty, being the only chapter to use a frequentist approach. Here, uncertainty about the prevalence of self-reported behaviours is addressed through the lens of sampling variability, with confidence intervals reflecting the variability that would be observed if the data collection process were repeated. This naturally leads to well-calibrated confidence intervals, under the assumption that the model is correctly specified. Two areas of concern are the validity of the model weights (they are calibrated to the demographics of the UK population, but do not account for other potential sources of bias, including correlation in inclusion probability with the outcome of interest), and the assumption of binomial sampling. As seen in Chapter 6 (Paper IV), overdispersion in epidemic prevalence survey data has substantial implications for the estimation of infection prevalence; it is reasonable to expect that overdispersion is also present in behavioural responses.

10.2.2 Consolidation, reproducibility, accessibility, and interdisciplinarity

As statistical epidemiology expands and draws on an ever-increasing variety of datasets and methods, there is a growing need for consolidation. Bringing approaches into common frameworks, clarifying their assumptions, and highlighting connections across applications will ultimately improve transparency and reproducibility. The work in this thesis contributes to this

goal by showing, for example, how various methods for modelling survey data are related, or how the SMC framework can be used to fit a range of models that account for epidemiological best practices, that would otherwise require bespoke fitting algorithms. Such consolidation strengthens the field by ensuring that methodological innovations accumulate and build on one another, rather than being developed in isolation.

Reproducibility and accessibility are also crucial for ensuring that methodological advances have a lasting impact. There has been a push towards open science in the scientific community more broadly [410], and, in my experience, epidemiology has been a leader in this regard, with the publication of well-documented open-source software being the norm for new methods. This thesis continues this tradition, with extensive online material accompanying Chapters 3-to-8, including code to reproduce all results and, in some cases, interactive applications to facilitate exploration of the methods and data. This not only allows other researchers to verify and build on the work presented here, but also makes these methods more immediately usable.

Finally, interdisciplinarity is becoming increasingly important in epidemiological modelling. Rapid advances in other fields, such as machine learning and statistics, can provide powerful new tools and perspectives for infectious disease modelling. This thesis draws on a range of ideas from across statistics and epidemiology, demonstrating how various concepts can be adapted and applied to epidemiological problems. By bridging the gap between disciplines, we can develop more robust and flexible models that are better suited to the complex challenges of statistical epidemiology.

10.2.3 Policy relevance of epidemiological modelling

To maximise their utility, epidemiological models must be both methodologically rigorous and relevant to policymakers. These two goals are sometimes seen as being in conflict, with methodological rigour being associated with complex models that are difficult to understand and communicate, while policy relevance is associated with simpler models that are easier to interpret but make unrealistic assumptions and simplifications. This thesis demonstrates that these two goals can be achieved simultaneously. We have shown that topics usually considered to be mostly theoretical, even philosophical, in nature, such as uncertainty quantification, can be directly tied to practical decision-making. Furthermore, we showed that relatively simple changes

to existing models, such as the estimation and marginalisation of smoothing parameters, can substantially improve the robustness of inferences without requiring major changes to existing workflows.

10.3 Limitations of this thesis

Like any piece of research, each chapter in this thesis has its own limitations, which are discussed in the respective chapters. Here, we recognise three overarching limitations of this thesis.

First, this thesis has a narrow **geographic focus**. All chapters use data from either the United Kingdom or Aotearoa New Zealand, two high-income Anglophone countries with substantial resources for infectious disease surveillance and research. While the methods we develop are broadly applicable, the specific applications and lessons learned may not generalise to other settings. A recent review found a limited number of infectious disease modelling studies in South Asia and Africa in particular, despite the relatively high burden of infectious diseases in these regions [23].

Second, this thesis is narrow in terms of **pathogens considered**. The only pathogen explicitly considered is SARS-CoV-2, the virus that causes COVID-19, which dominated infectious disease research from 2020 onwards, although we do consider other pathogens in the appendices. The methods developed here are broadly applicable to other pathogens, particularly those that are directly transmitted between humans with generation intervals on the order of days. Beyond this, the applicability to pathogens with much longer generation intervals (e.g., HIV) or indirect transmission (e.g., vector-borne diseases such as malaria or dengue) is less clear.

Finally, this thesis is narrow in terms of **epidemiological models**, focusing almost entirely on the renewal model. While we demonstrate the wide applicability and flexibility of this model, we do not compare it with alternative modelling frameworks, such as compartmental models, agent-based models, or network models, all of which have their own strengths and weaknesses.

10.4 Directions for future work

Addressing the limitations discussed above is crucial to broaden the applicability, and thus the impact, of the methods developed in this thesis. There are also many directions for future work that build directly on the specific chapters of this thesis, including, but not limited to:

- Chapter 3. Extending the decision-theoretic framework to account for "structural uncertainty" (including the use of reference distributions).
- Chapter 3. Investigating application-specific loss functions to enable more targeted uncertainty quantification.
- Chapter 4. Integrating the developed estimation and marginalisation techniques into existing tools to facilitate uptake by practitioners.
- Chapter 5. Using SMC-squared [344] to jointly estimate static parameters and latent states in the online/real-time setting, substantially improving the applicability and relative value of these methods for real-time outbreak responses.
- Chapter 6. Allowing for abrupt changes in growth rates to better capture sudden changes in transmission.
- Chapter 7. Using REACT-1 data in addition to reported cases and wastewater sampling data (in England) to model the dynamics of the viral shedding load, currently a major unknown in the literature. Doing so would also allow for the estimation of the absolute case ascertainment ratio, rather than just the relative case ascertainment ratio.
- Chapter 8. Further validating the scenarios in which mobility data are a good proxy for behaviour.

We conclude this thesis with a discussion of one particularly exciting direction for future work. Recent research in the machine-learning literature has explored the use of highly flexible deep-learning black-box predictive models for *inference* [216]. Such models are trained on vast quantities of data to predict future observations, which are then used to impute the "missing data" required to calculate the parameter of interest. By repeatedly imputing the missing data, a distribution for the parameter of interest can be constructed [98]. Model diagnostics to check

for coherence (or related conditions, such as martingale or conditionally identically distributed conditions) are currently under development [216, 223, 411].

This provides a clear pathway, starting from the framework introduced in Chapter 3, to the application of modern machine-learning methods in epidemiological modelling. Traditional machine-learning approaches have found limited success in epidemiology due to inherent data limitations (we only ever observe a single epidemic trajectory) and the need for robust uncertainty quantification. Machine learning in epidemiology has largely been focused on "dynamics-informed" models, effectively semi-mechanistic models fit using machine-learning methods [108]. Instead, a large predictive model could be trained on a very wide range of epidemic datasets. By taking the predictive approach to inference, this model could be used to estimate quantities such as R_t, with uncertainty quantification being validated by checking coherence diagnostics, and then improved through methods such as conformal prediction [412].

In this context, it appears we may be on the cusp of a new paradigm for infectious disease modelling. Such an approach reduces the need to specify arbitrary model structures by placing greater reliance on large quantities of data to inform the model. Furthermore, this approach naturally allows for "transfer learning" in the early stages of an outbreak of a novel pathogen, where data are inherently limited, potentially allowing for rapid and flexible inference even when little is known about the pathogen of interest. This is a promising direction for future research, and one that could substantially improve the robustness and relevance of infectious disease models for public health decision-making.

Bibliography

[1] M. Friedrich. "WHO's Top Health Threats for 2019". *JAMA* 321.11 (2019), p. 1041. ISSN: 0098-7484. DOI: 10.1001/jama.2019.1934.

[2] R. E. Baker, A. S. Mahmud, I. F. Miller, et al. "Infectious Disease in an Era of Global Change". *Nature Reviews Microbiology* 20.4 (2022), pp. 193–205. DOI: 10.1038/s41579-021-00639-z.

[3] K. E. Jones, N. G. Patel, M. A. Levy, et al. "Global Trends in Emerging Infectious Diseases". *Nature* 451.7181 (2008), pp. 990–993. ISSN: 1476-4687. DOI: 10.1038/nature06536.

[4] K. F. Smith, M. Goldberg, S. Rosenthal, et al. "Global Rise in Human Infectious Disease Outbreaks". *Journal of The Royal Society Interface* 11.101 (2014), p. 20140950. DOI: 10.1098/rsif.2014.0950.

[5] A. Tatem, D. Rogers, and S. Hay. "Global Transport Networks and Infectious Disease Spread". *Advances in Parasitology* 62 (2006), pp. 293–343. DOI: 10.1016/S0065-308X(05)62009-X.

[6] N. L. Gottdenker, D. G. Streicker, C. L. Faust, and C. R. Carroll. "Anthropogenic Land Use Change and Infectious Diseases: A Review of the Evidence". *EcoHealth* 11.4 (2014), pp. 619–632. ISSN: 1612-9210. DOI: 10.1007/s10393-014-0941-z.

[7] B. A. Jones, D. Grace, R. Kock, et al. "Zoonosis Emergence Linked to Agricultural Intensification and Environmental Change". *Proceedings of the National Academy of Sciences* 110.21 (2013), pp. 8399–8404. DOI: 10.1073/pnas.1208059110.

[8] C. J. Carlson, G. F. Albery, C. Merow, et al. "Climate Change Increases Cross-Species Viral Transmission Risk". *Nature* 607.7919 (2022), pp. 555–562. ISSN: 1476-4687. DOI: 10.1038/s41586-022-04788-w.

[9] S. S. Morse, J. A. Mazet, M. Woolhouse, et al. "Prediction and Prevention of the next Pandemic Zoonosis". *The Lancet* 380.9857 (2012), pp. 1956–1965. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(12)61684-5.

[10] S. D. Baral, T. Poteat, S. Strömdahl, et al. "Worldwide Burden of HIV in Transgender Women: A Systematic Review and Meta-Analysis". *The Lancet Infectious Diseases* 13.3 (2013), pp. 214–222. ISSN: 1473-3099. DOI: 10.1016/S1473-3099(12)70315-8.

[11] Z. A. Bhutta, J. Sommerfeld, Z. S. Lassi, et al. "Global Burden, Distribution, and Interventions for Infectious Diseases of Poverty". *Infectious Diseases of Poverty* 3.1 (2014), p. 21. ISSN: 2049-9957. DOI: 10.1186/2049-9957-3-21.

[12] N. Steyn, R. N. Binny, K. Hannah, et al. "Māori and Pacific People in New Zealand Have a Higher Risk of Hospitalisation for COVID-19". *New Zealand Medical Journal* 134.1538 (2021).

[13] A. S. Fauci and D. M. Morens. "The Perpetual Challenge of Infectious Diseases". *New England Journal of Medicine* 366.5 (2012), pp. 454–461. ISSN: 0028-4793. DOI: 10.1056/NEJMra1108296.

[14] N. G. Becker and T. Britton. "Statistical Studies of Infectious Disease Incidence". *Journal of the Royal Statistical Society Series B: Statistical Methodology* 61.2 (1999), pp. 287–307. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/1467-9868.00177.

[15] C. J. E. Metcalf and J. Lessler. "Opportunities and Challenges in Modeling Emerging Infectious Diseases". *Science* 357.6347 (2017), pp. 149–152. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aam8335.

[16] J. Wallinga and M. Lipsitch. "How Generation Intervals Shape the Relationship between Growth Rates and Reproductive Numbers". *Proceedings of the Royal Society B: Biological Sciences* 274.1609 (2006), pp. 599–604. DOI: 10.1098/rspb.2006.3754.

[17] R. Brookmeyer and M. H. Gail. *AIDS Epidemiology: A Quantitative Approach*. Oxford University Press, (1994). ISBN: 978-0-19-507641-7. DOI: 10.1093/oso/9780195076417.001.0001.

[18] A. James, M. J. Plank, R. N. Binny, et al. "A Structured Model for COVID-19 Spread: Modelling Age and Healthcare Inequities". *Mathematical Medicine and Biology: A Journal of the IMA* 38.3 (2021), pp. 299–313. ISSN: 1477-8602. DOI: 10.1093/imammb/dqab006.

[19] M. A. Johansson, K. M. Apfeldorf, S. Dobson, et al. "An Open Challenge to Advance Probabilistic Forecasting for Dengue Epidemics". *Proceedings of the National Academy of Sciences* 116.48 (2019), pp. 24268–24274. DOI: 10.1073/pnas.1909865116.

[20] S. Hendy, N. Steyn, A. James, et al. "Mathematical Modelling to Inform New Zealand's COVID-19 Response". *Journal of the Royal Society of New Zealand* 51 (2021), S86–S106. ISSN: 0303-6758, 1175-8899. DOI: 10.1080/03036758.2021.1876111.

[21] S. M. Moghadas, A. Shoukat, M. C. Fitzpatrick, et al. "Projecting Hospital Utilization during the COVID-19 Outbreaks in the United States". *Proceedings of the National Academy of Sciences* 117.16 (2020), pp. 9122–9126. DOI: 10.1073/pnas.2004064117.

[22] S. Flaxman, S. Mishra, A. Gandy, et al. "Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe". *Nature* 584.7820 (2020), pp. 257–261. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2405-7.

[23] P. Christen, M. H. A. Ahmed, B. Chua Wen Bing, et al. "Measuring the Growth of Infectious Disease Modelling Publications and Their Impact on Policymaking: A Large Language Model-assisted Bibliometric Review". *medRxiv* (2025). DOI: 10.1101/2025.06.12.25328864. Pre-published.

[24] D. De Angelis, A. M. Presanis, P. J. Birrell, et al. "Four Key Challenges in Infectious Disease Modelling Using Data from Multiple Sources". *Epidemics* 10 (2015), pp. 83–87. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2014.09.004.

[25] M. U. G. Kraemer, S. V. Scarpino, V. Marivate, et al. "Data Curation during a Pandemic and Lessons Learned from COVID-19". *Nature Computational Science* 1.1 (2021), pp. 9–10. ISSN: 2662-8457. DOI: 10.1038/s43588-020-00015-6.

[26] X. Li, F. Chadwick, and B. Swallow. "Advances in Approximate Bayesian Inference for Models in Epidemiology". *Epidemics* 53 (2025), p. 100855. ISSN: 17554365. DOI: 10.1016/j.epidem.2025.100855.

[27] M. U. G. Kraemer, J. L.-H. Tsui, S. Y. Chang, et al. "Artificial Intelligence for Modelling Infectious Disease Epidemics". *Nature* 638.8051 (2025), pp. 623–635. ISSN: 1476-4687. DOI: 10.1038/s41586-024-08564-w.

[28] J. Huang and J. S. Morris. "Infectious Disease Modeling". *Annual Review of Statistics and Its Application* 12 (2025), pp. 19–44. ISSN: 2326-8298, 2326-831X. DOI: 10.1146/annurev-statistics-112723-034351.

[29] M. S. Majumder and K. D. Mandl. "Early in the Epidemic: Impact of Preprints on Global Discourse about COVID-19 Transmissibility". *The Lancet Global Health* 8.5 (2020), e627–e630. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(20)30113-3.

[30] A. S. Henderson, R. I. Hickson, M. Furlong, et al. "Reproducibility of COVID-era Infectious Disease Models". *Epidemics* 46 (2024), p. 100743. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2024.100743.

[31] B. Jahn, S. Friedrich, J. Behnke, et al. "On the Role of Data, Statistics and Decisions in a Pandemic". *AStA Advances in Statistical Analysis* 106.3 (2022), pp. 349–382. ISSN: 1863-818X. DOI: 10.1007/s10182-022-00439-7.

[32] C. Mills, N. J. Irons, J. L.-H. Tsui, et al. "From Metric to Action: An Evaluation Framework to Translate Infectious Disease Forecasts into Policy Decisions". *medRxiv* (2025). DOI: 10.1101/2025.07.20.25331802. Pre-published.

[33] M. J. Penn, D. J. Laydon, J. Penn, et al. "Intrinsic Randomness in Epidemic Modelling beyond Statistical Uncertainty". *Communications Physics* 6.1 (2023), pp. 1–9. ISSN: 2399-3650. DOI: 10.1038/s42005-023-01265-2.

[34] R. McCabe, M. D. Kont, N. Schmit, et al. "Communicating Uncertainty in Epidemic Models". *Epidemics* 37 (2021), p. 100520. ISSN: 17554365. DOI: 10.1016/j.epidem.2021.100520.

[35] I. Holmdahl and C. Buckee. "Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us". *New England Journal of Medicine* 383.4 (2020), pp. 303–305. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMp2016822.

[36] K. M. Gostic, L. McGough, E. B. Baskerville, et al. "Practical Considerations for Measuring the Effective Reproductive Number, Rt". *PLOS Computational Biology* 16.12 (2020), e1008409. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008409.

[37] Y. Ye, A. Pandey, C. Bawden, et al. "Integrating Artificial Intelligence with Mechanistic Epidemiological Modeling: A Scoping Review of Opportunities and Challenges". *Nature Communications* 16.1 (2025), p. 581. ISSN: 2041-1723. DOI: 10.1038/s41467-024-55461-x.

[38] N. Steyn, F. B. Smith, C. Mills, et al. "A Decision-Theoretic Framework for Uncertainty Quantification in Epidemiological Modelling". *arXiv* (2025). DOI: 10.48550/ARXIV.2509.20013. Pre-published.

[39] N. Steyn and K. V. Parag. "Robust Uncertainty Quantification in Popular Estimators of the Instantaneous Reproduction Number". *American Journal of Epidemiology* (2025), kwaf165. ISSN: 0002-9262. DOI: 10.1093/aje/kwaf165.

[40] N. Steyn, K. V. Parag, R. N. Thompson, and C. A. Donnelly. "A Primer on Inference and Prediction With Epidemic Renewal Models and Sequential Monte Carlo". *Statistics in Medicine* 44.18–19 (2025), e70204. ISSN: 1097-0258. DOI: 10.1002/sim.70204.

[41] N. Steyn, M. Chadeau-Hyam, P. Elliott, and C. A. Donnelly. "A Bayesian Model for Repeated Cross-Sectional Epidemic Prevalence Survey Data". *medRxiv* (2025). DOI: 10.1101/2025.04.16.25325936. Pre-published.

[42] L. M. Watson, M. J. Plank, B. A. Armstrong, et al. "Jointly Estimating Epidemiological Dynamics of Covid-19 from Case and Wastewater Data in Aotearoa New Zealand". *Communications Medicine* 4.1 (2024), pp. 1–9. ISSN: 2730-664X. DOI: 10.1038/s43856-024-00570-3.

[43] N. Steyn, M. Chadeau-Hyam, M. Whitaker, et al. "Pandemic-Risk-Related Behaviour Change in England from June 2020 to March 2022: REACT-1 Study among over 2 Million People". *medRxiv* (2025). DOI: 10.1101/2025.03.03.25323250. Pre-published.

[44] N. Steyn, H. J. T. Unwin, J. Ponmattam, et al. "Regional and National Estimates of Children Affected by All-Cause and COVID-19-associated Orphanhood and Caregiver

Death in Brazil, by Age and Family Circumstance". *medRxiv* (2025). DOI: 10.1101/2025.01.31.25321479. Pre-published.

[45] I. Ogi-Gittins, N. Steyn, J. Polonsky, et al. "Simulation-Based Inference of the Time-Dependent Reproduction Number from Temporally Aggregated and under-Reported Disease Incidence Time Series Data". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 383.2293 (2025), p. 20240412. ISSN: 1364-503X, 1471-2962. DOI: 10.1098/rsta.2024.0412.

[46] R. N. Binny, P. Priest, N. P. French, et al. "Sensitivity of Reverse Transcription Polymerase Chain Reaction Tests for Severe Acute Respiratory Syndrome Coronavirus 2 Through Time". *The Journal of Infectious Diseases* 227.1 (2023), pp. 9–17. ISSN: 0022-1899. DOI: 10.1093/infdis/jiac317.

[47] The Royal Society. *COVID-19: Examining the Effectiveness of Non-Pharmaceutical Interventions*. (2023). ISBN: 978-1-78252-671-1. URL: https://www.royalsociety.org/npi-impact-on-covid-19.

[48] P. Elliott, M. Whitaker, D. Tang, et al. "Design and Implementation of a National SARS-CoV-2 Monitoring Program in England: REACT-1 Study". *American Journal of Public Health* 113.5 (2023), pp. 545–554. ISSN: 1541-0048. DOI: 10.2105/AJPH.2023.307230.

[49] P. Elliott, O. Eales, N. Steyn, et al. "Twin Peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 Epidemics in England". *Science* 376.6600 (2022), eabq4411. DOI: 10.1126/science.abq4411.

[50] G. Kayali. "The Forgotten History of Pre-Modern Epidemiology: Contribution of Ibn An-Nafis in the Islamic Golden Era". *Eastern Mediterranean Health Journal* 23.12 (2017), pp. 854–857. ISSN: 1020-3397, 1687-1634. DOI: 10.26719/2017.23.12.854.

[51] R. N. Thompson and E. Brooks-Pollock. "Detection, Forecasting and Control of Infectious Disease Epidemics: Modelling Outbreaks in Humans, Animals and Plants". *Philosophical Transactions of the Royal Society B: Biological Sciences* 374 (2019), p. 20190038. DOI: 10.1098/rstb.2019.0038.

[52] J. Graunt. "Natural and Political Observations Made upon the Bills of Mortality". *London: Printed by John Martyn* (1662).

[53] H. Connor. "John Graunt F.R.S. (1620-74): The Founding Father of Human Demography, Epidemiology and Vital Statistics". *Journal of Medical Biography* 32.1 (2024), pp. 57–69. ISSN: 0967-7720. DOI: 10.1177/09677720221079826.

[54] A. Hald. *A History of Probability and Statistics and Their Applications before 1750*. Wiley Series in Probability and Statistics. New York: Wiley, (2003). 586 pp. ISBN: 978-0-471-47129-5 978-0-471-72517-6.

[55] D. Bernoulli. "Essai d'une Nouvelle Analyse de La Mortalite Causee Par La Petite Verole". *Mem. Math. Phys. Acad. Roy. Sci* (1766).

[56] T. Bayes. "An Essay towards Solving a Problem in the Doctrine of Chances". *Philosophical Transactions of the Royal Society of London* 43 (1764), pp. 370–418.

[57] P. S. Laplace. "Mémoire Sur La Probabilité Des Causes Par Les Événements". *Mémoires de mathématique et de physique, presentés à l'Académie royale des sciences, par divers sçavans & lûs dans ses assemblées* 6 (1774), pp. 621–656.

[58] S. M. Stigler. "Laplace's 1774 Memoir on Inverse Probability". *Statistical Science* 1.3 (1986), pp. 359–363. ISSN: 08834237, 21688745. JSTOR: 2245475. URL: http://www.jstor.org/stable/2245475.

[59] J. M. Eyler. "The Changing Assessments of John Snow's and William Farr's Cholera Studies". *Sozial- und Präventivmedizin* 46.4 (2001), pp. 225–232. ISSN: 1420-911X. DOI: 10.1007/BF01593177.

[60] M. Santillana, A. Tuite, T. Nasserie, et al. "Relatedness of the Incidence Decay with Exponential Adjustment (IDEA) Model, "Farr's Law" and SIR Compartmental Difference Equation Models". *Infectious Disease Modelling* 3 (2018), pp. 1–12. ISSN: 2468-0427. DOI: 10.1016/j.idm.2018.03.001.

[61] Great Britain. General Register Office. *Annual Report of the Registrar-General of Births, Deaths, and Marriages in England.* London: H.M.S.O., (1840). URL: https://catalog.hathitrust.org/Record/008893397.

[62] K. Dietz. "THE FIRST EPIDEMIC MODEL: A HISTORICAL NOTE ON P.D. EN'KO". *Australian Journal of Statistics* 30A.1 (1988), pp. 56–65. ISSN: 0004-9581. DOI: 10.1111/j.1467-842X.1988.tb00464.x.

[63] I. M. Foppa. *A Historical Introduction to Mathematical Modeling of Infectious Diseases: Seminal Papers in Epidemiology.* London San Diego, CA Cambridge, MA Oxford, United Kingdom: Academic Press, (2017). 1 p. ISBN: 978-0-12-802260-3 978-0-12-802499-7.

[64] H. Abbey. "An Examination of the Reed-Frost Theory of Epidemics". *Human Biology* 24.3 (1952), p. 201. ISSN: 0018-7143. URL: https://www.proquest.com/scholarly-journals/examination-reed-frost-theory-epidemics/docview/1301828990/se-2?accountid=13042.

[65] W. O. Kermack and A. G. McKendrick. "A Contribution to the Mathematical Theory of Epidemics". *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772 (1927), pp. 700–721. DOI: 10.1098/rspa.1927.0118.

[66] A. J. Lotka. "Relation Between Birth Rates and Death Rates". *Science* 26.653 (1907), pp. 21–22. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.26.653.21.b.

[67] D. Champredon, J. Dushoff, and D. J. D. Earn. "Equivalence of the Erlang-Distributed SEIR Epidemic Model and the Renewal Equation". *SIAM Journal on Applied Mathematics* 78.6 (2018), pp. 3258–3278. ISSN: 0036-1399. DOI: 10.1137/18M1186411.

[68] S. Zabell. "R. A. Fisher on the History of Inverse Probability". *Statistical Science* 4.3 (1989). ISSN: 0883-4237. DOI: 10.1214/ss/1177012488.

[69] A. Hald. *A History of Parametric Statistical Inference from Bernoulli to Fischer: 1713-1935.* Sources and Studies in the History of Mathematics and Physical Sciences. New York: Springer, (2007). ISBN: 978-0-387-46408-4.

[70] S. L. Zabell. "The Rule of Succession". *Erkenntnis* 31.2 (1989), pp. 283–321. ISSN: 1572-8420. DOI: 10.1007/BF01236567.

[71] H. Jeffreys. "An Invariant Form for the Prior Probability in Estimation Problems". *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 186.1007 (1946), pp. 453–461. ISSN: 0080-4630, 2053-9169. DOI: 10.1098/rspa.1946.0056.

[72] N. Reid. "Asymptotics and the Theory of Inference". *The Annals of Statistics* 31.6 (2003). ISSN: 0090-5364. DOI: 10.1214/aos/1074290325.

[73] R. A. Fisher. *Statistical Methods For Research Workers.* Oliver & Boyd (Edinburgh), (1925).

[74] R. A. Fisher. *The Design of Experiments.* Macmillan Pub Co, (1935).

[75] J. Neyman and Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231.694–706 (1933), pp. 289–337. ISSN: 0264-3952, 2053-9258. DOI: 10.1098/rsta.1933.0009.

[76] L. J. Savage. "On Rereading R. A. Fisher". *The Annals of Statistics* 4.3 (1976), pp. 441–500. ISSN: 00905364, 21688966. JSTOR: 2958221. URL: http://www.jstor.org/stable/2958221.

[77] S. L. Zabell. "R. A. Fisher and Fiducial Argument". *Statistical Science* 7.3 (1992), pp. 369–387. ISSN: 08834237. JSTOR: 2246073. URL: http://www.jstor.org/stable/2246073.

[78] A. Wald. "Statistical Decision Functions Which Minimize the Maximum Risk". *The Annals of Mathematics* 46.2 (1945), p. 265. ISSN: 0003486X. DOI: 10.2307/1969022.

[79] A. Wald. "Statistical Decision Functions". *The Annals of Mathematical Statistics* 20.2 (1949), pp. 165–205. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177730030.

[80] L. J. Savage. "The Theory of Statistical Decision". *Journal of the American Statistical Association* 46.253 (1951), pp. 55–67. ISSN: 0162-1459. DOI: 10.2307/2280094.

[81] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory*. Wiley classics library ed. Wiley Classics Library. New York, NY: Wiley, (1961). 356 pp. ISBN: 978-0-471-38349-9.

[82] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. New York, NY: Springer, (1985). ISBN: 978-1-4419-3074-3 978-1-4757-4286-2. DOI: 10.1007/978-1-4757-4286-2.

[83] G. Baio. "Statistical Modeling for Health Economic Evaluations". *Annual Review of Statistics and Its Application* 5 (2018), pp. 289–309. ISSN: 2326-8298, 2326-831X. DOI: 10.1146/annurev-statistics-031017-100404.

[84] B. Efron. "Why Isn't Everyone a Bayesian?" *The American Statistician* 40.1 (1986), pp. 1–5. ISSN: 0003-1305, 1537-2731. DOI: 10.1080/00031305.1986.10475342.

[85] G. M. Martin, D. T. Frazier, and C. P. Robert. "Computing Bayes: From Then 'Til Now". *Statistical Science* 39.1 (2024), pp. 3–19. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/22-STS876.

[86] C. Robert and G. Casella. "A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data". *Statistical Science* 26.1 (2011), pp. 102–115. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/10-STS351.

[87] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. "The BUGS Project: Evolution, Critique and Future Directions". *Statistics in Medicine* 28.25 (2009), pp. 3049–3067. ISSN: 1097-0258. DOI: 10.1002/sim.3680.

[88] Stan Development Team. *Stan Reference Manual*. Version 2.36.0. (2024).

[89] P. D. O'Neill and G. O. Roberts. "Bayesian Inference for Partially Observed Stochastic Epidemics". *Journal of the Royal Statistical Society Series A: Statistics in Society* 162.1 (1999), pp. 121–129. ISSN: 0964-1998, 1467-985X. DOI: 10.1111/1467-985X.00125.

[90] L. Breiman. "Statistical Modeling: The Two Cultures". *Statistical Science* 16.3 (2001), pp. 199–215. ISSN: 08834237, 21688745. JSTOR: 2676681. URL: http://www.jstor.org/stable/2676681.

[91] A. P. Dawid. "Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach". *Journal of the Royal Statistical Society. Series A (General)* 147.2 (1984), p. 278. ISSN: 00359238. DOI: 10.2307/2981683.

[92] G. Shmueli. "To Explain or to Predict?" *Statistical Science* 25.3 (2010). ISSN: 0883-4237. DOI: 10.1214/10-STS330.

[93] V. Chernozhukov, D. Chetverikov, M. Demirer, et al. "Double/Debiased Machine Learning for Treatment and Structural Parameters". *The Econometrics Journal* 21.1 (2018), pp. C1–C68. ISSN: 1368-4221, 1368-423X. DOI: 10.1111/ectj.12097.

[94] P. McCullagh. "What Is a Statistical Model?" *The Annals of Statistics* 30.5 (2002), pp. 1225–1310. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1035844977.

[95] D. R. Cox. *Principles of Statistical Inference*. 1st ed. Cambridge University Press, (2006). ISBN: 978-0-521-86673-6 978-0-521-68567-2 978-0-511-81355-9. DOI: 10.1017/CBO9780511813559.

[96] C. Holmes. "On the Predictive Nature of Bayesian Inference" (Oxford Population Health). (2024). URL: https://www.youtube.com/watch?v=0k3Mtvb8zYw.

[97] C. P. Robert, ed. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. New York, NY: Springer New York, (2007). ISBN: 978-0-387-71598-8 978-0-387-71599-5. DOI: 10.1007/0-387-71599-1.

[98] E. Fong, C. Holmes, and S. G. Walker. "Martingale Posterior Distributions". *Journal of the Royal Statistical Society Series B: Statistical Methodology* 85.5 (2023), pp. 1357–1391. ISSN: 1369-7412. DOI: 10.1093/jrsssb/qkad005.

[99] S. Fortini and S. Petrone. "Exchangeability, Prediction and Predictive Modeling in Bayesian Statistics". *Statistical Science* 40.1 (2025). ISSN: 0883-4237. DOI: 10.1214/24-STS965.

[100] A. Gelman and C. Hennig. "Beyond Subjective and Objective in Statistics". *Journal of the Royal Statistical Society Series A: Statistics in Society* 180.4 (2017), pp. 967–1033. ISSN: 0964-1998, 1467-985X. DOI: 10.1111/rssa.12276.

[101] J. Berger. "The Case for Objective Bayesian Analysis". *Bayesian Analysis* 1.3 (2006), pp. 385–402. ISSN: 1936-0975, 1931-6690. DOI: 10.1214/06-BA115.

[102] K. Cranmer, J. Brehmer, and G. Louppe. "The Frontier of Simulation-Based Inference". *Proceedings of the National Academy of Sciences* 117.48 (2020), pp. 30055–30062. DOI: 10.1073/pnas.1912789117.

[103] M. A. Beaumont, W. Zhang, and D. J. Balding. "Approximate Bayesian Computation in Population Genetics". *Genetics* 162.4 (2002), pp. 2025–2035. ISSN: 1943-2631. DOI: 10.1093/genetics/162.4.2025.

[104] G. Papamakarios. "Neural Density Estimation and Likelihood-free Inference". *arXiv* (2019). DOI: 10.48550/arXiv.1910.13233. Pre-published.

[105] N. Banholzer, T. Mellan, H. J. T. Unwin, et al. "A Comparison of Short-Term Probabilistic Forecasts for the Incidence of COVID-19 Using Mechanistic and Statistical Time Series Models". *arXiv* (2023). DOI: 10.48550/arXiv.2305.00933. Pre-published.

[106] S. Kandula, T. Yamana, S. Pei, et al. "Evaluation of Mechanistic and Statistical Methods in Forecasting Influenza-like Illness". *Journal of The Royal Society Interface* 15.144 (2018), p. 20180174. DOI: 10.1098/rsif.2018.0174.

[107] C. Fraser. "Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic". *PLOS ONE* 2.8 (2007), e758. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0000758.

[108] J. M. G. Vilar and L. Saiz. "Dynamics-Informed Deconvolutional Neural Networks for Super-Resolution Identification of Regime Changes in Epidemiological Time Series". *Science Advances* 9.28 (2023), eadf0673. ISSN: 2375-2548. DOI: 10.1126/sciadv.adf0673.

[109] G. Panagopoulos, G. Nikolentzos, and M. Vazirgiannis. "Transfer Graph Neural Networks for Pandemic Forecasting". *Proceedings of the AAAI Conference on Artificial Intelligence* 35.6 (2021), pp. 4838–4845. ISSN: 2374-3468, 2159-5399. DOI: 10.1609/aaai.v35i6.16616.

[110] F. Bickford Smith, J. Kossen, E. Trollope, et al. "Rethinking Aleatoric and Epistemic Uncertainty". *arXiv* (2025). DOI: 10.48550/arXiv.2412.20892. Pre-published.

[111] L. D'Agostino McGowan, K. H. Grantz, and E. Murray. "Quantifying Uncertainty in Mechanistic Models of Infectious Disease". *American Journal of Epidemiology* 190.7 (2021), pp. 1377–1385. ISSN: 0002-9262, 1476-6256. DOI: 10.1093/aje/kwab013.

[112] J. Zelner, J. Riou, R. Etzioni, and A. Gelman. "Accounting for Uncertainty during a Pandemic". *Patterns* 2.8 (2021), p. 100310. ISSN: 26663899. DOI: 10.1016/j.patter.2021.100310.

[113] B. Swallow, P. Birrell, J. Blake, et al. "Challenges in Estimation, Uncertainty Quantification and Elicitation for Pandemic Modelling". *Epidemics* 38 (2022), p. 100547. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2022.100547.

[114] 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 (2013), pp. 1505–1512. ISSN: 0002-9262. DOI: 10.1093/aje/kwt133.

[115] K. V. Parag. "Improved Estimation of Time-Varying Reproduction Numbers at Low Case Incidence and between Epidemic Waves". *PLOS Computational Biology* 17.9 (2021), e1009347. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1009347.

[116] R. N. Thompson, J. E. Stockwin, R. D. van Gaalen, et al. "Improved Inference of Time-Varying Reproduction Numbers during Infectious Disease Outbreaks". *Epidemics* 29 (2019), p. 100356. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2019.100356.

[117] P. Nouvellet, A. Cori, T. Garske, et al. "A Simple Approach to Measure Transmissibility and Forecast Incidence". *Epidemics*. The RAPIDD Ebola Forecasting Challenge 22 (2018), pp. 29–35. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2017.02.012.

[118] K. V. Parag, B. J. Cowling, and C. A. Donnelly. "Deciphering Early-Warning Signals of SARS-CoV-2 Elimination and Resurgence from Limited Data at Multiple Scales". *Journal of The Royal Society Interface* 18.185 (2021), p. 20210569. ISSN: 1742-5662. DOI: 10.1098/rsif.2021.0569.

[119] S. Särkkä. *Bayesian Filtering and Smoothing*. 1st ed. Cambridge University Press, (2013). ISBN: 978-1-107-03065-7 978-1-139-34420-3 978-1-107-61928-9. DOI: 10.1017/CBO9781139344203.

[120] A. Doucet, N. De Freitas, and N. Gordon, eds. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. New York: Springer, (2001). 581 pp. ISBN: 978-0-387-95146-1.

[121] A. Doucet and A. M. Johansen. "A Tutorial on Particle Filtering and Smoothing : Fifteen Years Later". Oxford ; N.Y.: Oxford University Press, (2011), pp. 656–705. ISBN: 978-0-19-953290-2. URL: http://webcat.warwick.ac.uk/record=b2490036~S1.

[122] G. Riutort-Mayol, P.-C. Bürkner, M. R. Andersen, et al. "Practical Hilbert Space Approximate Bayesian Gaussian Processes for Probabilistic Programming". *Statistics and Computing* 33.1 (2022), p. 17. ISSN: 1573-1375. DOI: 10.1007/s11222-022-10167-2.

[123] N. J. Foti, J. Xu, D. Laird, and E. B. Fox. "Stochastic Variational Inference for Hidden Markov Models". *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani, M. Welling, C. Cortes, et al. Vol. 27. (2014). URL: https://proceedings.neurips.cc/paper_files/paper/2014/file/63ed2c35352753435bf749c40b0ce171-Paper.pdf.

[124] J. Liu, Z. Cai, P. Gustafson, and D. J. McDonald. "Rtestim: Time-varying Reproduction Number Estimation with Trend Filtering". *PLOS Computational Biology* 20.8 (2024), e1012324. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1012324.

[125] P. McCullagh and J. Nelder. *Generalized Linear Models*. 2nd ed. Routledge, (1989). ISBN: 978-0-203-75373-6. DOI: 10.1201/9780203753736.

[126] L. D. Brown, T. T. Cai, and A. DasGupta. "Interval Estimation for a Binomial Proportion". *Statistical Science* 16.2 (2001), pp. 101–117. ISSN: 0883-4237. JSTOR: 2676784. URL: https://www.jstor.org/stable/2676784.

[127] E. B. Wilson. "Probable Inference, the Law of Succession, and Statistical Inference". *Journal of the American Statistical Association* 22.158 (1927), pp. 209–212. ISSN: 0162-1459. DOI: 10.2307/2276774.

[128] A. Agresti and B. A. Coull. "Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions". *The American Statistician* 52.2 (1998), pp. 119–126. ISSN: 0003-1305. DOI: 10.2307/2685469.

[129] J. B. S. Haldane. "A Note on Inverse Probability". *Mathematical Proceedings of the Cambridge Philosophical Society* 28.1 (1932), pp. 55–61. ISSN: 0305-0041, 1469-8064. DOI: 10.1017/S0305004100010495.

[130] T. Lumley and A. Scott. "Fitting Regression Models to Survey Data". *Statistical Science* 32.2 (2017), pp. 265–278. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/16-STS605.

[131] A. Gelman. "Struggles with Survey Weighting and Regression Modeling". *Statistical Science* 22.2 (2007), pp. 153–164. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/088342306000000691.

[132] T. Lumley. *Survey: Analysis of Complex Survey Samples*. Version 4.2-1. (2023). URL: https://cran.r-project.org/web/packages/survey/index.html.

[133] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford Science Publications. Oxford: Oxford University Press, (1991). 757 pp. ISBN: 978-0-19-854599-6.

[134] O. Diekmann, J. Heesterbeek, and J. Metz. "On the Definition and the Computation of the Basic Reproduction Ratio R 0 in Models for Infectious Diseases in Heterogeneous Populations". *Journal of Mathematical Biology* 28.4 (1990). ISSN: 0303-6812, 1432-1416. DOI: 10.1007/BF00178324.

[135] O. Diekmann and H. J. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Chichester: Wiley, (2000). 320 pp. ISBN: 978-0-471-49241-2.

[136] J. Wallinga and P. Teunis. "Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures". *American Journal of Epidemiology* 160.6 (2004), pp. 509–516. ISSN: 0002-9262. DOI: 10.1093/aje/kwh255.

[137] A. James, M. J. Plank, S. Hendy, et al. "Model-Free Estimation of COVID-19 Transmission Dynamics from a Complete Outbreak". *PLOS ONE* 16.3 (2021), e0238800. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0238800.

[138] K. V. Parag, R. N. Thompson, and C. A. Donnelly. "Are Epidemic Growth Rates More Informative than Reproduction Numbers?" *Journal of the Royal Statistical Society Series A: Statistics in Society* 185 (2022), S5–S15. ISSN: 0964-1998. DOI: 10.1111/rssa.12867.

[139] S. Abbott, J. Hellewell, R. N. Thompson, et al. "Estimating the Time-Varying Reproduction Number of SARS-CoV-2 Using National and Subnational Case Counts". *Wellcome Open Research* 5 (2020), p. 112. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.16006.2.

[140] P. E. M. Fine. "The Interval between Successive Cases of an Infectious Disease". *American Journal of Epidemiology* 158.11 (2003), pp. 1039–1047. ISSN: 0002-9262. DOI: 10.1093/aje/kwg251.

[141] T. W. Russell, N. Golding, J. Hellewell, et al. "Reconstructing the Early Global Dynamics of Under-Ascertained COVID-19 Cases and Infections". *BMC Medicine* 18.1 (2020), p. 332. ISSN: 1741-7015. DOI: 10.1186/s12916-020-01790-9.

[142] NZ Ministry of Health. *Notifiable Diseases*. (2024). URL: https://www.health.govt.nz/regulation-legislation/notifiable-diseases.

[143] J. Murray and A. L. Cohen. "Infectious Disease Surveillance". *International Encyclopedia of Public Health* (2017), pp. 222–229. DOI: 10.1016/B978-0-12-803678-5.00517-8.

[144] E. Mathieu, H. Ritchie, L. Rodés-Guirao, et al. "Coronavirus Pandemic (COVID-19)". *Our World in Data* (2020). URL: https://ourworldindata.org/covid-cases.

[145] *National Flu and COVID-19 Surveillance Reports: 2021 to 2022 Season*. GOV.UK. (2022). URL: https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports-2021-to-2022-season.

[146] *Notifiable Infectious Disease Tables — CDC*. (2023). URL: https://www.cdc.gov/nndss/data-statistics/infectious-tables/index.html.

[147] M. Lipsitch, M. T. Bassett, J. S. Brownstein, et al. "Infectious Disease Surveillance Needs for the United States: Lessons from Covid-19". *Frontiers in Public Health* 12 (2024). ISSN: 2296-2565. DOI: 10.3389/fpubh.2024.1408193.

[148] S. M. Redmond, K. Alexander-Kisslig, S. C. Woodhall, et al. "Genital Chlamydia Prevalence in Europe and Non-European High Income Countries: Systematic Review and Meta-Analysis". *PLOS ONE* 10.1 (2015), e0115753. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0115753.

[149] R. P. Slaven, A. E. P. Stewart, M. Zerihun, et al. "A Cost-Analysis of Conducting Population-Based Prevalence Surveys for the Validation of the Elimination of Trachoma as a Public Health Problem in Amhara, Ethiopia". *PLOS Neglected Tropical Diseases* 14.9 (2020), e0008401. ISSN: 1935-2735. DOI: 10.1371/journal.pntd.0008401.

[150] K. B. Pouwels, T. House, E. Pritchard, et al. "Community Prevalence of SARS-CoV-2 in England from April to November, 2020: Results from the ONS Coronavirus Infection Survey". *The Lancet Public Health* 6.1 (2021), e30–e38. ISSN: 24682667. DOI: 10.1016/S2468-2667(20)30282-6.

[151] M. Chadeau-Hyam, H. Wang, O. Eales, et al. "SARS-CoV-2 Infection and Vaccine Effectiveness in England (REACT-1): A Series of Cross-Sectional Random Community Surveys". *The Lancet Respiratory Medicine* 10.4 (2022), pp. 355–366. ISSN: 2213-2600, 2213-2619. DOI: 10.1016/S2213-2600(21)00542-7.

[152] M. Whitaker, J. Elliott, B. Bodinier, et al. "Variant-Specific Symptoms of COVID-19 in a Study of 1,542,510 Adults in England". *Nature Communications* 13 (2022), p. 6856. ISSN: 2041-1723. DOI: 10.1038/s41467-022-34244-2.

[153] Google LLC. *Google COVID-19 Community Mobility Reports*. (2022). URL: https://www.google.com/covid19/mobility/.

[154] T. Hale, N. Angrist, R. Goldszmidt, et al. "A Global Panel Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker)". *Nature Human Behaviour* 5.4 (2021), pp. 529–538. ISSN: 2397-3374. DOI: 10.1038/s41562-021-01079-8.

[155] C. G. Daughton. "Wastewater Surveillance for Population-Wide Covid-19: The Present and Future". *Science of The Total Environment* 736 (2020), p. 139631. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2020.139631.

[156] A. Keshaviah, M. B. Diamond, M. J. Wade, et al. "Wastewater Monitoring Can Anchor Global Disease Surveillance Systems". *The Lancet Global Health* 11.6 (2023), e976–e981. ISSN: 2214-109X. DOI: 10.1016/S2214-109X(23)00170-5.

[157] J. Hewitt, S. Trowsdale, B. A. Armstrong, et al. "Sensitivity of Wastewater-Based Epidemiology for Detection of SARS-CoV-2 RNA in a Low Prevalence Setting". *Water Research* 211 (2022), p. 118032. ISSN: 0043-1354. DOI: 10.1016/j.watres.2021.118032.

[158] A. Venugopal, H. Ganesan, S. S. Sudalaimuthu Raja, et al. "Novel Wastewater Surveillance Strategy for Early Detection of Coronavirus Disease 2019 Hotspots". *Current Opinion in Environmental Science & Health* 17 (2020), pp. 8–13. ISSN: 2468-5844. DOI: 10.1016/j.coesh.2020.05.003.

[159] M. J. Wade, A. Lo Jacomo, E. Armenise, et al. "Understanding and Managing Uncertainty and Variability for Wastewater Monitoring beyond the Pandemic: Lessons Learned from the United Kingdom National COVID-19 Surveillance Programmes". *Journal of Hazardous Materials* 424 (2022), p. 127456. ISSN: 0304-3894. DOI: 10.1016/j.jhazmat.2021.127456.

[160] C. Mills, M. Chadeau-Hyam, P. Elliott, and C. A. Donnelly. "The Utility of Wastewater Surveillance for Monitoring SARS-CoV-2 Prevalence". *PNAS Nexus* 3.10 (2024). ISSN: 2752-6542. DOI: 10.1093/pnasnexus/pgae438.

[161] H. Wilde, W. B. Perry, O. Jones, et al. "Accounting for Dilution of SARS-CoV-2 in Wastewater Samples Using Physico-Chemical Markers". *Water* 14.18 (2022), p. 2885. ISSN: 2073-4441. DOI: 10.3390/w14182885.

[162] H. Dutta, G. Kaushik, and V. Dutta. "Wastewater-Based Epidemiology: A New Frontier for Tracking Environmental Persistence and Community Transmission of COVID-19". *Environmental Science and Pollution Research* 29.57 (2022), pp. 85688–85699. ISSN: 1614-7499. DOI: 10.1007/s11356-021-17419-0.

[163] *COVID-19 - Mobility Trends Reports*. Apple. URL: https://covid19.apple.com/mobility.

[164] *Data For Good at Meta Colocation Maps*. URL: https://dataforgood.facebook.com/dfg/tools/colocation-maps.

[165] M. Campbell, L. Marek, J. Wiki, et al. "National Movement Patterns during the COVID-19 Pandemic in New Zealand: The Unexplored Role of Neighbourhood Deprivation". *J Epidemiol Community Health* 75.9 (2021), pp. 903–905. ISSN: 0143-005X, 1470-2738. DOI: 10.1136/jech-2020-216108.

[166] K. Ågren, P. Bjelkmar, and E. Allison. "The Use of Anonymized and Aggregated Telecom Mobility Data by a Public Health Agency during the COVID-19 Pandemic: Learnings from Both the Operator and Agency Perspective". *Data & Policy* 3 (2021), e17. ISSN: 2632-3249. DOI: 10.1017/dap.2021.11.

[167] P. Nouvellet, S. Bhatia, A. Cori, et al. "Reduction in Mobility and COVID-19 Transmission". *Nature Communications* 12.1 (2021), p. 1090. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21358-2.

[168] H. J. T. Unwin, S. Mishra, V. C. Bradley, et al. "State-Level Tracking of COVID-19 in the United States". *Nature Communications* 11.1 (2020), p. 6189. ISSN: 2041-1723. DOI: 10.1038/s41467-020-19652-6.

[169] Y. Deng, H. Lin, D. He, and Y. Zhao. "Trending on the Use of Google Mobility Data in COVID-19 Mathematical Models". *Advances in Continuous and Discrete Models* 2024.1 (2024), p. 21. ISSN: 2731-4235. DOI: 10.1186/s13662-024-03816-5.

[170] A. Wesolowski, C. O. Buckee, K. Engø-Monsen, and C. J. E. Metcalf. "Connecting Mobility to Infectious Diseases: The Promise and Limits of Mobile Phone Data". *The Journal of Infectious Diseases* 214 (2016), S414–S420. ISSN: 0022-1899. DOI: 10.1093/infdis/jiw273.

[171] D. Cucinotta and M. Vanelli. "WHO Declares COVID-19 a Pandemic". *Acta Biomedica Atenei Parmensis* 91.1 (2020), pp. 157–160. ISSN: 2531-6745. DOI: 10.23750/abm.v91i1.9397.

[172] *WHO Chief Declares End to COVID-19 as a Global Health Emergency — UN News*. (2023). URL: https://news.un.org/en/story/2023/05/1136367.

[173] N. Haug, L. Geyrhofer, A. Londei, et al. "Ranking the Effectiveness of Worldwide COVID-19 Government Interventions". *Nature Human Behaviour* 4.12 (2020), pp. 1303–1312. ISSN: 2397-3374. DOI: 10.1038/s41562-020-01009-0.

[174] O. Eales, A. J. Page, S. N. Tang, et al. "SARS-CoV-2 Lineage Dynamics in England from January to March 2021 Inferred from Representative Community Samples". *medRxiv* (2021). DOI: 10.1101/2021.05.08.21256867. Pre-published.

[175] N. Steyn, S. Hendy, M. Plank, and R. Binny. *Technical Report: Modelling the Potential Spread of COVID-19 during the August 2021 Outbreak*. (2021). URL: https://www.covid19modelling.ac.nz/modelling-the-potential-spread-of-covid-19-during-the-august-2021-outbreak/.

[176] N. Ferguson, D. Laydon, G. Nedjati Gilani, et al. *Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand*. Imperial College London, (2020). DOI: 10.25561/77482.

[177] A. James, S. C. Hendy, M. J. Plank, and N. Steyn. "Suppression and Mitigation Strategies for Control of COVID-19 in New Zealand". *medRxiv* (2020). DOI: 10.1101/2020.03.26.20044677. Pre-published.

[178] Stats NZ. *National Population Estimates: At 30 June 2022*. URL: https://www.stats.govt.nz/information-releases/national-population-estimates-at-30-june-2022.

[179] D. Williams, M. Dilcher, H. Dong, et al. "Lessons from a System- Wide Response to a Measles Outbreak, Canterbury, February–April 2019". 133.1522 (2020).

[180] AshleyG. Jordan, R. J. Sadler, K. Sawford, et al. "Mycoplasma Bovis Outbreak in New Zealand Cattle: An Assessment of Transmission Trends Using Surveillance Data". *Transboundary and Emerging Diseases* 68.6 (2021), pp. 3381–3395. ISSN: 1865-1682. DOI: 10.1111/tbed.13941.

[181] R. N. Binny, M. G. Baker, S. C. Hendy, et al. "Early Intervention Is the Key to Success in COVID-19 Control". *Royal Society Open Science* 8.11 (2021), p. 210488. ISSN: 2054-5703. DOI: 10.1098/rsos.210488.

[182] E. Corlett. "New Zealand Covid Elimination Strategy to Be Phased out, Ardern Says". *The Guardian. World news* (2021). ISSN: 0261-3077. URL: https://www.theguardian.com/world/2021/oct/04/new-zealand-covid-strategy-in-transition-ardern-says-as-auckland-awaits-lockdown-decision.

[183] The New Zealand Government. *History of the COVID-19 Alert System*. Unite against COVID-19. (2022). URL: https://covid19.govt.nz/about-our-covid-19-response/history-of-the-covid-19-alert-system/.

[184] N. Steyn, M. J. Plank, A. James, et al. "Managing the Risk of a COVID-19 Outbreak from Border Arrivals". *Journal of The Royal Society Interface* 18.177 (2021), p. 20210063. DOI: 10.1098/rsif.2021.0063.

[185] NZ Ministry of Health. *COVID-19 Data for New Zealand*. (2025). URL: https://github.com/minhealthnz/nz-covid-data.

[186] *Population Estimates for the UK, England, Wales, Scotland, and Northern Ireland - Office for National Statistics*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2022.

[187] P. J. Lillie, A. Samson, A. Li, et al. "Novel Coronavirus Disease (Covid-19): The First Two Patients in the UK with Person to Person Transmission". *The Journal of Infection* 80.5 (2020), pp. 578–606. ISSN: 0163-4453. DOI: 10.1016/j.jinf.2020.02.020.

[188] P. Elliott, D. Haw, H. Wang, et al. "Exponential Growth, High Prevalence of SARS-CoV-2, and Vaccine Effectiveness Associated with the Delta Variant". *Science* 374.6574 (2021), eabl9551. DOI: 10.1126/science.abl9551.

[189] N. G. Davies, R. C. Barnard, C. I. Jarvis, et al. "Association of Tiered Restrictions and a Second Lockdown with COVID-19 Deaths and Hospital Admissions in England: A Modelling Study". *The Lancet Infectious Diseases* 21.4 (2021), pp. 482–492. ISSN: 14733099. DOI: 10.1016/S1473-3099(20)30984-1.

[190] O. J. Watson, G. Barnsley, J. Toor, et al. "Global Impact of the First Year of COVID-19 Vaccination: A Mathematical Modelling Study". *The Lancet Infectious Diseases* 22.9 (2022), pp. 1293–1302. ISSN: 14733099. DOI: 10.1016/S1473-3099(22)00320-6.

[191] E. Pritchard, P. C. Matthews, N. Stoesser, et al. "Impact of Vaccination on New SARS-CoV-2 Infections in the United Kingdom". *Nature Medicine* 27.8 (2021), pp. 1370–1378. ISSN: 1078-8956, 1546-170X. DOI: 10.1038/s41591-021-01410-w.

[192] T. Maishman, S. Schaap, D. S. Silk, et al. "Statistical Methods Used to Combine the Effective Reproduction Number, R(t), and Other Related Measures of COVID-19 in the UK". *Statistical Methods in Medical Research* 31.9 (2022), pp. 1757–1777. ISSN: 0962-2802. DOI: 10.1177/09622802221109506.

[193] A. Cori and A. Kucharski. "Inference of Epidemic Dynamics in the COVID-19 Era and Beyond". *Epidemics* 48 (2024), p. 100784. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2024.100784.

[194] A. Lloyd. "Sensitivity of Model-Based Epidemiological Parameter Estimation to Model Assumptions". *Mathematical and Statistical Estimation Approaches in Epidemiology*. Ed. by G. Chowell, J. M. Hyman, L. M. A. Bettencourt, and C. Castillo-Chavez. Dordrecht: Springer Netherlands, (2009), pp. 123–141. ISBN: 978-90-481-2313-1. DOI: 10.1007/978-90-481-2313-1_6.

[195] R. N. Thompson, T. D. Hollingsworth, V. Isham, et al. "Key Questions for Modelling COVID-19 Exit Strategies". *Proceedings of the Royal Society B: Biological Sciences* 287.1932 (2020), p. 20201405. DOI: 10.1098/rspb.2020.1405.

[196] B. Case, K. C. Dye-Braumuller, C. Evans, et al. "Adapting Vector Surveillance Using Bayesian Experimental Design: An Application to an Ongoing Tick Monitoring Program in the Southeastern United States". *Ticks and Tick-borne Diseases* 15.3 (2024), p. 102329. ISSN: 1877959X. DOI: 10.1016/j.ttbdis.2024.102329.

[197] L. Berger, N. Berger, V. Bosetti, et al. "Rational Policymaking during a Pandemic". *Proceedings of the National Academy of Sciences* 118.4 (2021), e2012704118. DOI: 10.1073/pnas.2012704118.

[198] M. J. Kochenderfer, C. Amato, G. Chowdhary, et al. *Decision Making Under Uncertainty: Theory and Application*. The MIT Press, (2015). ISBN: 978-0-262-33170-8. DOI: 10.7551/mitpress/10187.001.0001.

[199] C. Jackson, A. Presanis, S. Conti, and D. De Angelis. "Value of Information: Sensitivity Analysis and Research Design in Bayesian Evidence Synthesis". *Journal of the American Statistical Association* 114.528 (2019), pp. 1436–1449. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2018.1562932.

[200] C. H. Jackson, G. Baio, A. Heath, et al. "Value of Information Analysis in Models to Inform Health Policy". *Annual Review of Statistics and Its Application* 9.1 (2022), pp. 95–118. ISSN: 2326-8298, 2326-831X. DOI: 10.1146/annurev-statistics-040120-010730.

[201] A. Heath, N. Kunst, and C. Jackson, eds. *Value of Information for Healthcare Decision Making*. First edition. Biostatistics Series. Boca Raton: CRC Press, (2024). 1 p. ISBN: 978-1-00-315610-9 978-1-00-382558-6 978-1-00-382557-9. DOI: 10.1201/9781003156109.

[202] E. G. Ryan, C. C. Drovandi, J. M. McGree, and A. N. Pettitt. "A Review of Modern Computational Algorithms for Bayesian Optimal Design". *International Statistical Review* 84.1 (2016), pp. 128–154. ISSN: 1751-5823. DOI: 10.1111/insr.12107.

[203] Q. Tan, C. Zhang, J. Xia, et al. "Information-Guided Adaptive Learning Approach for Active Surveillance of Infectious Diseases". *Infectious Disease Modelling* 10.1 (2025), pp. 257–267. ISSN: 2468-0427. DOI: 10.1016/j.idm.2024.10.005.

[204] M. Chatzimanolakis, P. Weber, G. Arampatzis, et al. "Optimal Allocation of Limited Test Resources for the Quantification of COVID-19 Infections". *Swiss Medical Weekly* 150.5153 (2020), w20445–w20445. ISSN: 1424-3997. DOI: 10.4414/smw.2020.20445.

[205] K. V. Parag, C. A. Donnelly, and A. E. Zarebski. "Quantifying the Information in Noisy Epidemic Curves". *Nature Computational Science* 2.9 (2022), pp. 584–594. ISSN: 2662-8457. DOI: 10.1038/s43588-022-00313-1.

[206] D. V. Lindley. "On a Measure of the Information Provided by an Experiment". *The Annals of Mathematical Statistics* 27.4 (1956), pp. 986–1005. ISSN: 00034851, 21688990. JSTOR: 2237191. URL: http://www.jstor.org/stable/2237191.

[207] N. I. Bosse, H. Gruson, A. Cori, et al. "Evaluating Forecasts with Scoringutils in R". *arXiv* (2024). DOI: 10.48550/arXiv.2205.07090. Pre-published.

[208] J. Robins and L. Wasserman. "Conditioning, Likelihood, and Coherence: A Review of Some Foundational Concepts". *Journal of the American Statistical Association* 95.452 (2000), pp. 1340–1346. ISSN: 0162-1459. DOI: 10.1080/01621459.2000.10474344.

[209] T. Gneiting and A. E. Raftery. "Strictly Proper Scoring Rules, Prediction, and Estimation". *Journal of the American Statistical Association* 102.477 (2007), pp. 359–378. ISSN: 0162-1459. DOI: 10.1198/016214506000001437.

[210] T. J. Hastie, R. Tibshirani, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer Series in Statistics. New York: Springer, (2009). ISBN: 978-0-387-84858-7.

[211] M. H. DeGroot. "Uncertainty, Information, and Sequential Experiments". *The Annals of Mathematical Statistics* 33.2 (1962), pp. 404–419. ISSN: 00034851. JSTOR: 2237520. URL: http://www.jstor.org/stable/2237520.

[212] A. P. Dawid. *Coherent Measures of Discrepancy, Uncertainty and Dependence, with Applications to Bayesian Predictive Experimental Design*. Technical report, University College London., (1998).

[213] T. Gneiting, D. Wolffram, J. Resin, et al. "Model Diagnostics and Forecast Evaluation for Quantiles". *Annual Review of Statistics and Its Application* 10 (2023), pp. 597–621. ISSN: 2326-8298, 2326-831X. DOI: 10.1146/annurev-statistics-032921-020240.

[214] K. R. Moran, G. Fairchild, N. Generous, et al. "Epidemic Forecasting Is Messier Than Weather Forecasting: The Role of Human Behavior and Internet Data Streams in Epidemic Forecast". *The Journal of Infectious Diseases* 214 (2016), S404–S408. ISSN: 0022-1899. DOI: 10.1093/infdis/jiw375.

[215] P. G. Bissiri, C. C. Holmes, and S. G. Walker. "A General Framework for Updating Belief Distributions". *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 78.5 (2016), pp. 1103–1130. ISSN: 1369-7412. JSTOR: 44682909. URL: https://www.jstor.org/stable/44682909.

[216] V. Shirvaikar. "Prediction-Powered Machine Learning for Model Selection and Uncertainty". University of Oxford, (2025).

[217] X. Bertels, P. Demeyer, S. Van den Bogaert, et al. "Factors Influencing SARS-CoV-2 RNA Concentrations in Wastewater up to the Sampling Stage: A Systematic Review". *Science of The Total Environment* 820 (2022), p. 153290. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2022.153290.

[218] C. Jackson, R. Johnson, A. De Nazelle, et al. "A Guide to Value of Information Methods for Prioritising Research in Health Impact Modelling". *Epidemiologic Methods* 10.1 (2021), p. 20210012. ISSN: 2161-962X. DOI: 10.1515/em-2021-0012.

[219] E. Fenwick, L. Steuten, S. Knies, et al. "Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force". *Value in Health* 23.2 (2020), pp. 139–150. ISSN: 10983015. DOI: 10.1016/j.jval.2020.01.001.

[220] T. Rainforth, A. Foster, D. R. Ivanova, and F. Bickford Smith. "Modern Bayesian Experimental Design". *Statistical Science* 39.1 (2024). ISSN: 0883-4237. DOI: 10.1214/23-STS915.

[221] A. Alexanderian. "A Brief Note on the Bayesian D-optimality Criterion". *arXiv* (2023). DOI: 10.48550/arXiv.2212.11466. Pre-published.

[222] S. Beregi and K. V. Parag. "Optimal Algorithms for Controlling Infectious Diseases in Real Time Using Noisy Infection Data". *PLOS Computational Biology* 21.9 (2025), e1013426. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1013426.

[223] M. Battiston and L. Cappello. "New (and Old) Predictive Schemes with a.c.i.d. Sequences". *arXiv* (2025). DOI: 10.48550/arXiv.2507.21874. Pre-published.

[224] C. H. Jackson, L. Bojke, S. G. Thompson, et al. "A Framework for Addressing Structural Uncertainty in Decision Models". *Medical Decision Making* 31.4 (2011), pp. 662–674. ISSN: 0272-989X, 1552-681X. DOI: 10.1177/0272989x11406986.

[225] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. "Bayesian Model Averaging: A Tutorial". *Statistical Science* 14.4 (1999). ISSN: 0883-4237. DOI: 10.1214/ss/1009212519.

[226] M. C. Kennedy and A. O'Hagan. "Bayesian Calibration of Computer Models". *Journal of the Royal Statistical Society Series B: Statistical Methodology* 63.3 (2001), pp. 425–464. ISSN: 1369-7412. DOI: 10.1111/1467-9868.00294.

[227] R. McCabe and C. A. Donnelly. "Disease Transmission and Control Modelling at the Science–Policy Interface". *Interface Focus* 11.6 (2021), p. 20210013. DOI: 10.1098/rsfs.2021.0013.

[228] NZ Ministry of Health. *COVID-19 Update 30 August 4pm.* (2021). URL: https://www.health.govt.nz/news/covid-19-update-30-august-4pm.

[229] NZ Ministry of Health. *COVID-19 Update 11 October 2021 4pm.* (2021). URL: https://www.health.govt.nz/news/covid-19-update-11-october-2021-4pm.

[230] Center for Infectious Disease Research and Policy. *No 'reset' with Ebola Outbreak, WHO Official Says.* (2019). URL: https://www.cidrap.umn.edu/ebola/no-reset-ebola-outbreak-who-official-says.

[231] Prime Minister's Office. *Prime Minister's Statement on Coronavirus (COVID-19): 10 May 2020.* GOV.UK. (2020). URL: https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-10-may-2020.

[232] L. Pellis, P. J. Birrell, J. Blake, et al. "Estimation of Reproduction Numbers in Real Time: Conceptual and Statistical Challenges". *Journal of the Royal Statistical Society Series A: Statistics in Society* 185 (2022), S112–S130. ISSN: 0964-1998. DOI: 10.1111/rssa.12955.

[233] "Schools in North-West of England Postpone Reopening Plans after New Coronavirus Data". *The Guardian. Education* (2020). ISSN: 0261-3077. URL: https://www.theguardian.com/education/2020/jun/06/schools-north-west-england-postpone-reopening-coronavirus.

[234] The Royal Society. *Reproduction Number (R) and Growth Rate (r) of the COVID-19 Epidemic in the UK.* (2020).

[235] O. Gressani, J. Wallinga, C. L. Althaus, et al. "EpiLPS: A Fast and Flexible Bayesian Tool for Estimation of the Time-Varying Reproduction Number". *PLOS Computational Biology* 18.10 (2022), e1010618. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010618.

[236] N. Hens, M. Van Ranst, M. Aerts, et al. "Estimating the Effective Reproduction Number for Pandemic Influenza from Notification Data Made Publicly Available in Real Time: A Multi-Country Analysis for Influenza A/H1N1v 2009". *Vaccine* 29.5 (2011), pp. 896–904. ISSN: 0264-410X. DOI: 10.1016/j.vaccine.2010.05.010.

[237] C. Fraser, D. A. T. Cummings, D. Klinkenberg, et al. "Influenza Transmission in Households During the 1918 Pandemic". *American Journal of Epidemiology* 174.5 (2011), pp. 505–514. ISSN: 0002-9262. DOI: 10.1093/aje/kwr122.

[238] S. Koyama, T. Horie, and S. Shinomoto. "Estimating the Time-Varying Reproduction Number of COVID-19 with a State-Space Method". *PLOS Computational Biology* 17.1 (2021), e1008679. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008679.

[239] A. Azmon, C. Faes, and N. Hens. "On the Estimation of the Reproduction Number Based on Misreported Epidemic Data". *Statistics in Medicine* 33.7 (2014), pp. 1176–1192. ISSN: 1097-0258. DOI: 10.1002/sim.6015.

[240] R. Creswell, M. Robinson, D. Gavaghan, et al. "A Bayesian Nonparametric Method for Detecting Rapid Changes in Disease Transmission". *Journal of Theoretical Biology* 558 (2023), p. 111351. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2022.111351.

[241] O. Eales, K. E. C. Ainslie, C. E. Walters, et al. "Appropriately Smoothing Prevalence Data to Inform Estimates of Growth Rate and Reproduction Number". *Epidemics* 40 (2022), p. 100604. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2022.100604.

[242] K. V. Parag and C. A. Donnelly. "Using Information Theory to Optimise Epidemic Models for Real-Time Prediction and Estimation". *PLOS Computational Biology* 16.7 (2020), e1007990. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007990.

[243] NZ Ministry of Health. *COVID-19 Data for New Zealand*. (2024). URL: https://github.com/minhealthnz/nz-covid-data.

[244] N. Steyn, S. Hendy, M. Plank, and R. Binny. "Technical Report: Update to Modelling 7 September 2021" (2021). URL: https://www.covid19modelling.ac.nz/modelling-the-potential-spread-of-covid-19-during-the-august-2021-outbreak/.

[245] NZ Ministry of Health. *COVID-19 Update 1 September 2021*. (2021). URL: https://www.health.govt.nz/news/covid-19-update-1-september-2021.

[246] NZ Ministry of Health. *COVID-19 Update 2 September 2021*. (2021). URL: https://www.health.govt.nz/news/covid-19-update-2-september-2021.

[247] NZ Ministry of Health. *COVID-19 Update 13 September 2021 4pm*. (2021). URL: https://www.health.govt.nz/news/covid-19-update-13-september-2021-4pm.

[248] NZ Ministry of Health. *COVID-19 Update 18 October 2021 4pm*. (2021). URL: https://www.health.govt.nz/news/covid-19-update-18-october-2021-4pm.

[249] NZ Ministry of Health. *COVID-19 Update 19 October 2021*. (2021). URL: https://www.health.govt.nz/news/covid-19-update-19-october-2021.

[250] *Covid Crunch: Auckland Lockdown Cost $8 Billion*. NZ Herald. (2022). URL: https://www.nzherald.co.nz/nz/covid-19-delta-outbreak-auckland-lockdown-cost-8-billion/AGURSE2ZSR475B4FZ2LBFRKPZI/.

[251] X. Yang, S. Wang, Y. Xing, et al. "Bayesian Data Assimilation for Estimating Instantaneous Reproduction Numbers during Epidemics: Applications to COVID-19". *PLOS Computational Biology* 18.2 (2022), e1009807. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1009807.

[252] M. J. Plank, W. S. Hart, J. Polonsky, et al. "Estimation of End-of-Outbreak Probabilities in the Presence of Delayed and Incomplete Case Reporting". *Proceedings of the*

Royal Society B: Biological Sciences 292.2039 (2025), p. 20242825. DOI: 10.1098/rspb.2024.2825.

[253] L. Y. H. Chan, B. Yuan, and M. Convertino. "COVID-19 Non-Pharmaceutical Intervention Portfolio Effectiveness and Risk Communication Predominance". *Scientific Reports* 11.1 (2021), p. 10605. ISSN: 2045-2322. DOI: 10.1038/s41598-021-88309-1.

[254] C. A. Ogwara, J. W. Ronberg, S. M. Cox, et al. "Impact of Public Health Policy and Mobility Change on Transmission Potential of Severe Acute Respiratory Syndrome Coronavirus 2 in Rhode Island, March 2020 – November 2021". *Pathogens and Global Health* 118.1 (2024), pp. 65–79. ISSN: 2047-7724. DOI: 10.1080/20477724.2023.2201984.

[255] X. Liu, X. Xu, G. Li, et al. "Differential Impact of Non-Pharmaceutical Public Health Interventions on COVID-19 Epidemics in the United States". *BMC Public Health* 21.1 (2021), p. 965. ISSN: 1471-2458. DOI: 10.1186/s12889-021-10950-2.

[256] V. Barros, I. Manes, V. Akinwande, et al. "A Causal Inference Approach for Estimating Effects of Non-Pharmaceutical Interventions during Covid-19 Pandemic". *PLOS ONE* 17.9 (2022), e0265289. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0265289.

[257] S. K. Ofori, C. A. Ogwara, S. Kwon, et al. "SARS-CoV-2 Transmission Potential and Rural-Urban Disease Burden Disparities across Alabama, Louisiana, and Mississippi, March 2020 — May 2021". *Annals of Epidemiology* 71 (2022), pp. 1–8. ISSN: 1047-2797. DOI: 10.1016/j.annepidem.2022.04.006.

[258] Y. Bo, C. Guo, C. Lin, et al. "Effectiveness of Non-Pharmaceutical Interventions on COVID-19 Transmission in 190 Countries from 23 January to 13 April 2020". *International Journal of Infectious Diseases* 102 (2021), pp. 247–253. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2020.10.066.

[259] D. Rubin, J. Huang, B. T. Fisher, et al. "Association of Social Distancing, Population Density, and Temperature With the Instantaneous Reproduction Number of SARS-CoV-2 in Counties Across the United States". *JAMA Network Open* 3.7 (2020), e2016099. ISSN: 2574-3805. DOI: 10.1001/jamanetworkopen.2020.16099.

[260] M. Kendall, L. Milsom, L. Abeler-Dörner, et al. "Epidemiological Changes on the Isle of Wight after the Launch of the NHS Test and Trace Programme: A Preliminary Analysis". *The Lancet Digital Health* 2.12 (2020), e658–e666. ISSN: 2589-7500. DOI: 10.1016/S2589-7500(20)30241-7.

[261] R. E. Baker, W. Yang, G. A. Vecchi, et al. "Assessing the Influence of Climate on Wintertime SARS-CoV-2 Outbreaks". *Nature Communications* 12.1 (2021), p. 846. ISSN: 2041-1723. DOI: 10.1038/s41467-021-20991-1.

[262] N. I. Bosse, S. Abbott, A. Cori, et al. "Scoring Epidemiological Forecasts on Transformed Scales". *PLOS Computational Biology* 19.8 (2023), e1011393. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1011393.

[263] N. Bosse, S. Abbott, H. Gruson, et al. *Scoringutils: Utilities for Scoring and Assessing Predictions*. Version 1.2.2. (2023). URL: https://cloud.r-project.org/web/packages/scoringutils/index.html.

[264] K. V. Parag, B. J. Cowling, and B. C. Lambert. "Angular Reproduction Numbers Improve Estimates of Transmissibility When Disease Generation Times Are Misspecified or Time-Varying". *Proceedings of the Royal Society B: Biological Sciences* 290.2007 (2023), p. 20231664. DOI: 10.1098/rspb.2023.1664.

[265] G. Freeguard. "The Story of the R Number: How an Obscure Epidemiological Figure Took over Our Lives. Part 4: The Politics of R". *Significance* 21.4 (2024), pp. 19–22. ISSN: 1740-9705, 1740-9713. DOI: 10.1093/jrssig/qmae058.

[266] J. A. Lewnard and A. L. Reingold. "Emerging Challenges and Opportunities in Infectious Disease Epidemiology". *American Journal of Epidemiology* 188.5 (2019), pp. 873–882. ISSN: 0002-9262. DOI: 10.1093/aje/kwy264.

[267] O. Diekmann. "Limiting Behaviour in an Epidemic Model". *Nonlinear Analysis: Theory, Methods & Applications* 1.5 (1977), pp. 459–470. ISSN: 0362-546X. DOI: 10.1016/0362-546X(77)90011-6.

[268] S. Bhatt, N. Ferguson, S. Flaxman, et al. "Semi-Mechanistic Bayesian Modelling of COVID-19 with Renewal Processes". *Journal of the Royal Statistical Society Series A: Statistics in Society* 186.4 (2023), pp. 601–615. ISSN: 0964-1998. DOI: 10.1093/jrsssa/qnad030.

[269] J. Bracher, D. Wolffram, J. Deuschel, et al. "A Pre-Registered Short-Term Forecasting Study of COVID-19 in Germany and Poland during the Second Wave". *Nature Communications* 12.1 (2021), p. 5173. ISSN: 2041-1723. DOI: 10.1038/s41467-021-25207-0.

[270] B. A. Djaafara, N. Imai, E. Hamblion, et al. "A Quantitative Framework for Defining the End of an Infectious Disease Outbreak: Application to Ebola Virus Disease". *American Journal of Epidemiology* 190.4 (2021), pp. 642–651. ISSN: 0002-9262. DOI: 10.1093/aje/kwaa212.

[271] R. Thompson, W. Hart, M. Keita, et al. "Using Real-Time Modelling to Inform the 2017 Ebola Outbreak Response in DR Congo". *Nature Communications* 15.1 (2024), p. 5667. ISSN: 2041-1723. DOI: 10.1038/s41467-024-49888-5.

[272] R. K. Nash, P. Nouvellet, and A. Cori. "Real-Time Estimation of the Epidemic Reproduction Number: Scoping Review of the Applications and Challenges". *PLOS Digital Health* 1.6 (2022). Ed. by M. Tizzoni, e0000052. ISSN: 2767-3170. DOI: 10.1371/journal.pdig.0000052.

[273] T. McKinley, A. R. Cook, and R. Deardon. "Inference in Epidemic Models without Likelihoods". *The International Journal of Biostatistics* 5.1 (2009). ISSN: 1557-4679. DOI: 10.2202/1557-4679.1171.

[274] C. Dai, J. Heng, P. E. Jacob, and N. Whiteley. "An Invitation to Sequential Monte Carlo Samplers". *Journal of the American Statistical Association* 117.539 (2022), pp. 1587–1600. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2022.2087659.

[275] A. Safarishahrbijari, A. Teyhouee, C. Waldner, et al. "Predictive Accuracy of Particle Filtering in Dynamic Models Supporting Outbreak Projections". *BMC Infectious Diseases* 17.1 (2017), p. 648. ISSN: 1471-2334. DOI: 10.1186/s12879-017-2726-9.

[276] X. Li, V. Patel, L. Duan, et al. "Real-Time Epidemiology and Acute Care Need Monitoring and Forecasting for COVID-19 via Bayesian Sequential Monte Carlo-Leveraged Transmission Models". *International Journal of Environmental Research and Public Health* 21.2 (2024), p. 193. ISSN: 1660-4601. DOI: 10.3390/ijerph21020193.

[277] J. Welding and P. Neal. "Real Time Analysis of Epidemic Data". *arXiv* (2019). DOI: 10.48550/arXiv.1909.11560. Pre-published.

[278] G. Storvik, A. Diz-Lois Palomares, S. Engebretsen, et al. "A Sequential Monte Carlo Approach to Estimate a Time-Varying Reproduction Number in Infectious Disease Models: The Covid-19 Case". *Journal of the Royal Statistical Society Series A: Statistics in Society* 186.4 (2023), pp. 616–632. ISSN: 0964-1998. DOI: 10.1093/jrsssa/qnad043.

[279] D. M. Sheinson, J. Niemi, and W. Meiring. "Comparison of the Performance of Particle Filter Algorithms Applied to Tracking of a Disease Epidemic". *Mathematical Biosciences* 255 (2014), pp. 21–32. ISSN: 0025-5564. DOI: 10.1016/j.mbs.2014.06.018.

[280] Y. S. Won, W.-S. Son, S. Choi, and J.-H. Kim. "Estimating the Instantaneous Reproduction Number (Rt) by Using Particle Filter". *Infectious Disease Modelling* 8.4 (2023), pp. 1002–1014. ISSN: 2468-0427. DOI: 10.1016/j.idm.2023.08.003.

[281] W. Yang, A. Karspeck, and J. Shaman. "Comparison of Filtering Methods for the Modeling and Retrospective Forecasting of Influenza Epidemics". *PLOS Computational Biology* 10.4 (2014), e1003583. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003583.

[282] L. Rimella, C. Jewell, and P. Fearnhead. "Approximating Optimal SMC Proposal Distributions in Individual-Based Epidemic Models". *arXiv* (2023). DOI: 10.48550/arXiv.2206.05161. Pre-published.

[283] D. Temfack and J. Wyse. "A Review of Sequential Monte Carlo Methods for Real-Time Disease Modeling". *arXiv* (2024). DOI: 10.48550/ARXIV.2408.15739. Pre-published.

[284] A. Endo, E. van Leeuwen, and M. Baguelin. "Introduction to Particle Markov-chain Monte Carlo for Disease Dynamics Modellers". *Epidemics* 29 (2019), p. 100363. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2019.100363.

[285] J. A. Scott, A. Gandy, S. Mishra, et al. *Epidemia: Modeling of Epidemics Using Hierarchical Bayesian Models*. Version R package version 1.0.0. URL: https://imperialcollegelondon.github.io/epidemia/.

[286] A. Lison, S. Abbott, J. Huisman, and T. Stadler. "Generative Bayesian Modeling to Nowcast the Effective Reproduction Number from Line List Data with Missing Symptom Onset Dates". *PLOS Computational Biology* 20.4 (2024), e1012021. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1012021.

[287] T. Kypraios, P. Neal, and D. Prangle. "A Tutorial Introduction to Bayesian Inference for Stochastic Epidemic Models Using Approximate Bayesian Computation". *Mathematical Biosciences*. 50th Anniversary Issue 287 (2017), pp. 42–53. ISSN: 0025-5564. DOI: 10.1016/j.mbs.2016.07.001.

[288] T. J. McKinley, I. Vernon, I. Andrianakis, et al. "Approximate Bayesian Computation and Simulation-Based Inference for Complex Stochastic Epidemic Models". *Statistical Science* 33.1 (2018), pp. 4–18. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/17-STS618.

[289] N. Kantas, A. Doucet, S. S. Singh, et al. "On Particle Methods for Parameter Estimation in State-Space Models". *Statistical Science* 30.3 (2015), pp. 328–351. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/14-STS511.

[290] A. A. King, D. Nguyen, and E. L. Ionides. "Statistical Inference for Partially Observed Markov Processes via the *R* Package **Pomp**". *Journal of Statistical Software* 69.12 (2016). ISSN: 1548-7660. DOI: 10.18637/jss.v069.i12.

[291] S. Cauchemez, P.-Y. Boëlle, G. Thomas, and A.-J. Valleron. "Estimating in Real Time the Efficacy of Measures to Control Emerging Communicable Diseases". *American Journal of Epidemiology* 164.6 (2006), pp. 591–597. ISSN: 1476-6256, 0002-9262. DOI: 10.1093/aje/kwj274.

[292] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. "Superspreading and the Effect of Individual Variation on Disease Emergence". *Nature* 438.7066 (2005), pp. 355–359. ISSN: 1476-4687. DOI: 10.1038/nature04153.

[293] I. Ogi-Gittins, W. S. Hart, J. Song, et al. "A Simulation-Based Approach for Estimating the Time-Dependent Reproduction Number from Temporally Aggregated Disease Incidence Time Series Data". *Epidemics* 47 (2024), p. 100773. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2024.100773.

[294] R. K. Nash, S. Bhatt, A. Cori, and P. Nouvellet. "Estimating the Epidemic Reproduction Number from Temporally Aggregated Incidence Data: A Statistical Modelling Approach and Software Tool". *PLOS Computational Biology* 19.8 (2023), e1011439. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1011439.

[295] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation". *IEE Proceedings F (Radar and Signal Processing)* 140.2 (1993), pp. 107–113. ISSN: 2053-9045. DOI: 10.1049/ip-f-2.1993.0015.

[296] N. Chopin and O. Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer Series in Statistics. Cham: Springer International Publishing, (2020). ISBN: 978-3-030-47844-5 978-3-030-47845-2. DOI: 10.1007/978-3-030-47845-2.

[297] M. A. Vink, M. C. J. Bootsma, and J. Wallinga. "Serial Intervals of Respiratory Infectious Diseases: A Systematic Review and Analysis". *American Journal of Epidemiology* 180.9 (2014), pp. 865–875. ISSN: 1476-6256. DOI: 10.1093/aje/kwu209.

[298] V. Elvira, L. Martino, and C. P. Robert. "Rethinking the Effective Sample Size". *International Statistical Review* 90.3 (2022), pp. 525–550. ISSN: 0306-7734, 1751-5823. DOI: 10.1111/insr.12500.

[299] M. K. Pitt and N. Shephard. "Filtering via Simulation: Auxiliary Particle Filters". *Journal of the American Statistical Association* 94.446 (1999), pp. 590–599. ISSN: 0162-1459. DOI: 10.2307/2670179.

[300] C. Andrieu, A. Doucet, and R. Holenstein. "Particle Markov Chain Monte Carlo Methods". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3 (2010), pp. 269–342. ISSN: 1467-9868. DOI: 10.1111/j.1467-9868.2009.00736.x.

[301] A. Gelman. *Bayesian Data Analysis*. Third edition. Chapman & Hall/CRC Texts in Statistical Science. Boca Raton: CRC Press, (2014).

[302] M. Hürzeler and H. R. Künsch. "Approximating and Maximising the Likelihood for a General State-Space Model". *Sequential Monte Carlo Methods in Practice*. Ed. by A. Doucet, N. de Freitas, and N. Gordon. Statistics for Engineering and Information Science. New York, NY: Springer, (2001), pp. 159–175. ISBN: 978-1-4757-3437-9. DOI: 10.1007/978-1-4757-3437-9_8.

[303] A. Gelman and D. B. Rubin. "Inference from Iterative Simulation Using Multiple Sequences". *Statistical Science* 7.4 (1992), pp. 457–472. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/ss/1177011136.

[304] H. Ge, K. Xu, and Z. Ghahramani. "Turing: A Language for Flexible Probabilistic Inference". *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain* (2018), pp. 1682–1690. URL: http://proceedings.mlr.press/v84/ge18b.html.

[305] A. Doucet, M. K. Pitt, G. Deligiannidis, and R. Kohn. "Efficient Implementation of Markov Chain Monte Carlo When Using an Unbiased Likelihood Estimator". *Biometrika* 102.2 (2015), pp. 295–313. ISSN: 0006-3444. DOI: 10.1093/biomet/asu075.

[306] C. Andrieu and J. Thoms. "A Tutorial on Adaptive MCMC". *Statistics and Computing* 18.4 (2008), pp. 343–373. ISSN: 1573-1375. DOI: 10.1007/s11222-008-9110-y.

[307] D. B. Rubin. "Inference and Missing Data". *Biometrika* 63.3 (1976), pp. 581–592. ISSN: 0006-3444. DOI: 10.1093/biomet/63.3.581.

[308] P. Klepac, S. Funk, T. D. Hollingsworth, et al. "Six Challenges in the Eradication of Infectious Diseases". *Epidemics* 10 (2015), pp. 97–101. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2014.12.001.

[309] A. Jordan, F. Krüger, and S. Lerch. "Evaluating Probabilistic Forecasts with scoringRules". *Journal of Statistical Software* 90 (2019), pp. 1–37. ISSN: 1548-7660. DOI: 10.18637/jss.v090.i12.

[310] L. F. White and M. Pagano. "Reporting Errors in Infectious Disease Outbreaks, with an Application to Pandemic Influenza A/H1N1". *Epidemiologic Perspectives & Innovations* 7.1 (2010), p. 12. ISSN: 1742-5573. DOI: 10.1186/1742-5573-7-12.

[311] K. Sherratt, S. Abbott, S. R. Meakin, et al. "Exploring Surveillance Data Biases When Estimating the Reproduction Number: With Insights into Subpopulation Transmission of COVID-19 in England". *Philosophical Transactions of the Royal Society B: Biological Sciences* 376.1829 (2021), p. 20200283. DOI: 10.1098/rstb.2020.0283.

[312] M. G. Roberts and H. Nishiura. "Early Estimation of the Reproduction Number in the Presence of Imported Cases: Pandemic Influenza H1N1-2009 in New Zealand". *PLOS ONE* 6.5 (2011), e17835. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0017835.

[313] T. S. Churcher, J. M. Cohen, J. Novotny, et al. "Measuring the Path toward Malaria Elimination". *Science* 344.6189 (2014), pp. 1230–1232. DOI: 10.1126/science.1251449.

[314] J. F. Lawless. "Adjustments for Reporting Delays and the Prediction of Occurred but Not Reported Events". *The Canadian Journal of Statistics* 22.1 (1994), pp. 15–31. ISSN: 0319-5724. DOI: 10.2307/3315820.

[315] L. Alvarez, M. Colom, and J.-M. Morel. "Removing Weekly Administrative Noise in the Daily Count of COVID-19 New Cases. Application to the Computation of Rt". *medRxiv* (2020). DOI: 10.1101/2020.11.16.20232405. Pre-published.

[316] L. Bengtsson, J. Gaudart, X. Lu, et al. "Using Mobile Phone Data to Predict the Spatial Spread of Cholera". *Scientific Reports* 5.1 (2015), p. 8923. ISSN: 2045-2322. DOI: 10.1038/srep08923.

[317] *COVID-19 Community Mobility Report*. COVID-19 Community Mobility Report. URL: https://www.google.com/covid19/mobility?hl=en.

[318] D. A. Broniatowski, M. J. Paul, and M. Dredze. "National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic". *PLOS ONE* 8.12 (2013), e83672. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0083672.

[319] V. C. Bradley, S. Kuriwaki, M. Isakov, et al. "Unrepresentative Big Surveys Significantly Overestimated US Vaccine Uptake". *Nature* 600.7890 (2021), pp. 695–700. ISSN: 1476-4687. DOI: 10.1038/s41586-021-04198-4.

[320] S. Abbott and S. Funk. *Estimating Epidemiological Quantities from Repeated Cross-Sectional Prevalence Measurements*. preprint. Epidemiology, (2022). DOI: 10.1101/2022.03.29.22273101.

[321] T. Ward, M. Fyles, A. Glaser, et al. "The Real-Time Infection Hospitalisation and Fatality Risk across the COVID-19 Pandemic in England". *Nature Communications* 15.1 (2024), p. 4633. ISSN: 2041-1723. DOI: 10.1038/s41467-024-47199-3.

[322] O. Eales, D. Haw, H. Wang, et al. "Dynamics of SARS-CoV-2 Infection Hospitalisation and Infection Fatality Ratios over 23 Months in England". *PLOS Biology* 21.5 (2023), e3002118. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.3002118.

[323] *Total Cost of the COVID-19 Infection Survey - Office for National Statistics*. URL: https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/totalcostofthecovid19infectionsurvey.

[324] A. Agresti and B. A. Coull. "Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions". *The American Statistician* 52.2 (1998), pp. 119–126. ISSN: 0003-1305. DOI: 10.2307/2685469.

[325] K. Charniga, S. W. Park, A. R. Akhmetzhanov, et al. "Best Practices for Estimating and Reporting Epidemiological Delay Distributions of Infectious Diseases". *PLOS Computational Biology* 20.10 (2024), e1012520. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1012520.

[326] X. Xu, Y. Wu, A. G. Kummer, et al. "Assessing Changes in Incubation Period, Serial Interval, and Generation Time of SARS-CoV-2 Variants of Concern: A Systematic Review and Meta-Analysis". *BMC Medicine* 21.1 (2023), p. 374. ISSN: 1741-7015. DOI: 10.1186/s12916-023-03070-8.

[327] J. Hellewell, T. W. Russell, R. Matthews, et al. "Estimating the Effectiveness of Routine Asymptomatic PCR Testing at Different Frequencies for the Detection of SARS-CoV-2 Infections". *BMC Medicine* 19.1 (2021), p. 106. ISSN: 1741-7015. DOI: 10.1186/s12916-021-01982-x.

[328] A. Luceño and F. D. Ceballos. "Describing Extra-Binomial Variation with Partially Correlated Models". *Communications in Statistics - Theory and Methods* 24.6 (1995), pp. 1637–1653. ISSN: 0361-0926. DOI: 10.1080/03610929508831576.

[329] B. Wagner, P. Riggs, and S. Mikulich-Gilbertson. "The Importance of Distribution-Choice in Modeling Substance Use Data: A Comparison of Negative Binomial, Beta Binomial, and Zero-Inflated Distributions". *The American Journal of Drug and Alcohol Abuse* 41.6 (2015), pp. 489–497. ISSN: 0095-2990, 1097-9891. DOI: 10.3109/00952990.2015.1056447.

[330] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. "Julia: A Fresh Approach to Numerical Computing". *SIAM Review* 59.1 (2017), pp. 65–98. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/141000671.

[331] R Core Team. *R: A Language and Environment for Statistical Computing*. manual. Vienna, Austria: R Foundation for Statistical Computing, (2021). URL: https://www.R-project.org/.

[332] J. Gabry, R. Češnovar, A. Johnson, and S. Bronder. *Cmdstanr: R Interface to 'CmdStan'*. manual. (2024). URL: https://mc-stan.org/cmdstanr/.

[333] A. Solin and S. Särkkä. "Hilbert Space Methods for Reduced-Rank Gaussian Process Regression". *Statistics and Computing* 30.2 (2020), pp. 419–446. ISSN: 0960-3174, 1573-1375. DOI: 10.1007/s11222-019-09886-w.

[334] M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur. "Gaussian Processes and Kernel Methods: A Review on Connections and Equivalences". *arXiv* (2018). DOI: 10.48550/arXiv.1807.02582. Pre-published.

[335] P. Calistri, L. Amato, I. Puglia, et al. "Infection Sustained by Lineage B.1.1.7 of SARS-CoV-2 Is Characterised by Longer Persistence and Higher Viral RNA Loads in Nasopharyngeal Swabs". *International Journal of Infectious Diseases* 105 (2021), pp. 753–755. ISSN: 1201-9712. DOI: 10.1016/j.ijid.2021.03.005.

[336] J. A. Hay, L. Kennedy-Shaffer, S. Kanjilal, et al. "Estimating Epidemiologic Dynamics from Cross-Sectional Viral Load Distributions". *Science* 373.6552 (2021), eabh0635. DOI: 10.1126/science.abh0635.

[337] F. E. J. Harrell. *Hmisc: Harrell Miscellaneous*. Version 5.1-3. (2024). URL: https://hbiostat.org/r/hmisc/.

[338] T. Sharot. "Weighting Survey Results". *Journal of the Market Research Society* 28.3 (1986), pp. 269–284.

[339] B. Cazelles, C. Champagne, and J. Dureau. "Accounting for Non-Stationarity in Epidemiology by Embedding Time-Varying Parameters in Stochastic Models". *PLOS Computational Biology* 14.8 (2018), e1006211. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006211.

[340] UK Health Security Agency. *The R Value and Growth Rate*. (2022). URL: https://www.gov.uk/guidance/the-r-value-and-growth-rate.

[341] E. K. Brockhaus, D. Wolffram, T. Stadler, et al. "Why Are Different Estimates of the Effective Reproductive Number so Different? A Case Study on COVID-19 in Germany". *PLOS Computational Biology* 19.11 (2023), e1011653. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1011653.

[342] L. M. Kucirka, S. A. Lauer, O. Laeyendecker, et al. "Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure". *Annals of Internal Medicine* 173.4 (2020), pp. 262–267. ISSN: 0003-4819. DOI: 10.7326/M20-1495.

[343] Z. Zhang, Q. Bi, S. Fang, et al. "Insight into the Practical Performance of RT-PCR Testing for SARS-CoV-2 Using Serological Data: A Cohort Study". *The Lancet Microbe* 2.2 (2021), e79–e87. ISSN: 2666-5247. DOI: 10.1016/S2666-5247(20)30200-7.

[344] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. "SMC2: An Efficient Algorithm for Sequential Analysis of State Space Models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.3 (2013), pp. 397–426. ISSN: 1467-9868. DOI: 10.1111/j.1467-9868.2012.01046.x.

[345] A. Hadji and B. Szabó. "Can We Trust Bayesian Uncertainty Quantification from Gaussian Process Priors with Squared Exponential Covariance Kernel?" *SIAM/ASA Journal on Uncertainty Quantification* 9.1 (2021), pp. 185–230. ISSN: 2166-2525. DOI: 10.1137/19m1253010.

[346] K. B. Pouwels, D. W. Eyre, T. House, et al. "Improving the Representativeness of UK's National COVID-19 Infection Survey through Spatio-Temporal Regression and Post-Stratification". *Nature Communications* 15.1 (2024), p. 5340. ISSN: 2041-1723. DOI: 10.1038/s41467-024-49201-4.

[347] A. Gelman, J. Lax, J. Phillips, et al. "Using Multilevel Regression and Poststratification to Estimate Dynamic Public Opinion" (2018). URL: http://www.stat.columbia.edu/~gelman/research/unpublished/MRT(1).pdf.

[348] NZ Ministry of Health. *COVID-19 Data for New Zealand*. (2023). URL: https://github.com/minhealthnz/nz-covid-data.

[349] ESR. *COVID-19 Data Repository by the Institute of Environmental Science and Research*. (2023). URL: https://github.com/ESR-NZ/covid_in_wastewater.

[350] L. M. Watson, M. J. Plank, B. A. Armstrong, et al. *NZ Wastewater Modelling Code*. (2024). URL: https://github.com/nicsteyn2/NZWastewaterModelling.

[351] E. Colman, G. A. Puspitarani, J. Enright, and R. R. Kao. "Ascertainment Rate of SARS-CoV-2 Infections from Healthcare and Community Testing in the UK". *Journal of Theoretical Biology* 558 (2023), p. 111333. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2022.111333.

[352] G. Vattiato, A. Lustig, O. J. Maclaren, and M. J. Plank. "Modelling the Dynamics of Infection, Waning of Immunity and Re-Infection with the Omicron Variant of SARS-CoV-2 in Aotearoa New Zealand". *Epidemics* 41 (2022), p. 100657. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2022.100657.

[353] F. S. Dawood, C. A. Porucznik, V. Veguilla, et al. "Incidence Rates, Household Infection Risk, and Clinical Characteristics of SARS-CoV-2 Infection Among Children and Adults in Utah and New York City, New York". *JAMA Pediatrics* 176.1 (2022), pp. 59–67. ISSN: 2168-6203. DOI: 10.1001/jamapediatrics.2021.4217.

[354] S. Riley, C. Atchison, D. Ashby, et al. "REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 Virus: Study Protocol". *Wellcome Open Research* 5 (2021), p. 200. ISSN: 2398-502X. DOI: 10.12688/wellcomeopenres.16228.2.

[355] S. Nourbakhsh, A. Fazil, M. Li, et al. "A Wastewater-Based Epidemic Model for SARS-CoV-2 with Application to Three Canadian Cities". *Epidemics* 39 (2022), p. 100560. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2022.100560.

[356] J. S. Huisman, J. Scire, D. C. Angst, et al. "Estimation and Worldwide Monitoring of the Effective Reproductive Number of SARS-CoV-2". *eLife* 11 (2022), e71345. ISSN: 2050-084X. DOI: 10.7554/eLife.71345.

[357] J. S. Huisman, J. Scire, L. Caduff, et al. "Wastewater-Based Estimation of the Effective Reproductive Number of SARS-CoV-2". *Environmental Health Perspectives* 130.5 (2022). ISSN: 0091-6765, 1552-9924. DOI: 10.1289/ehp10050.

[358] M. Asadi, F. F. Oloye, Y. Xie, et al. "A Wastewater-Based Risk Index for SARS-CoV-2 Infections among Three Cities on the Canadian Prairie". *Science of The Total Environment* 876 (2023), p. 162800. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2023.162800.

[359] D. L. Wannigama, M. Amarasiri, P. Hongsing, et al. "COVID-19 Monitoring with Sparse Sampling of Sewered and Non-Sewered Wastewater in Urban and Rural Communities". *iScience* 26.7 (2023), p. 107019. ISSN: 2589-0042. DOI: 10.1016/j.isci.2023.107019.

[360] J. Scire, J. S. Huisman, A. Grosu, et al. "estimateR: An R Package to Estimate and Monitor the Effective Reproductive Number". *BMC Bioinformatics* 24.1 (2023), p. 310. ISSN: 1471-2105. DOI: 10.1186/s12859-023-05428-4.

[361] G. Jiang, J. Wu, J. Weidhaas, et al. "Artificial Neural Network-Based Estimation of COVID-19 Case Numbers and Effective Reproduction Rate Using Wastewater-Based Epidemiology". *Water Research* 218 (2022), p. 118451. ISSN: 0043-1354. DOI: 10.1016/j.watres.2022.118451.

[362] B. Pell, S. Brozak, T. Phan, et al. "The Emergence of a Virus Variant: Dynamics of a Competition Model with Cross-Immunity Time-Delay Validated by Wastewater Surveillance Data for COVID-19". *Journal of Mathematical Biology* 86.5 (2023), p. 63. ISSN: 1432-1416. DOI: 10.1007/s00285-023-01900-0.

[363] V. Kisand, P. Laas, K. Palmik-Das, et al. "Prediction of COVID-19 Positive Cases, a Nation-Wide SARS-CoV-2 Wastewater-Based Epidemiology Study". *Water Research* 231 (2023), p. 119617. ISSN: 0043-1354. DOI: 10.1016/j.watres.2023.119617.

[364] E. L. P. E. Geubbels, J. A. Backer, F. Bakhshi-Raiez, et al. "The Daily Updated Dutch National Database on COVID-19 Epidemiology, Vaccination and Sewage Surveillance". *Scientific Data* 10.1 (2023), p. 469. ISSN: 2052-4463. DOI: 10.1038/s41597-023-02232-w.

[365] Z. Fang, A. M. I. Roberts, C.-D. Mayer, et al. "Wastewater Monitoring of COVID-19: A Perspective from Scotland". *Journal of Water and Health* 20.12 (2022), pp. 1688–1700. ISSN: 1477-8920, 1996-7829. DOI: 10.2166/wh.2022.082.

[366] O. McManus, L. E. Christiansen, M. Nauta, et al. "Predicting COVID-19 Incidence Using Wastewater Surveillance Data, Denmark, October 2021–June 2022". *Emerging Infectious Diseases* 29.8 (2023). ISSN: 1080-6040, 1080-6059. DOI: 10.3201/eid2908.221634.

[367] X. Bertels, S. Hanoteaux, R. Janssens, et al. "Time Series Modelling for Wastewater-Based Epidemiology of COVID-19: A Nationwide Study in 40 Wastewater Treatment Plants of Belgium, February 2021 to June 2022". *Science of The Total Environment* 899 (2023), p. 165603. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2023.165603.

[368] L. M. Watson, M. J. Plank, B. A. Armstrong, et al. *NZ Wastewater Modelling Code (Zenodo Backup)*. (2024). URL: https://doi.org/%2010.5281/zenodo.11081779.

[369] N. Golding, D. J. Price, G. Ryan, et al. "A Modelling Approach to Estimate the Transmissibility of SARS-CoV-2 during Periods of High, Low, and Zero Case Incidence". *eLife* (2023). DOI: 10.7554/eLife.78089.

[370] G. Medema, F. Been, L. Heijnen, and S. Petterson. "Implementation of Environmental Surveillance for SARS-CoV-2 Virus to Support Public Health Decisions: Opportunities and Challenges". *Current Opinion in Environmental Science & Health* 17 (2020), pp. 49–71. ISSN: 2468-5844. DOI: 10.1016/j.coesh.2020.09.006.

[371] M. Nauta, O. McManus, K. Træholt Franck, et al. "Early Detection of Local SARS-CoV-2 Outbreaks by Wastewater Surveillance: A Feasibility Study". *Epidemiology and Infection* 151 (2023). ISSN: 0950-2688, 1469-4409. DOI: 10.1017/s0950268823000146.

[372] A. Lustig, G. Vattiato, O. Maclaren, et al. "Modelling the Impact of the Omicron BA.5 Subvariant in New Zealand". *Journal of The Royal Society Interface* 20.199 (2023), p. 20220698. DOI: 10.1098/rsif.2022.0698.

[373] J. A. Backer, D. Eggink, S. P. Andeweg, et al. "Shorter Serial Intervals in SARS-CoV-2 Cases with Omicron BA.1 Variant Compared with Delta Variant, the Netherlands, 13 to 26 December 2021". *Eurosurveillance* 27.6 (2022), p. 2200042. ISSN: 1560-7917. DOI: 10.2807/1560-7917.ES.2022.27.6.2200042.

[374] S. Abbott, K. Sherratt, M. Gerstung, and S. Funk. "Estimation of the Test to Test Distribution as a Proxy for Generation Interval Distribution for the Omicron Variant in England". *medRxiv* (2022). DOI: 10.1101/2022.01.08.22268920. Pre-published.

[375] D. Kim, S. T. Ali, S. Kim, et al. "Estimation of Serial Interval and Reproduction Number to Quantify the Transmissibility of SARS-CoV-2 Omicron Variant in South Korea". *Viruses* 14.3 (2022), p. 533. ISSN: 1999-4915. DOI: 10.3390/v14030533.

[376] J. Douglas, D. Winter, A. McNeill, et al. "Tracing the International Arrivals of SARS-CoV-2 Omicron Variants after Aotearoa New Zealand Reopened Its Border". *Nature Communications* 13.1 (2022), p. 6484. ISSN: 2041-1723. DOI: 10.1038/s41467-022-34186-9.

[377] P. Kilaru, D. Hill, K. Anderson, et al. "Wastewater Surveillance for Infectious Disease: A Systematic Review". *American Journal of Epidemiology* 192.2 (2023), pp. 305–322. ISSN: 0002-9262. DOI: 10.1093/aje/kwac175.

[378] M. Bunce, J. L. Geoghegan, D. Winter, et al. "Exploring the Depth and Breadth of the Genomics Toolbox during the COVID-19 Pandemic: Insights from Aotearoa New Zealand". *BMC Medicine* 21.1 (2023). ISSN: 1741-7015. DOI: 10.1186/s12916-023-02909-4.

[379] D. Toribio-Avedillo, C. Gómez-Gómez, L. Sala-Comorera, et al. "Monitoring Influenza and Respiratory Syncytial Virus in Wastewater. Beyond COVID-19". *Science of The Total Environment* 892 (2023), p. 164495. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2023.164495.

[380] S. M. Prasek, I. L. Pepper, G. K. Innes, et al. "Variant-Specific SARS-CoV-2 Shedding Rates in Wastewater". *Science of The Total Environment* 857 (2023), p. 159165. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2022.159165.

[381] S. Riley, K. E. C. Ainslie, O. Eales, et al. "Resurgence of SARS-CoV-2: Detection by Community Viral Surveillance". *Science* 372.6545 (2021), pp. 990–995. DOI: 10.1126/science.abf0874.

[382] Q. S. Huang, T. Wood, L. Jelley, et al. "Impact of the COVID-19 Nonpharmaceutical Interventions on Influenza and Other Respiratory Viral Infections in New Zealand". *Nature Communications* 12.1 (2021), p. 1001. ISSN: 2041-1723. DOI: 10.1038/s41467-021-21157-9.

[383] M. C. Zambon, J. D. Stockton, J. P. Clewley, and D. M. Fleming. "Contribution of Influenza and Respiratory Syncytial Virus to Community Cases of Influenza-like Illness: An Observational Study". *The Lancet* 358.9291 (2001), pp. 1410–1416. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(01)06528-X.

[384] O. Eales, M. J. Plank, B. J. Cowling, et al. "Key Challenges for Respiratory Virus Surveillance While Transitioning out of Acute Phase of COVID-19 Pandemic". *Emerging Infectious Diseases* 30.2 (2024). ISSN: 1080-6040, 1080-6059. DOI: 10.3201/eid3002.230768.

[385] H. Schenk, P. Heidinger, H. Insam, et al. "Prediction of Hospitalisations Based on Wastewater-Based SARS-CoV-2 Epidemiology". *Science of The Total Environment* 873 (2023), p. 162149. ISSN: 0048-9697. DOI: 10.1016/j.scitotenv.2023.162149.

[386] M. J. Plank, L. Watson, and O. J. Maclaren. "Near-Term Forecasting of Covid-19 Cases and Hospitalisations in Aotearoa New Zealand". *PLOS Computational Biology* 20.1 (2024). Ed. by A. Perkins, e1011752. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1011752.

[387] R. Moss, A. Zarebski, P. Dawson, and J. M. McCaw. "Retrospective Forecasting of the 2010–2014 Melbourne Influenza Seasons Using Multiple Surveillance Systems". *Epidemiology & Infection* 145.1 (2017), pp. 156–169. ISSN: 0950-2688, 1469-4409. DOI: 10.1017/S0950268816002053.

[388] E. S. Knock, L. K. Whittles, J. A. Lees, et al. "Key Epidemiological Drivers and Impact of Interventions in the 2020 SARS-CoV-2 Epidemic in England". *Science Translational Medicine* (2021). DOI: 10.1126/scitranslmed.abg4262.

[389] S. Riley, O. Eales, C. E. Walters, et al. "REACT-1 Round 8 Final Report: High Average Prevalence with Regional Heterogeneity of Trends in SARS-CoV-2 Infection in the Community in England during January 2021". *medRxiv* (2021). DOI: 10.1101/2021.01.28.21250606. Pre-published.

[390] P. Elliott, B. Bodinier, O. Eales, et al. "Rapid Increase in Omicron Infections in England during December 2021: REACT-1 Study". *Science* 375.6587 (2022), pp. 1406–1411. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abn8347.

[391] S. N. Williams, C. J. Armitage, T. Tampe, and K. Dienes. "Public Perceptions and Experiences of Social Distancing and Social Isolation during the COVID-19 Pandemic: A UK-based Focus Group Study". *BMJ Open* 10.7 (2020), e039334. ISSN: 2044-6055, 2044-6055. DOI: 10.1136/bmjopen-2020-039334.

[392] B. T. Snoeijer, M. Burger, S. Sun, et al. "Measuring the Effect of Non-Pharmaceutical Interventions (NPIs) on Mobility during the COVID-19 Pandemic Using Global Mobility Data". *npj Digital Medicine* 4.1 (2021), pp. 1–12. ISSN: 2398-6352. DOI: 10.1038/s41746-021-00451-2.

[393] N. Askitas, K. Tatsiramos, and B. Verheyden. "Estimating Worldwide Effects of Non-Pharmaceutical Interventions on COVID-19 Incidence and Population Mobility Patterns Using a Multiple-Event Study". *Scientific Reports* 11.1 (2021), p. 1972. ISSN: 2045-2322. DOI: 10.1038/s41598-021-81442-x.

[394] D. McLennan, S. Noble, M. Noble, et al. *The English Indices of Deprivation 2019*. Ministry of Housing, Communities, and Local Government, (2019). URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/833951/IoD2019_Technical_Report.pdf.

[395] L. Breiman, A. Cutler, A. Liaw, and M. Wiener. *randomForest: Breiman and Cutlers Random Forests for Classification and Regression*. Version 4.7-1.2. (2024). URL: https://cran.r-project.org/web/packages/randomForest/index.html.

[396] Prime Minister's Office. *Prime Minister Confirms Move to Plan B in England*. GOV.UK. (2021). URL: https://www.gov.uk/government/news/prime-minister-confirms-move-to-plan-b-in-england.

[397] H. Seale, C. E. F. Dyer, I. Abdi, et al. "Improving the Impact of Non-Pharmaceutical Interventions during COVID-19: Examining the Factors That Influence Engagement and the Impact on Individuals". *BMC Infectious Diseases* 20.1 (2020), p. 607. ISSN: 1471-2334. DOI: 10.1186/s12879-020-05340-9.

[398] P. G. Hansen, E. G. Larsen, and C. D. Gundersen. "Reporting on One's Behavior: A Survey Experiment on the Nonvalidity of Self-Reported COVID-19 Hygiene-Relevant Routine Behaviors". *Behavioural Public Policy* 6.1 (2022), pp. 34–51. ISSN: 2398-063X, 2398-0648. DOI: 10.1017/bpp.2021.13.

[399] R. Davies, F. Mowbray, A. F. Martin, et al. "A Systematic Review of Observational Methods Used to Quantify Personal Protective Behaviours among Members of the Public during the COVID-19 Pandemic, and the Concordance between Observational and Self-Report Measures in Infectious Disease Health Protection". *BMC Public Health* 22.1 (2022), p. 1436. ISSN: 1471-2458. DOI: 10.1186/s12889-022-13819-0.

[400] R. Davies, J. Weinman, and G. J. Rubin. "Observed and Self-Reported COVID-19 Health Protection Behaviours on a University Campus and the Impact of a Single Simple Intervention". *Journal of Public Health* 45.3 (2023), pp. 676–679. ISSN: 1741-3842. DOI: 10.1093/pubmed/fdac147.

[401] M. Savage, R. McKie, and P. Inman. "From Mere Advice to Full Lockdown: The Week When It All Changed". *The Guardian. World news* (2020). ISSN: 0261-3077. URL: https://www.theguardian.com/world/2020/mar/22/from-mere-advice-to-full-lockdown-the-week-when-it-all-changed-fast.

[402] L. E. Smith, H. W. W. Potts, R. Amlot, et al. "Engagement with Protective Behaviours in the UK during the COVID-19 Pandemic: A Series of Cross-Sectional Surveys (the COVID-19 Rapid Survey of Adherence to Interventions and Responses [CORSAIR] Study)". *BMC Public Health* 22.1 (2022), p. 475. ISSN: 1471-2458. DOI: 10.1186/s12889-022-12777-x.

[403] L. Wright, A. Steptoe, and D. Fancourt. "Trajectories of Compliance With COVID-19 Related Guidelines: Longitudinal Analyses of 50,000 UK Adults". *Annals of Behavioral Medicine* 56.8 (2022), pp. 781–790. ISSN: 0883-6612. DOI: 10.1093/abm/kaac023.

[404] F. V. Surano, M. Porfiri, and A. Rizzo. "Analysis of Lockdown Perception in the United States during the COVID-19 Pandemic". *The European Physical Journal Special Topics* 231.9 (2022), pp. 1625–1633. ISSN: 1951-6355, 1951-6401. DOI: 10.1140/epjs/s11734-021-00265-z.

[405] World Health Organization. *Avian Influenza A(H5N1) – United States of America*. (2024). URL: https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON512.

[406] UK Health Security Agency. *WHO Declares Mpox Outbreak a Public Health Emergency of International Concern*. GOV.UK. (2024). URL: https://www.gov.uk/government/news/who-declares-mpox-outbreak-a-public-health-emergency-of-international-concern.

[407] E. C. Goligher, A. Heath, and M. O. Harhay. "Bayesian Statistics for Clinical Research". *The Lancet* 404.10457 (2024), pp. 1067–1076. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(24)01295-9.

[408] D. Freedman. "Wald Lecture: On the Bernstein-von Mises Theorem with Infinite-Dimensional Parameters". *The Annals of Statistics* 27.4 (1999), pp. 1119–1141. ISSN: 0090-5364, 2168-8966. DOI: 10.1214/aos/1017938917.

[409] B. J. K. Kleijn and A. W. van der Vaart. "The Bernstein-Von-Mises Theorem under Misspecification". *Electronic Journal of Statistics* 6 (none 2012), pp. 354–381. ISSN: 1935-7524, 1935-7524. DOI: 10.1214/12-EJS675.

[410] R. T. Thibault, O. B. Amaral, F. Argolo, et al. "Open Science 2.0: Towards a Truly Collaborative Research Ecosystem". *PLOS Biology* 21.10 (2023), e3002362. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.3002362.

[411] B. Mlodozeniec, D. Krueger, and R. Turner. "Implicitly Bayesian Prediction Rules in Deep Learning". *Proceedings of the 6th Symposium on Advances in Approximate Bayesian Inference*. Vol. 253. Proceedings of Machine Learning Research. PMLR, (2024), pp. 79–110. URL: https://proceedings.mlr.press/v253/mlodozeniec24a.html.

[412] A. N. Angelopoulos and S. Bates. "Conformal Prediction: A Gentle Introduction". *Foundations and Trends® in Machine Learning* 16.4 (2023), pp. 494–591. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000101.

[413] S. T. Ali, L. Wang, E. H. Y. Lau, et al. "Serial Interval of SARS-CoV-2 Was Shortened over Time by Nonpharmaceutical Interventions". *Science* 369.6507 (2020), pp. 1106–1109. DOI: 10.1126/science.abc9004.

[414] Z. J. Madewell, Y. Yang, I. M. Longini, et al. "Rapid Review and Meta-Analysis of Serial Intervals for SARS-CoV-2 Delta and Omicron Variants". *BMC Infectious Diseases* 23.1 (2023), p. 429. ISSN: 1471-2334. DOI: 10.1186/s12879-023-08407-5.

[415] K. Ito, C. Piantham, and H. Nishiura. "Estimating Relative Generation Times and Reproduction Numbers of Omicron BA.1 and BA.2 with Respect to Delta Variant in Denmark". *Mathematical Biosciences and Engineering* 19.9 (2022), pp. 9005–9017. ISSN: 1551-0018. DOI: 10.3934/mbe.2022418.

[416] J. L. Juul, K. Græsbøll, L. E. Christiansen, and S. Lehmann. "Fixed-Time Descriptive Statistics Underestimate Extremes of Epidemic Curve Ensembles". *Nature Physics* 17.1 (2021), pp. 5–8. ISSN: 1745-2481. DOI: 10.1038/s41567-020-01121-y.

[417] M. Plummer, N. Best, K. Cowles, et al. *Coda: Output Analysis and Diagnostics for MCMC*. Version 0.19-4.1. (2024). URL: https://cran.r-project.org/web/packages/coda/index.html.

Appendices

Appendix A

Supplementary material for Chapter 4: Robust uncertainty quantification in popular estimators of the instantaneous reproduction number

A.1 Derivations

A.1.1 General notes

Before considering specific derivations, we highlight some useful tools that are common between estimators.

Predictive decomposition of the likelihood

Letting θ be the parameter(s) of interest (typically k for EpiEstim or η for EpiFilter), the log-likelihood of θ at time-step t is defined as:

$$\ell(\theta|C_{1:t}) = \log P(C_{1:t}|\theta). \tag{A.1}$$

This can be decomposed into the sum of the one-step-ahead predictive log-likelihoods:

$$\ell(\theta|C_{1:t}) = \sum_{s=1}^{t} \log P(C_s|C_{1:s-1}, \theta). \tag{A.2}$$

That is, the likelihood of observing all data $C_{1:t}$ is equal to the product of the likelihoods of observing each data point C_s given the data up to that point $C_{1:s-1}$.

Grid-based approximations

We frequently leverage grid-based approximations to various distributions in EpiFilter, and to a lesser extent in EpiEstim. These are necessary (in some cases) when considering three specific variables: the reproduction number R_t, EpiFilter's smoothing η, and predictive cases C_{t+1}.

We denote the grid of values for R_t using $\mathcal{R}$, typically using one of length $|\mathcal{R}| = 1000$:

$$\mathcal{R} = \{0.01, 0.02, \ldots, 10.000\}.$$

We denote the grid of values for η using $\mathcal{E}$, typically using one of length $|\mathcal{E}| = 1000$:

$$\mathcal{E} = \{0.001, 0.002, \ldots, 1.000\}.$$

Finally, we denote the grid of values for predictive cases C_{t+1} using $\mathcal{C}$, and typically use:

$$\mathcal{C} = \{0, 1, 2, \ldots, 10 \max C_{1:T}\},$$

where $10 \max C_{1:T}$ (where $C_{1:T}$ is the observed cases series) is chosen to be sufficiently large to capture the majority of the predictive distribution. This is checked by ensuring that the sum of the predictive distribution over $\mathcal{C}$ is close to 1.

A grid-based approximation to a distribution for R_t (for example) is then given by the set of values $\{P(R_t = r|\cdots)\}_{r\in\mathcal{R}}$. We test the sensitivity of our results to the choice of grids in Supplementary Section A.9.

Wind-in periods

Wind-in periods may be necessary for two reasons. Firstly, if a model is initialised mid-outbreak, then reported cases at time t (for low t) depend on reported cases prior to the initialisation of the model. A wind-in period is thus necessary to ensure a sufficient number of past data points are available to inform the model, the necessary duration of which depends on the specific application. Secondly, the derivation of a one-step-ahead likelihood requires the predictive distribution of R_{t+1}, which is derived from the posterior distribution of R_t, which requires at least one day of prior data. Thus, model likelihoods can only be calculated from time-step $t = 3$ onwards.

When fitting to data from the 2021 August COVID-19 outbreak in Auckland, New Zealand, the model is initialised alongside the first reported case, thus only a 3-day wind-in period is necessary. For simulated data, we assume that the model was initialised mid-outbreak, and thus use a 10-day wind-in period.

For simulated results, we also normalise the generation time distribution to sum to 1 over $[1, t]$, reducing the impact of a wind-in period that does not cover the entire support of the generation time distribution.

A.1.2 EpiEstim

Model description

EpiEstim assumes that R_t is fixed over the interval $[t-k+1, t]$. The likelihood of observing $C_{t-k+1:t}$ given R_t, k, and $C_{1:t-k}$ is then given by the product of the Poisson renewal model likelihood (Equation 4.1) for each day s in the interval $[t-k+1, t]$:

$$P(C_{t-k+1:t}|R_t, C_{1:t-k}, k) = \prod_{s=t-k+1}^{t} P(C_s|R_t, \Lambda_s) = \prod_{s=t-k+1}^{t} \frac{(R_t\Lambda_s)^{C_s} e^{-R_t\Lambda_s}}{C_s!}. \tag{A.3}$$

A conjugate (shape-rate-parameterised) Gamma(α, β) prior distribution is assumed for R_t, thus the posterior distribution for R_t given k and $C_{1:t}$ is also Gamma-distributed:

$$\begin{aligned} P(R_t|C_{1:t}, k) &\propto P(C_{t-k+1:t}|R_t, C_{1:t-k}, k)P(R_t) \\ &\propto \left(R_t^{\sum C_s} e^{-R_t \sum \Lambda_s}\right)\left(R_t^{\alpha-1} e^{-\beta R_t}\right) \\ &= R_t^{\alpha+\sum C_s-1} e^{-R_t(\beta+\sum \Lambda_s)} \\ &\sim \text{Gamma}\left(\alpha_{t,k} = \alpha + \sum_{s=t-k+1}^{t} C_s, \quad \beta_{t,k} = \beta + \sum_{s=t-k+1}^{t} \Lambda_s\right). \end{aligned} \tag{A.4}$$

The shape parameter $\alpha_{t,k}$ at time t is the shape parameter of the prior distribution α plus the sum of the previous k days' case counts. The rate parameter $\beta_{t,k}$ at time t is the rate parameter β of the prior distribution plus the sum of the previous k days' force of infection[1]. We make regular use of $\alpha_{t,k}$ and $\beta_{t,k}$ in the following derivations.

[1]Through this lens, EpiEstim's prior parameters α and β can be viewed as the assumed number of observed infections and force-of-infection before any data are collected.

Likelihood for k

We use the predictive decomposition of the likelihood (Equation A.2) to derive the likelihood of k given $C_{1:t}$. First note that:

$$P(C_s|C_{1:s-1},k) = \int P(C_s|R_s,C_{1:s-1},k)P(R_s|C_{1:s-1},k)dR_s. \tag{A.5}$$

The first term in the integral is the Poisson renewal model likelihood (Equation 4.1), and the second term is the predictive distribution for R_s given $C_{1:s-1}$ and k. We find this predictive distribution now.

Applying Bayes' theorem, leveraging EpiEstim's assumed independence of R_s from $C_{1:s-k}$, and substituting in the known distributions gives:

$$\begin{aligned} P(R_s|C_{1:s-1},k) &\propto P(C_{t-k+1:s-1}|R_s,C_{1:s-k},k)P(R_s) \\ &= P(R_s)\prod_{u=s-k+1}^{s-1} P(C_u|R_s,C_{1:u-1},k) \\ &\propto R_s^{\alpha-1}e^{-\beta R_s}\prod_{u=s-k+1}^{s-1} R_s^{C_u}e^{-R_s\Lambda_s} \\ &= R_s^{\alpha+\sum_{u=s-k+1}^{s-1}C_u-1}e^{-R_s\left(\beta+\sum_{u=s-k+1}^{s-1}\Lambda_u\right)}. \end{aligned} \tag{A.6}$$

Thus, the predictive distribution for R_s given $C_{1:s-1}$ and k is Gamma-distributed with shape parameter $\alpha_{s-1,k-1}$ and rate parameter $\beta_{s-1,k-1}$ (where $\alpha_{s-1,k-1}$ and $\beta_{s-1,k-1}$ are once again defined in Equation 4.2).

This is not a surprising result. By (EpiEstim's) definition, R_s is only informed by data between time-steps $s-k+1$ and s, however the predictive distribution for R_s must not include C_s, so the parameters of the posterior distribution for R_s are the sum over $k-1$ days of data. This is the key difference between our approach to "selecting" k and that advocated by the accumulated prediction error (APE) approach of [242] (Supplementary Section A.8), where the APE approach temporarily assumes that R_s is fixed for the preceding $k+1$ days, so that the predictive distribution can be informed by k-days of data.

The predictive distribution for C_s conditional on $C_{1:s-1}$ and k is a Poisson-Gamma mixture, and thus is negative binomial:

$$C_s|C_{1:s-1},k \sim \text{NegativeBinomial}\left(r=\alpha_{s-1,k-1}, p=\frac{\beta_{s-1,k-1}}{\Lambda_s+\beta_{s-1,k-1}}\right). \tag{A.7}$$

Finally, the log-likelihood of k given $C_{1:t}$ is given by the sum of the log-likelihoods of each day s in the interval $[t-k+1, t]$:

$$\ell(k|C_{1:t}) = \sum_{s=1}^{t} \log \text{NegBinPMF}\left(C_s | r = a_{s-1,k-1}, p = \frac{\beta_{s-1,k-1}}{\Lambda_s + \beta_{s-1,k-1}}\right). \tag{A.8}$$

Posterior distribution for k

The log-likelihood $\ell(k|C_{1:t})$ (Equation A.8) is sufficient for calculating quantities such as the maximum likelihood estimate of k. However, in order to marginalise over k, we require the posterior distribution for k given $C_{1:t}$.

We typically use a Uniform prior distribution over a sufficiently wide range of k values, say $k \in \{1, 2, \ldots, 30\}$, although other discrete prior distributions are possible. The posterior distribution for k is then proportional to the exponential of $\ell(k|C1:t)$:

$$P(k|C_{1:t}) = \begin{cases} c \exp\left(\ell(k|C_{1:t})\right) & k \in \{1, 2, \ldots, 30\} \\ 0 & \text{otherwise,} \end{cases} \tag{A.9}$$

where the normalising constant c is chosen such that $\sum_{k=1}^{30} P(k|C_{1:t}) = 1$. We note here that this posterior distribution is exact and does not rely upon any approximation.

Marginalising over k

To present estimates of R_t that do not depend on our choice of k, we marginalise over k using the posterior distribution derived above (Equation A.9):

$$P(R_t|C_{1:t}) = \sum_{k} P(R_t|C_{1:t}, k) P(k|C_{1:t}). \tag{A.10}$$

The first term in the summation is given by Equation A.4 and the second term is given by Equation A.9.

We typically calculate this over a grid $\mathcal{R}$ of R_t values as defined in Supplementary Section A.1.1. The resulting posterior mean and credible intervals for R_t are then calculated from this grid and are statements about R_t given the data $C_{1:t}$ (and implicitly the model structure), but independent of any specific choice of k.

Posterior predictive distribution

We previously derived the predictive distribution for C_t given $C_{1:t-1}$, conditional on some value of k (Equation A.7). We can also marginalise over k to obtain the predictive distribution for C_t given $C_{1:t-1}$:

$$P(C_t|C_{1:t-1}) = \sum_k P(C_t|C_{1:t-1}, k)P(k|C_{1:t-1}). \tag{A.11}$$

The first term in the summation is given by Equation A.7 and the second term is given by Equation A.9.

We typically calculate this over a grid $\mathcal{C}$ of potential C_t values as defined in Supplementary Section A.1.1.

Practical notes

At time-step $s = 1$, there is no past data to inform the predictive distribution for R_s (Equation A.6). Therefore, R_1 is independent of k and provides no information about the posterior distribution for k. When calculating these distributions, we typically begin related summations at $s = 2$.

A.1.3 EpiFilter

Model description

EpiFilter models the evolution of R_t using a Gaussian random walk with standard deviation $\eta\sqrt{R_{t-1}}$:

$$R_t|R_{t-1} \sim \text{Normal}\left(R_{t-1}, \eta\sqrt{R_{t-1}}\right). \tag{A.12}$$

Given the posterior distribution for R_{t-1} at time-step $t-1$, the posterior distribution for R_t is found by applying the Bayesian filtering equations. Specifically, we first find the update distribution:

$$P(R_t|C_{1:t-1}, \eta) = \int P(R_t|R_{t-1}, \eta)P(R_{t-1}|C_{1:t-1}, \eta)\, dR_{t-1}, \tag{A.13}$$

and then use the Poisson renewal likelihood to find the filtering distribution:

$$P(R_t|C_{1:t}, \eta) \propto P(C_t|R_t, C_{1:t-1})P(R_t|C_{1:t-1}, \eta). \tag{A.14}$$

Analytical forms for these distributions are not readily available, so they are approximated using a grid of R_t values $\mathcal{R}$. This replaces the integral in Equation A.13 with a summation

over $R_{t-1} \in \mathcal{R}$, and allows us to normalise the filtering distribution in Equation A.14 by summing over the grid.

Likelihood for η

We use a similar approach to EpiEstim to derive the likelihood of η given $C_{1:t}$. Re-writing Equation A.5 in terms of EpiFilter's η:

$$P(R_s|C_{1:s-1},\eta) = \int P(C_s|R_s, C_{1:s-1},\eta)P(R_s|C_{1:s-1},\eta)dR_s. \tag{A.15}$$

The first term in the integral is the Poisson renewal model likelihood (Equation 4.1), and the second term is the update distribution for R_s given $C_{1:s-1}$ and η, which is a by-product from solving the filtering equations (specifically, Equation A.13). Leveraging the grid-based approximation, we write this as:

$$P(R_s|C_{1:s-1},\eta) = \sum_{r\in\mathcal{R}} P(R_s|R_{s-1}=r,\eta)P(R_{s-1}=r|C_{1:s-1},\eta). \tag{A.16}$$

Then the probability of observing C_s given $C_{1:s-1}$ and η is found by marginalising over R_s:

$$P(C_s|C_{1:s-1},\eta) = \sum_{r\in\mathcal{R}} P(C_s|R_s=r,C_{1:s-1})P(R_s=r|C_{1:s-1},\eta). \tag{A.17}$$

Thus, the log-likelihood of η at time-step t can be written as:

$$\ell(\eta|C_{1:t}) = \sum_{s=1}^{t} \log\left(\sum_{r\in\mathcal{R}} P(C_s|R_s=r,C_{1:s-1})\left[\sum_{r'\in\mathcal{R}} P(R_s=r|R_{s-1}=r',\eta)P(R_{s-1}=r'|C_{1:s-1},\eta)\right]\right). \tag{A.18}$$

Posterior distribution for η

The log-likelihood $\ell(\eta|C_{1:t})$ (Equation A.18) is sufficient for calculating quantities such as the maximum likelihood estimate of η. However, in order to marginalise over η, we require the posterior distribution for η given $C_{1:t}$.

We typically use a Uniform prior distribution over a sufficiently wide range of η values, say $\eta \in (0, 0.5)$. The posterior distribution for η is then proportional to the exponential of $\ell(\eta|C_{1:t})$:

$$P(\eta|C_{1:t}) = \begin{cases} c\exp\left(\ell(\eta|C_{1:t})\right) & \eta \in (0, 0.5) \\ 0 & \text{otherwise.} \end{cases} \tag{A.19}$$

As no analytical form for this distribution is available, we use another grid-approximation (this time over $\eta \in \mathcal{E}$), chosen such that $\sum_{\eta\in\mathcal{E}} P(\eta|C_{1:t}) = 1$. This posterior distribution is not exact in the same sense as the equivalent distribution for k in EpiEstim, but we show that it is a good approximation for sufficiently large $|\mathcal{E}|$ (Supplementary Section A.1.1).

Marginalising over η

To present estimates of R_t that do not depend on our choice of η, we marginalise over η using the posterior distribution derived above (Equation A.19):

$$P(R_t|C_{1:t}) = \sum_{\eta \in \mathcal{E}} P(R_t|C_{1:t}, \eta) P(\eta|C_{1:t}). \tag{A.20}$$

The first term in the summation is given by Equation A.14 and the second term is given by Equation A.19. We typically calculate this over a grid $\mathcal{R}$ of R_t values.

Posterior predictive distribution

We marginalise over η in Equation A.17 to find the marginal posterior predictive distribution:

$$P(C_t|C_{1:t-1}) = \sum_{\eta \in \mathcal{E}} P(C_t|C_{1:t-1}, \eta) P(\eta|C_{1:t-1}). \tag{A.21}$$

The first term in the summation is given by Equation A.17 and the second term is given by Equation A.19.

Marginal smoothing posterior distribution

We focus on real-time analysis in this paper, and thus focus on EpiFilter's *filtering distribution*, the posterior distribution for R_t given data up-to time t. The full implementation of EpiFilter also includes a *smoothing distribution* that incorporates both past and future data in estimates of R_t. It is easy to marginalise out η from this smoothing distribution:

$$P(R_t|C_{1:T}) = \int P(R_t|C_{1:T}, \eta) P(\eta|C_{1:T}) \, d\eta, \tag{A.22}$$

where $P(\eta|C_{1:T})$ is simply the posterior distribution for η at the final time-step T. That is, the predictive decomposition of the likelihood produces the correct likelihood for η regardless of whether we are considering the filtering or smoothing distribution.

This marginal smoothing posterior distribution is used and compared with other methods in Supplementary Section A.2.

A.1.4 Continuous ranked probability score

The continuous ranked probability score (CRPS) is a measure of the accuracy of a probabilistic forecast. For a given time-step t, the CRPS is defined as:

$$CRPS(F_t, C_t) = \int (F_t(y) - \mathbb{I}(y \geq C_t))^2 dy, \tag{A.23}$$

where $F_t(y)$ is the predictive distribution for C_t at time t, and $\mathbb{I}(y \geq C_t)$ is the indicator function that is 1 if $y \geq C_t$ and 0 otherwise. The CRPS is a measure of the distance between the predictive distribution and the observed value. The CRPS for a model is the average CRPS over all time-steps in the dataset, with lower scores representing better forecasts.

The CRPS is a strictly proper scoring rule, meaning that the optimal forecast is the true distribution. It is also a strictly consistent scoring rule, meaning that the optimal forecast converges to the true distribution as the number of observations increases.

We calculate the CRPS using the following formula:

$$CRPS(F_t, C_t) = \sum_{y \in \mathcal{C}} (F_t(y) - \mathbb{I}(y \geq C_t))^2, \tag{A.24}$$

where $F_t(y)$ is the CDF for the predictive distribution of C_t (found either analytically in the case of default EpiEstim, or by taking the cumulative sum over the grid-approximation in the case of marginalised EpiEstim and EpiFilter).

CRPS should only be compared between models that have been fit to the same data, as it is a measure of the accuracy of a forecast, not the quality of a model. In Supplementary Section A.5 we average results over multiple simulated epidemics. To ensure that CRPS values are comparable, in this section, we report the average of CRPS scores relative to the default EpiEstim model.

A.2 Smoothing posterior distributions and comparisons with additional methods

The material in this supplementary section has been published as a live document online. We provide a copy here for record, although the reader may find the online document more up-to-date.

A.2.1 Background

Many state-of-the-art R_t estimation methods target the smoothing distribution, rather than the filtering distribution (see below). Examples include the smoothing version of EpiFilter [115], EpiNow2 [139], EpiLPS [235], and rtestim [124]. We compare and contrast these methods on simulated data here.

For each method, we:

- Find the posterior distribution for R_t
- Find the posterior predictive distribution for observed cases $\tilde{C}_t$
- Calculate the calibration of the implied 95% credible intervals for R_t and C_t
- Calculate the CRPS of the posterior predictive distribution for observed cases

We use $\tilde{C}_t$ to highlight when we are treating reported cases as a random variable, instead of as observed data.

Unlike the other methods listed, rtestim is frequentist in nature. In this case, we replace posterior means/modes and credible intervals with estimates and confidence intervals. All of the above metrics still apply and carry comparable interpretations.

The posterior predictive distribution obtained from smoothing methods typically provides within-sample estimates. That is, the observed value C_t is used to find the posterior predictive distribution of $\tilde{C}_t$. This can occur either indirectly (observed C_t informs R_t estimates, which then are used to predict $\tilde{C}_t$) or directly (observed C_t informs estimates of latent infection incidence, from which $\tilde{C}_t$ are generated). This changes the interpretation of CRPS from a score of future predictions to a score of how well the data-generating mechanism is modelled.

In many cases, existing methodology had to be extended to find the posterior predictive distributions or to calculate the CRPS. Model descriptions are otherwise kept to a minimum. Where possible, default options of these models are used, although minor modifications are made where defaults lead to obviously suboptimal results.

Code to explicitly reproduce these results and figures is provided on GitHub.

Filtering versus smoothing

The **filtering posterior distribution** depends only on past data and is written $P(R_t|C_{1:t})$, whereas the **smoothing posterior distribution** depends on all data and is written $P(R_t|C_{1:T})$ [119].

When producing estimates at the most recent time-step, only the filtering distribution is accessible, as future data are not yet available, whereas the smoothing distribution is generally preferred for retrospective analysis, as it incorporates all available data. Note that, at time $t = T$, the filtering and smoothing distributions are equivalent.

In the main paper, we focus on methods for estimating R_t in real-time. That is, we target the filtering distribution rather than the smoothing distribution. In this supplementary section we consider the latter.

A.2.2 Methods

EpiFilter (smoothing)

We outline how EpiFilter can be used to find the marginal smoothing posterior distribution in Supplementary Section A.1.3.

The posterior smoothing distribution for observed case incidence is found by marginalising over R_t from the smoothing posterior distribution for R_t (instead of the predictive distribution for R_t):

$$P(\tilde{C}_t|C_{1:T}) = \int P(\tilde{C}_t|R_t, C_{1:T})P(R_t|C_{1:T})\ dR_t.$$

For simplicity, we use the renewal model for $P(\tilde{C}_t|R_t, C_{1:T})$, even though this ignores future reported case incidence. That is, future case data $C_{t:T}$ only features in $P(\tilde{C}_t|C_{1:T})$ via R_t.

EpiNow2

Rather than using a Gaussian random walk (as in EpiFilter) or fixed sliding windows (as in EpiEstim), EpiNow2 models the evolution of R_t using a Gaussian process. The default implementation then assumes that latent infection incidence follows a deterministic renewal model. Reported cases are assumed to be negative binomially distributed around the true infection, with a day-of-the-week effect that is estimated during the fitting process. That is, EpiNow2 accounts for both process noise (in the evolution of R_t) and observation noise (in the distribution for reported cases).

Smoothness in R_t is primarily controlled by the Gaussian process kernel (default Matérn 3/2, with lengthscale ℓ and magnitude α). In particular, prior assumptions about the lengthscale ℓ have a significant impact on the smoothness of resulting estimates. It is also expected that prior assumptions about the observation overdispersion will have a secondary effect on the smoothness of R_t estimates (as this impacts the trade-off between process and observation noise).

The EpiNow2 package provides pre-built functionality to extract central estimates and credible intervals for R_t and observed cases. As far as we are aware, the package does not provide a method to calculate the CRPS, so we provide our own sample-based implementation.

Samples from the posterior distribution for reported cases $\{x_t^{(i)}\}_{i=1}^N$ at time t are extracted from the Stan fit object. We then use the sample-based CRPS estimator:

$$\text{CRPS}_t = \frac{1}{N}\sum_{i=1}^{n}|x_t^{(i)} - C_t| - \frac{1}{2N^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_t^{(i)} - x_t^{(j)}|.$$

Calculating this for each time-step t and taking the average gives the CRPS for observed case incidence. Code to reproduce this is available on GitHub.

By default, EpiNow2 uses a log normal prior distribution for the lengthscale ℓ with mean 21 days and standard deviation 7 days. We also test an alternative and less informative inverse-gamma prior distribution provided with the package, which performs considerably better.

EpiLPS (MAP)

EpiLPS models latent infection incidence using Bayesian P-splines. Given infection incidence $\mu(t)$ and overdispersion parameter ρ, reported cases are assumed to be negative binomially distributed around $\mu(t)$. In the maximum a posteriori (MAP) version of EpiLPS, central estimates of R_t are calculated using a plug-in estimate of the posterior mean incidence $\hat{\mu}(t)$ and uncertainty is derived using a delta method. Like EpiNow2, EpiLPS explicitly accounts for both process noise (in the splines used to model infection incidence) and observation noise (in the distribution for reported cases).

Smoothness in R_t is primarily controlled by parameter λ, where larger values penalise sharp changes in infection incidence. A hierarchical prior distribution is assumed for λ. Prior assumptions about the overdispersion parameter ρ also likely impact the smoothness of R_t estimates, as this controls how much noise is associated with the observation process rather than the underlying epidemic process. The MAP version of EpiLPS selects optimal values of λ and ρ using an optimization routine.

While $K = 30$ is used as the default number of splines, we find that this leads to inaccurate inference on our example datasets. For our examples, we increase this to $K = 100$, which allows for more flexible inference.

We provide two extensions to the EpiLPS package:

1. Methods to sample from the posterior distribution for reported cases.
2. A CRPS estimator for the posterior distribution for reported cases.

Infection incidence is defined as $\mu(t) = \exp(\theta^T b(t))$, where θ is an estimated vector of spline coefficients and $b(t)$ are the basis functions evaluated at time t. The MAP version of EpiLPS uses a multivariate Gaussian approximation to θ at the MAP estimate of λ. We sample from this distribution:

$$\theta^{(i)} \sim MVN(\hat{\theta}, Q_\lambda^{-1}),$$

and use these samples to generate samples of latent infection incidence:

$$\mu^{(i)}(t) = \exp(\theta^{(i)T} b(t)),$$

from which samples of observed data are generated:

$$C_t^{(i)} \sim \text{NegBin}(\mu^{(i)}(t), \rho),$$

where ρ is the MAP value of the overdispersion parameter. This is repeated N times (default $N = 1000$) at each value of t. The mean of these N samples is reported as the central estimate, with 95% credible intervals obtained by taking the 2.5th and 97.5th percentile values.

The CRPS is calculated similarly, first by sampling $\mu^{(i)}(t)$ as above and then using the *crps_nbinom()* function from the *scoringRules* package in R to calculate the CRPS value for each sampled $\mu^{(i)}(t)$ and C_t. The mean of these N CRPS values is reported as the CRPS for the t^{th} observed case incidence.

Some code is available on GitHub, although copyright limitations prevent us from providing the full implementation, in which case we outline the steps required to reproduce the analysis.

EpiLPS (MALA)

The Metropolis-adjusted Langevin algorithm (MALA) version of EpiLPS replaces the Laplace approximation and optimisation routine with a full MCMC-type sampler. This has the advantage of returning a posterior distribution on λ and ρ, as well as properly marginalising out uncertainty about these quantities. As the true posterior distribution of spline coefficients θ is targeted, we also expect the posterior distributions for R_t and observed cases incidence to be more accurate. This comes at a cost of slightly increased computational complexity, although we do not find this to be prohibitive.

The default version of EpiLPS(MALA) does not return the MCMC sampling object. In order to access this, we created a customised version of *estimRmcmc.R* that includes *MCMC=MCMCout* in the *"outputlist"* of the function. This is the sole change required to this script, although running it requires having local copies of some files from the EpiLPS package. A list of the required files is given here.

We make the same two extensions to EpiLPS(MALA) as we did to EpiLPS(MAP):

1. Methods to sample from the posterior distribution for reported cases.
2. A CRPS estimator for the posterior distribution for reported cases.

Samples of reported case incidence are extracted from the MCMC sampling object. For each sample $i = 1, \ldots, N$ and time-step $t = 1, \ldots, T$ we sample:

$$C_t^{(i)} \sim \text{NegBin}(\exp(\theta^{(i)T} b(t)), \rho^{(i)}),$$

where $\theta^{(i)}$ and $\rho^{(i)}$ are samples from the MCMC sampling object. Note the use of $\rho^{(i)}$ instead of the MAP value of ρ, ensuring we are appropriately accounting for uncertainty in this parameter.

The CRPS is calculated from sampled $C_t^{(i)}$ using the same approach as for EpiNow2, relying upon the sample-based CRPS estimator (see above).

Some code is available on GitHub, although copyright limitations prevent us from providing the full implementation, in which case we outline the steps required to reproduce our analysis.

rtestim

In contrast to the aforementioned Bayesian methods, rtestim is grounded in a frequentist framework. R_t is modelled using piecewise cubic functions with ℓ_1 regularisation on the divided differences. This regularisation enforces sparsity in changes in R_t, allowing for locally adaptive smoothness, a key advantage over the other methods considered here that assume global smoothness.

A tuning parameter λ controls the strength of this regularisation, with larger values enforcing smoother estimates of R_t. The optimal value of this parameter is automatically chosen by rtestim using cross-validation.

rtestim uses the delta method to calculate confidence intervals. Built-in functions are provided to calculate these for R_t and observed case incidence, for any user-specified significance level. To match the other methods we use a 95% confidence level.

While the frequentist framework of rtestim does not admit a posterior distribution for observed case incidence, we can still use CRPS to measure how well the confidence intervals approximate the observed distribution of the data. To do this, we treat the bounds of the confidence intervals at different significance levels as defining quantiles of an empirical CDF. That is, at each time-step t, we find $x_t^{(i)}$ such that $q^{(i)} = F(x_i^{(t)})$ for $q^{(i)} = 0.01, 0.02, \ldots, 0.99$. We then approximate

the CRPS numerically at time-step t using the trapezoidal rule:

$$CRPS_t = \sum_{i=1}^{n-1} \frac{1}{2}\left(x_t^{(i+1)} - x_t^{(i)}\right)\left[\left(q^{(i+1)} - \mathbb{I}(x_t^{(i+1)} \geq y_t)\right)^2 + \left(q^{(i)} - \mathbb{I}(x_t^{(i)} \geq y_t)\right)^2\right].$$

Averaging $CRPS_t$ over all time-steps t gives the CRPS for observed case incidence. Code to reproduce this is available on GitHub.

Additional methodological notes

Each method handles the wind-in period differently, particularly when t is small compared to generation times. This can become fairly complicated, and is not the point of this supplementary section. When estimating or calculating coverage and CRPS, we use time-steps $t = 10, \ldots, T$ to avoid differences from this period impacting our results.

Posterior parameter values were calculated using a grid-based approach for EpiFilter. For EpiLPS (MALA) and EpiNow2, a sampling approach was required. The mean, 2.5th and 97.5th quantiles of these samples were calculated to estimate the posterior mean and 95% credible intervals. Posterior mode values were calculated using a kernel density estimate from the default *density()* function in R.

A.2.3 Results

Theoretical links between Gaussian random walks, Gaussian processes, splines, and piecewise polynomials, as well as the fact that all models use the same underlying renewal model for the epidemic process, suggest that the methods should perform similarly in practice. However, we find this not to be the case.

Table A.1 presents the coverage of R_t and reported cases, as well as the CRPS for reported cases for each method on each of the three simulations considered in the main text. Figure A.1 presents the corresponding estimates. While the methods produce mostly well-calibrated 95% uncertainty intervals for observed case data, coverage of R_t estimates varies significantly, as does the CRPS for observed case incidence.

The smoothed version of EpiFilter consistently outperforms (i.e. exhibits coverage of 95% credible intervals closest to 95%, and the lowest CRPS values) the other methods on these simulated datasets. This is expected in the random walk simulation, where the dynamic model

Table A.1: Coverage of 95% credible intervals (confidence intervals for rtestim) for R_t and observed case incidence, and CRPS for observed case incidence, for each method on each simulation.

Method	Rt coverage	Ct coverage	CRPS
Random walk simulation			
EpiFilter (smoothing)	93.4%	97.8%	4.69
EpiNow2 (default)	45.1%	98.9%	8.98
EpiNow2 (inv-gamma)	69.2%	97.8%	8.05
EpiLPS (MAP)	75.8%	100.0%	9.84
EpiLPS (MALA)	91.2%	97.8%	6.41
rtestim	100.0%	100.0%	10.15
Sinusoidal simulation			
EpiFilter (smoothing)	94.5%	100.0%	3.79
EpiNow2 (default)	73.6%	100.0%	14.24
EpiNow2 (inv-gamma)	76.9%	98.9%	5.62
EpiLPS (MAP)	91.2%	100.0%	9.15
EpiLPS (MALA)	92.3%	100.0%	4.85
rtestim	100.0%	100.0%	9.26
Step-change simulation			
EpiFilter (smoothing)	94.5%	100.0%	3.10
EpiNow2 (default)	75.8%	97.8%	4.95
EpiNow2 (inv-gamma)	80.2%	97.8%	4.65
EpiLPS (MAP)	85.7%	100.0%	5.71
EpiLPS (MALA)	89.0%	98.9%	3.82
rtestim	100.0%	100.0%	6.08

employed by EpiFilter matches the simulated data. The better performance of EpiFilter on the other simulations is likely explained by two factors: (1) EpiFilter does not account for observation noise (which is absent in the simulated data), and (2) the flat prior distribution on η is less informative than the default prior distributions used by other methods. These are both testable hypotheses that could be explored in future work to better understand the trade-offs between smoothing, model assumptions, and robustness of performance.

EpiLPS and EpiNow2 both assume that reported case incidence are negative binomially-distributed about some smooth true infection incidence (upon which the renewal model is placed), thus explicitly modelling observation noise. Process noise is derived from either the smoothness of the splines (EpiLPS) or the Gaussian process (EpiNow2). While the overdispersion parameter is estimated in both methods, the negative binomial distribution implies a

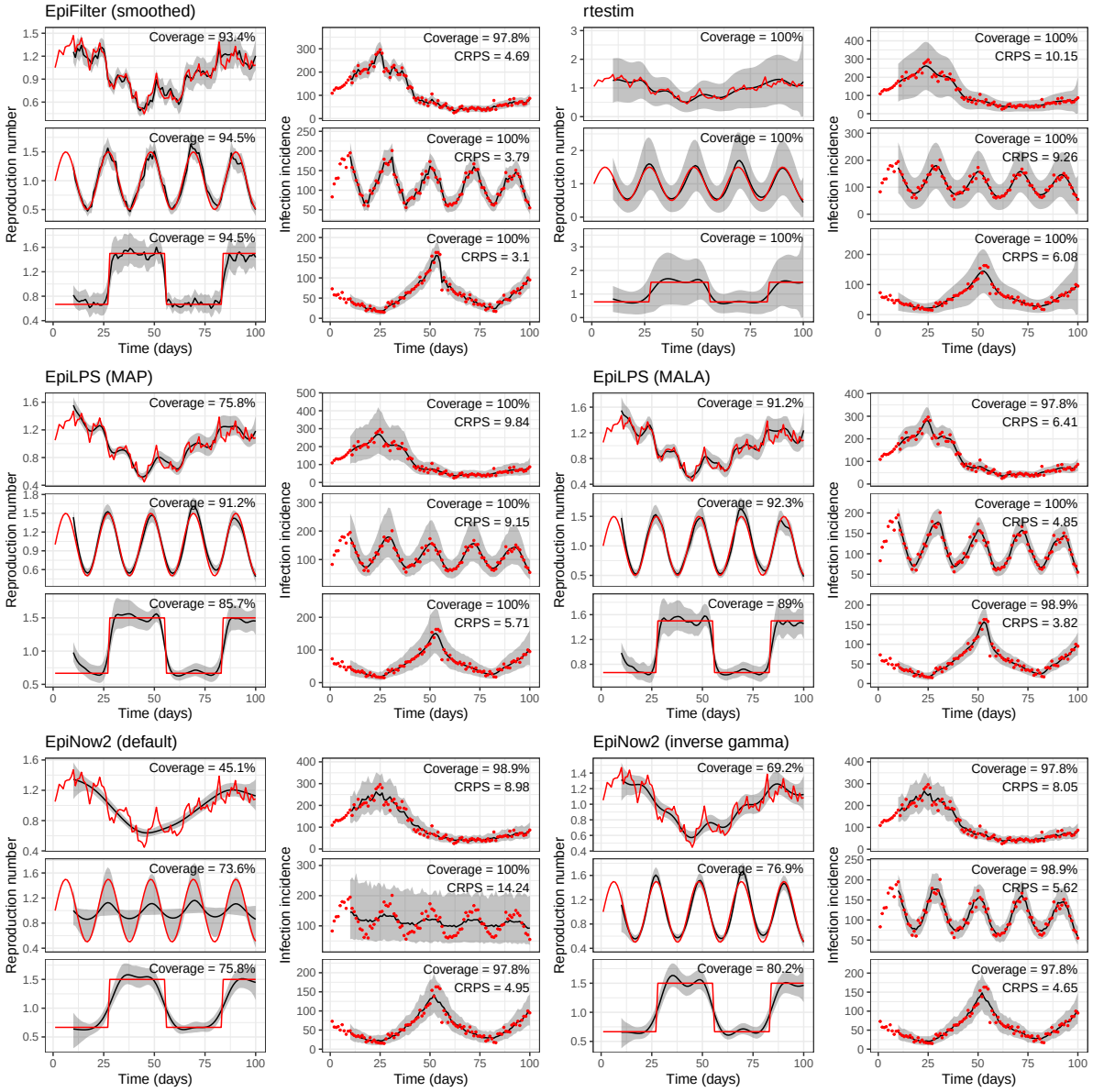

Figure A.1: Estimates of R_t and observed case incidence for each method on each simulation.

lower bound of Poisson observation noise. In the simulated data, however, all noise is assumed to arise from the underlying process. By enforcing the inclusion of observation noise, EpiLPS and EpiNow2 underestimate process noise, and thus produce overly smooth estimates of R_t. This is likely to be less impactful on real-world datasets, where observation noise is typically a significant factor.

It is possible to use CRPS as model selection criterion instead of cross-validation. Early code to do this is provided on GitHub.

The MALA version of EpiLPS outperforms the MAP version. This is unsurprising given that, in the MALA version, uncertainty associated with λ and ρ is fully marginalised out, whereas the MAP version selects optimal point estimates of these parameters. This lends support to the argument in the main paper for the marginalisation of these parameters over selection. The MALA version also targets the true posterior distribution for $\mu(t)$, whereas the MAP version uses a Laplace approximation, which is particularly advantageous when comparing CRPS values.

Figure A.1 also highlights critical oversmoothing in default EpiNow2 on the sinusoidal simulation. The posterior distribution for lengthscale ℓ is bi-modal, with one mode at a small value of ℓ (where the model follows the data) and another mode at a larger value of ℓ, where the model estimates flat incidence with large reporting noise accounting for changes in observed cases.

Prior distributions on smoothing parameters

While all methods listed estimate smoothing parameter(s) from the data, they handle this in different ways. Our implementation of EpiFilter, EpiNow2, and EpiLPS(MALA) marginalise this parameter out. EpiLPS(MAP) and rtestim select optimal point values of this parameter.

Of particular note is EpiNow2's prior on lengthscale ℓ. Prior and posterior modes and credible intervals are given in Table A.2. The default log-normal prior distribution often leads to over-smoothing, with an extreme example being the bimodal posterior on the sinusoidal simulation - where one mode implies that R_t is (nearly) fixed, and all variation in reported cases is assigned to the observation process.

By default, we place a uniform prior distribution on EpiFilter's η parameter. While we avoid claiming that the uniform prior distribution is uninformative, we do highlight a key advantage of this: the MAP is equal to the MLE. By the predictive decomposition of the likelihood, this is also the value of η that optimises one-step-ahead predictions. That is, using a flat prior distribution enforces a MAP that is optimal for one-step-ahead forecasting. Furthermore, if the model is correctly specified, this implies that the MAP is optimal for any n-step-ahead forecast.

Table A.2: MAP and 95% credible intervals for the parameter prior and posterior distributions, where they exist. The upper portion of the table refers to smoothing parameters and the lower portion refers to estimated overdispersion. Numerical instabilities lead to large variations in statistics in some cases which are marked by a "-". As rtestim is frequentist, there are no prior assumptions on λ, and the values are chosen by cross-validation instead of being MAP estimates.

Parameter	Prior Mode (95% Cr.I)	Posterior mode and 95% Cr.I. / rtestim optimal value Random walk	 Sinusoidal	 Step-change
EpiFilter η	U(0,1) *flat* (0.025, 0.975)	0.11 (0.08, 0.15)	0.16 (0.13, 0.19)	0.14 (0.11, 0.18)
EpiNow2 ℓ	LogN($\mu = 21$, $\sigma = 7$) 17.9 (10.5, 37.6)	16.9 (9.64, 31.7)	7.07 (5.33, 34.3)	10 (6.43, 21.6)
EpiNow2 (inv-gam) ℓ	InvGam(1.5, 3.4) 1.38 (0.736, 31.9)	1.35 (0.641, 7.46)	1.99 (1.1, 4.77)	1.77 (0.852, 6.4)
EpiLPS(MAP) λ	*Hierarchical* 0 (0, 3.48)	129	77.1	99.4
EpiLPS(MALA) λ	*Hierarchical* 0 (0, 3.48)	103 (55.5, 380)	76.5 (44.5, 189)	79.9 (29.6, 468)
rtestim λ	Not applicable	1.41	1.41	1.21
EpiNow2 ϕ	$\phi^{-1/2} \sim HalfN(0,1)$ -	65.6 (42.3, 124)	0 (7.37, 10200)	68.5 (41.7, 253)
EpiNow2 (inv-gam) ϕ	$\phi^{-1/2} \sim HalfN(0,1)$ -	86 (52.6, 199)	0 (205, 108000)	0 (54, 718)
EpiLPS(MAP) ρ	$\Gamma(10^{-4}, 10^{-4})$ -	25.1	25.2	25.2
EpiLPS(MALA) ρ	$\Gamma(10^{-4}, 10^{-4})$ -	180 (84.4, 4070)	937 (313, 17500)	243 (105, 14200)

A.3 Simulated data

We test our methods on simulated data from three dynamic models.

Gaussian random walk

The first model is a Gaussian random walk with standard deviation $\eta\sqrt{R_{t-1}}$, matching the dynamic model assumed by EpiFilter. Simulations are initialised with default values of $R_0 = 1$, $C_{:0} = 100$, and $\eta = 0.1$. To simulate $C_{1:100}$, we iteratively sample $R_t \sim N(R_{t-1}, \eta\sqrt{R_{t-1}})$ and $C_t \sim \text{Poisson}(R_t\Lambda_t)$.

To ensure that simulations are realistic, we reject and re-simulate any simulated epidemic where daily cases exceed 5000 or fall below 5. We also truncate R_t on $[0.1, 5.0]$ to ensure that results are not biased by unrealistic values of R_t.

Sinusoidal model

The second model is a sinusoidal model, which assumes that R_t follows a deterministic sinusoidal function of time. Default values are a period of $\omega = 21$ days, an amplitude of $A = 0.5$, an initial value of $R_0 = 1$, and initial cases $C_0 = 100$. On day t, R_t is assumed to be:

$$R_t^{sin} = R_0 + A\sin\left(\frac{2\pi}{\omega}t\right), \tag{A.25}$$

and C_t is then sampled from the Poisson renewal model with rate $R_t^{sin}\Lambda_t$.

Step-change model

The third model is a step-change model, which assumes that R_t alternates between two values R_a and R_b every T_{step} days. Default values are $R_a = 0.67$, $R_b = 1.5$, $T_{step} = 28$ days, and initial cases $C_0 = 100$. Depending on t, C_t is sampled from the Poisson renewal model with rate $R_a\Lambda_t$ or $R_b\Lambda_t$.

Additional details

To ensure that simulations are realistic, we assume that cases prior to $t = 0$ are all equal to C_0. This ensures the calculation of Λ_t is well-defined (i.e. there are past infections covering the entire generation time distribution). These assumed prior cases are not included in the simulated data, and are only used to calculate Λ_t.

While we do not consider this in the main paper, we do consider observation noise in the supplementary material. This is simulated by assuming some under-reporting rate p (default $p = 0.5$), initialising the model with $C_0 = 100/p$, generating interim cases $\tilde{C}_{1:T}$ as above, and sampling $C_t \sim \text{Binomial}(\tilde{C}_t, p)$. Additional reporting noise is included by sampling from a beta-binomial distribution with $\alpha = Np$ and $\beta = N(1-p)$, where N controls the amount of over-dispersion in reporting (default $N = 100$).

A.4 Serial intervals

A.4.1 Serial interval uncertainty

We have thus far focused on uncertainty in the smoothing parameter while assuming that the serial interval is known. Here we demonstrate how these methods can be extended to also account for uncertainty in the serial interval.

Letting θ denote an arbitrary smoothing parameter and ϕ denote the serial interval, we target the following posterior distribution for R_t:

$$P(R_t|C_{1:t}) = \int P(R_t|C_{1:t}, \phi)P(\phi)\ d\phi. \tag{A.26}$$

Until now, we have been implicitly targeting $P(R_t|C_{1:t}, \phi)$, even if we haven't been explicitly writing ϕ in the conditional.

We can write this as an expectation which can be approximated by Monte Carlo methods:

$$\begin{aligned} P(R_t|C_{1:t}) &= E_\phi \left[\int P(R_t|C_{1:t}, \phi, \theta)P(\theta|C_{1:t}, \phi)\ d\theta \right] \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \left[\int P(R_t|C_{1:t}, \phi_i, \theta)P(\theta|C_{1:t}, \phi_i)\ d\theta \right], \quad \phi_i \sim P(\phi). \end{aligned} \tag{A.27}$$

A similar averaging approach is used to develop the marginal posterior distribution for predictive cases.

We demonstrate these methods by assuming that ϕ represents a random Gamma distribution with shape and scale parameters:

$$\alpha = 2.3669\xi, \theta = 2.7463\xi, \quad \text{where } \xi^2 \sim N(1, 0.167)$$

$E[\phi]$ is a Gamma distribution with mean 6.5 days and standard deviation 4.2 days, with 99% of the mass on a mean between 3.25 and 13 days (0.5 and 1.5 times the expected mean).

Figure A.2 demonstrates the effect of applying Equation A.27 to the random walk simulation, using 50 draws from the random Gamma distribution described above for ϕ. As $E[\phi]$ matches the true serial interval distribution pointwise, the error averages out, and we obtain (to Monte Carlo error), matching results whether we include uncertainty about ϕ or not.

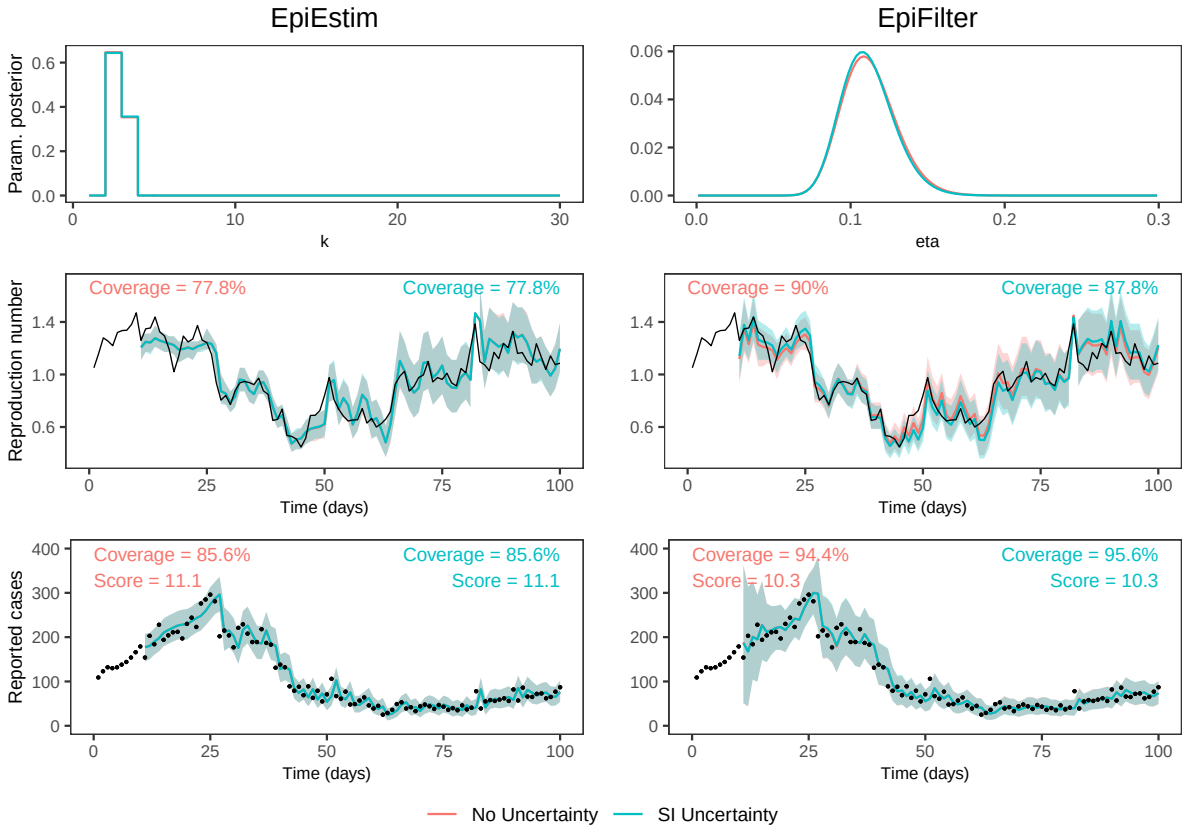

Figure A.2: Results from fitting the marginalised models while also accounting for uncertainty about the serial interval. Symmetry about the expected serial interval distribution produces very similar results whether we include uncertainty about the serial interval or not.

The apparent irrelevance of the serial interval in this case is expected, but will not always be the case. This also relies upon $P(\phi)$ being centred on the true ϕ. If this is not the case, then we have serial interval misspecification, which we explore in the next section.

A.4.2 Serial interval misspecification

Serial interval misspecification is known to bias estimates of R_t. Even in scenarios where high-quality data are available to estimate this distribution, unexpected biases can occur. During the COVID-19 pandemic, non-pharmaceutical interventions were found to shorten the generation time distribution of SARS-CoV-2 [413], and different SARS-CoV-2 variants also appeared to feature different generation times [414]. Other work has attempted to find solutions to this problem [264], we simply demonstrate the effects here.

We fit the marginalised models to the standard random walk simulation, assuming different serial interval distributions. The standard (correct) interval with mean 6.5 days and standard deviation 4.2 days is compared to results using a shorter interval with mean and standard deviation 4.2 days (an exponential distribution), and a longer interval with mean 9.75 days (1.5x the standard mean) and standard deviation 4.2 days. The interval is modified only during model fitting, the simulation itself remains unchanged.

Figure A.3 and Table A.3 highlight that, while the posterior distribution of the smoothing parameter depends slightly on the serial interval, and R_t estimates depend strongly on the serial interval, the posterior predictive distributions are largely independent of this choice. This is an expected result, as R_t and the serial interval are not jointly identifiable from reported case data alone. Diagnostics based on the predictive distribution of observed cases are unable to highlight serial interval misspecification.

Table A.3: Coverage of 95% credible intervals and CRPS values for the marginalised models fit to the random walk simulation with different serial intervals.

	Rt coverage		Predictive coverage		CRPS	
Serial interval	EpiEstim	EpiFilter	EpiEstim	EpiFilter	EpiEstim	EpiFilter
Standard (correct)	77.8%	90.0%	85.6%	94.4%	11.1	10.3
Exponential (shorter mean)	74.4%	86.7%	86.7%	94.4%	10.8	10.8
Longer mean	64.4%	72.2%	85.6%	96.7%	10.8	10.3

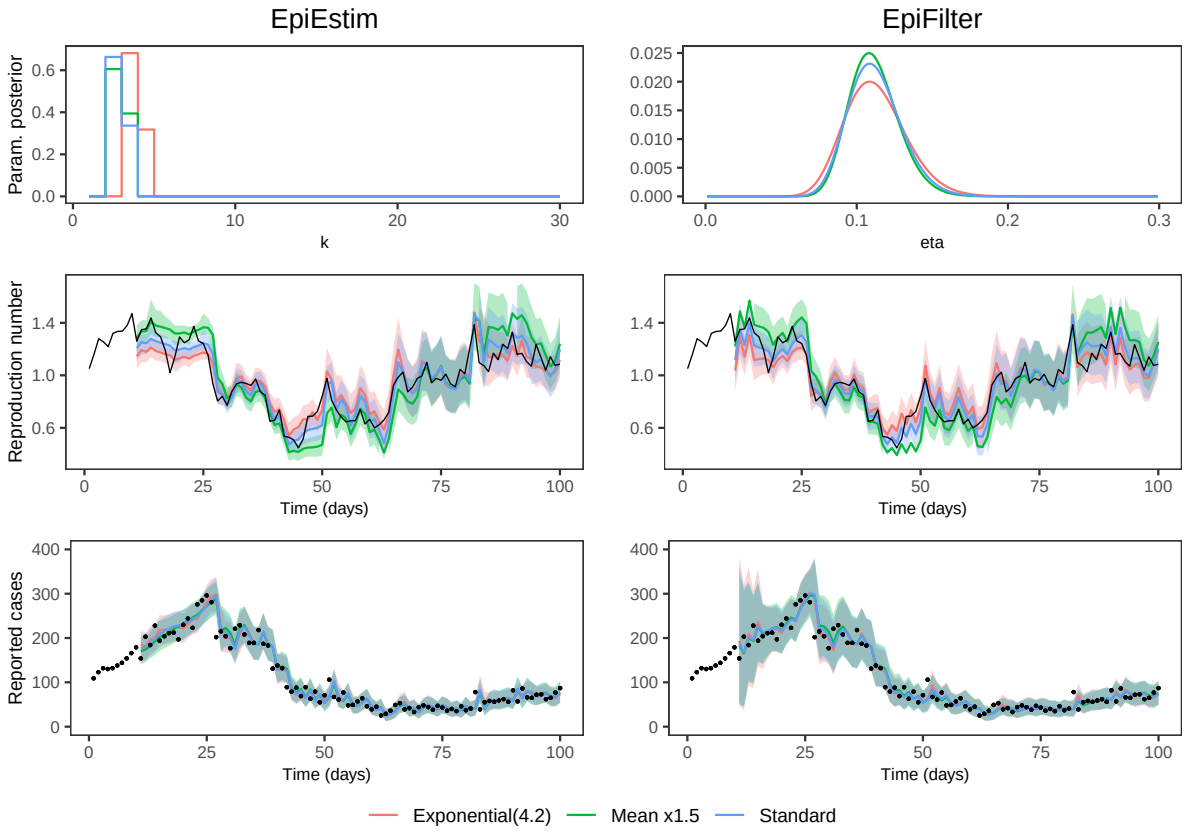

Figure A.3: Results from fitting the marginalised models to the random walk simulation with different serial intervals. The standard serial interval has a mean of 6.5 days and standard deviation of 4.2 days. The exponential serial interval has a mean of 4.2 days and standard deviation of 4.2 days. The longer serial interval has a mean of 9.75 days and standard deviation of 4.2 days.

A.5 Additional simulated results

A.5.1 Varying sample sizes

Common intuition suggests that the quality of inferences should improve as the sample size (i.e. the number of daily cases) increases. This is true for a correctly-specified model, but model misspecification, including in the choice of smoothing parameter, can lead to inferences that deteriorate with sample size (that is, the model becomes more confidently incorrect).

To demonstrate, we initialise the sinusoidal simulation with $C_0 = 25, 50, \ldots, 1600, 3200$ cases and fit the four models. Figure A.4 reports the coverage of the 95% credible intervals for both R_t and predictive cases, showing that while EpiFilter is largely robust to sample size, EpiEstim's coverage worsens as sample size increases in both the default and marginalised models. At a sufficiently large number of daily infections, EpiEstim's coverage increases again, as the posterior distribution of k becomes more concentrated around $k = 1$, and the model returns largely unsmoothed estimates of R_t. This causes the CRPS score to worsen dramatically, as credible intervals become extremely wide, and the model becomes extremely under-confident in its estimates. This highlights that problems caused by incorrect smoothing cannot be solved by increasing the number of daily cases, making marginalisation a crucial tool for both small and large epidemics.

A.5.2 Varying epidemic dynamics

We now test the models over a range of epidemic dynamics. The simulated rate of change of R_t is increased by (a) increasing the standard deviation of the random walk (by increasing η), (b) decreasing the period of the sinusoidal curve, and (c) increasing the number of evenly-spaced step-changes. The rate of change of R_t is decreased by doing the opposite. We fit the four models to 10 realisations of each dynamic model at each parameter value and report the average coverage, parameter estimates, and relative CRPS scores in Figure A.5.

As the rate at which R_t varies increases, the coverage of models with default smoothing parameters decreases, as they increasingly oversmooth. Applying our marginalisation technique improves coverage, with marginalised EpiFilter being the only inference approach that achieves (approximately) correct coverage under all considered epidemic dynamics. The CRPS score is

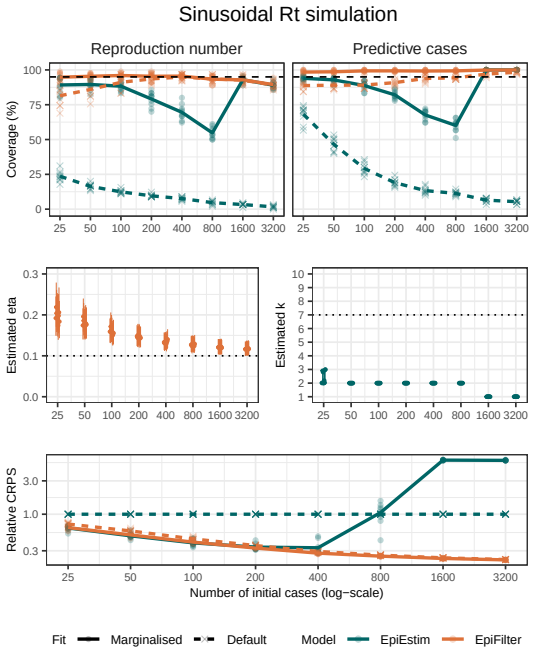

Figure A.4: Coverage of the 95% credible intervals for R_t and predictive cases (first row), posterior mean and 95% credible intervals for estimated parameters (second row), and the relative CRPS score (compared to default EpiEstim) for the four models fit to data from the sinusoidal simulations with varying initial number of cases C_0. A total of 10 simulations at each value of C_0 were performed and the results averaged to eliminate some stochasticity (solid lines). Results from individual simulations are reported in lighter points.

almost exclusively better for our marginalised models, except when epidemic dynamics result in the posterior distribution for EpiEstim's k having a large amount of mass at $k = 1$.

Default parameter values are shown in dotted lines in the second row of Figure A.5, highlighting that $\eta = 0.1$ is an acceptable point estimate for EpiFilter when R_t follows a sinusoidal curve with period greater than 30 days, or that $k = 7$ is an acceptable point estimate for EpiEstim when R_t follows a random walk with $\eta = 0.05$, for example. Finally, the marginalisation of EpiFilter, when fit to the random walk simulation, is consistently able to recover the true value of η.

Practical implications

We also consider the proportion of time that each model is correctly and incorrectly confident in the sign of $R_t - 1$ (Figure A.6). A model is confident in epidemic growth if the 95% credible interval for R_t is entirely above 1. We say it is correctly confident if the true value of R_t is also greater than 1, and incorrectly confident otherwise. A similar argument applies for confidence

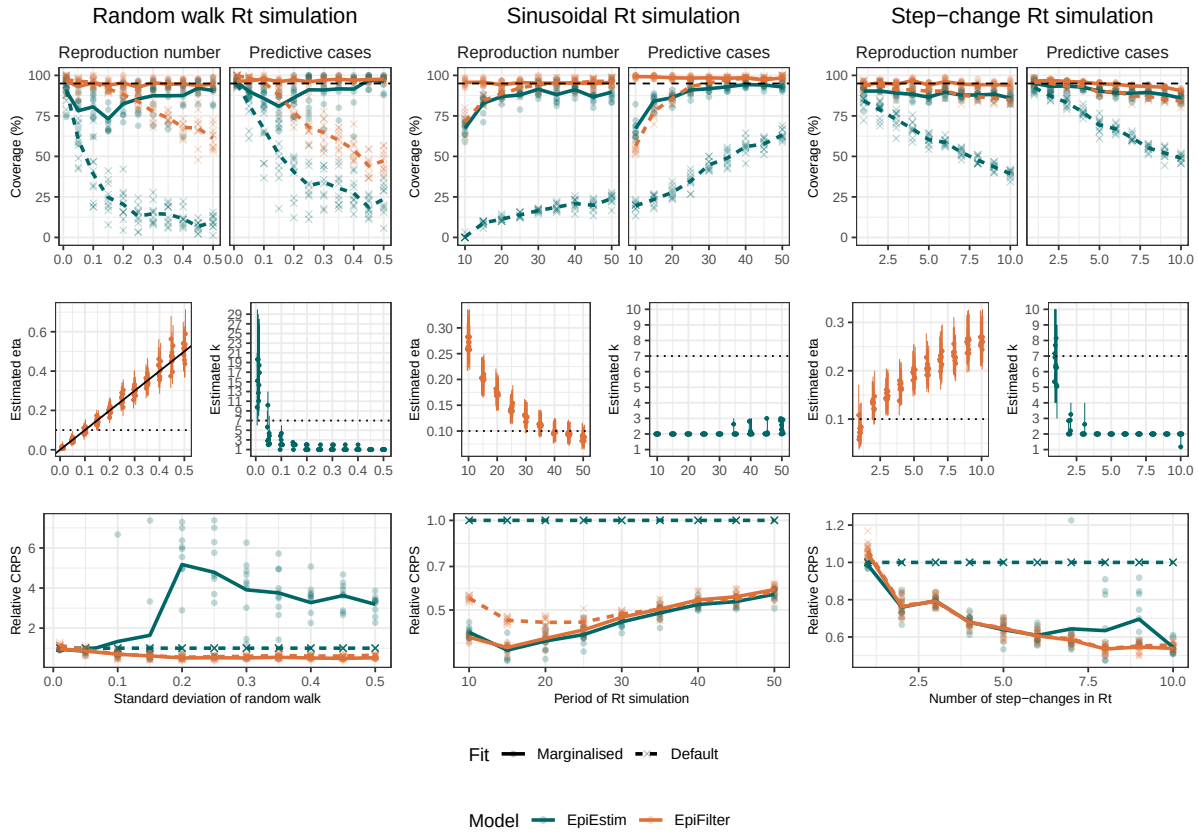

Figure A.5: Coverage of 95% credible intervals for R_t and predictive cases (first row), parameter estimates (second row), and relative CRPS compared to default EpiEstim (third row) for the four models fit to data from the three dynamic simulations with varying parameters. A total of 10 simulations at each parameter value were performed and the results averaged to eliminate some stochasticity (solid lines). Results from individual simulations are reported in lighter points. Default parameter values are shown in dotted lines, and the true value of the parameter (when EpiFilter is fit to the random walk simulation) is shown in a solid line.

in epidemic decline.

These results demonstrate that, while marginalisation generally increases the proportion of time in which models are correctly confident in the sign of $R_t - 1$, the crucial advantage is in reducing the proportion of time that the model is incorrectly confident in the sign of $R_t - 1$, which can be as high as 40% for default EpiEstim (in the sinusoidal R_t simulation with a rapidly changing R_t). This has important implications for public health decision-making and communication, which often uses whether $R_t > 1$ or not to indicate the state of the epidemic and to motivate interventions.

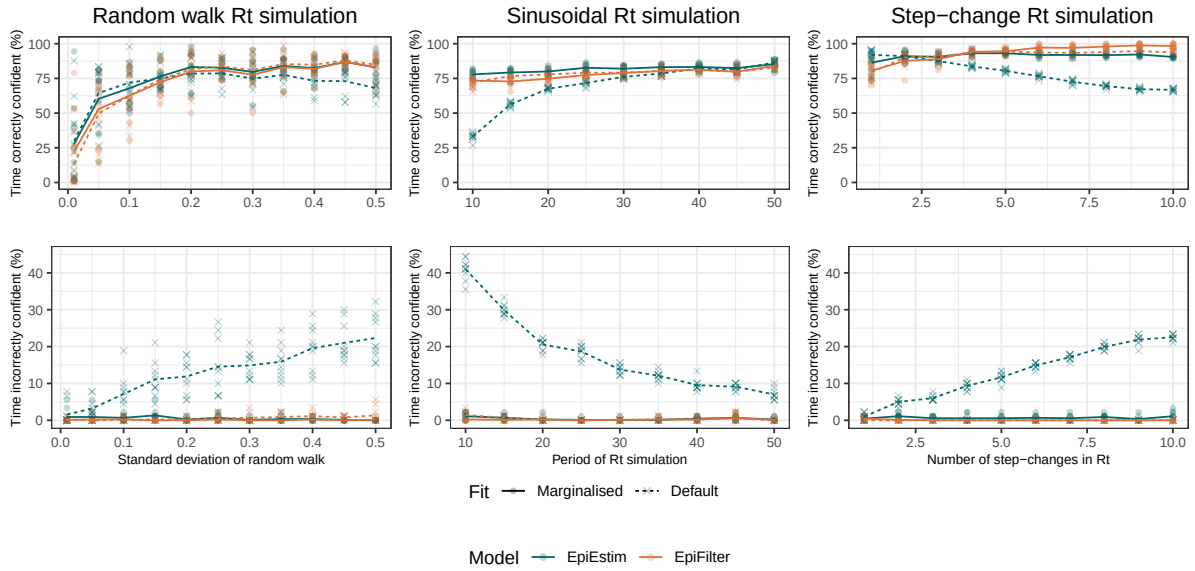

Figure A.6: The proportion of time each model is correctly confident in the sign of R_t (first row) and proportion of time each model is incorrectly confident in the sign of R_t (second row) as a function of epidemic dynamics. A total of 10 simulations at each parameter value were performed and the results averaged to eliminate some stochasticity (solid lines). Results from individual simulations are reported in lighter points.

A.5.3 Observation-based noise

Introducing observation-based noise to the simulations in Figure 4.2 results in posterior distributions for smoothing parameters that imply less smooth estimates (higher estimated η, lower estimated k), as the additional noise in the data is incorrectly assumed to arise from the epidemic process. We reproduce Figure 4.2 with simulated observation noise in Figure A.7. While marginalisation in this case still improves coverage of predictive cases (particularly in EpiFilter, where coverage of predictive cases remains approximately correct in all three simulations), the same guarantee no longer applies to R_t estimates.

Comparing Figure 4.2 to Figure A.7 also highlights that, while coverage of R_t is generally worse in the presence of observation-based noise, the decrease is much more marked in EpiFilter than in EpiEstim. In the sinusoidal and step-change examples, when observation noise is included, marginalised EpiEstim's coverage of R_t becomes comparable to that of marginalised EpiFilter's.

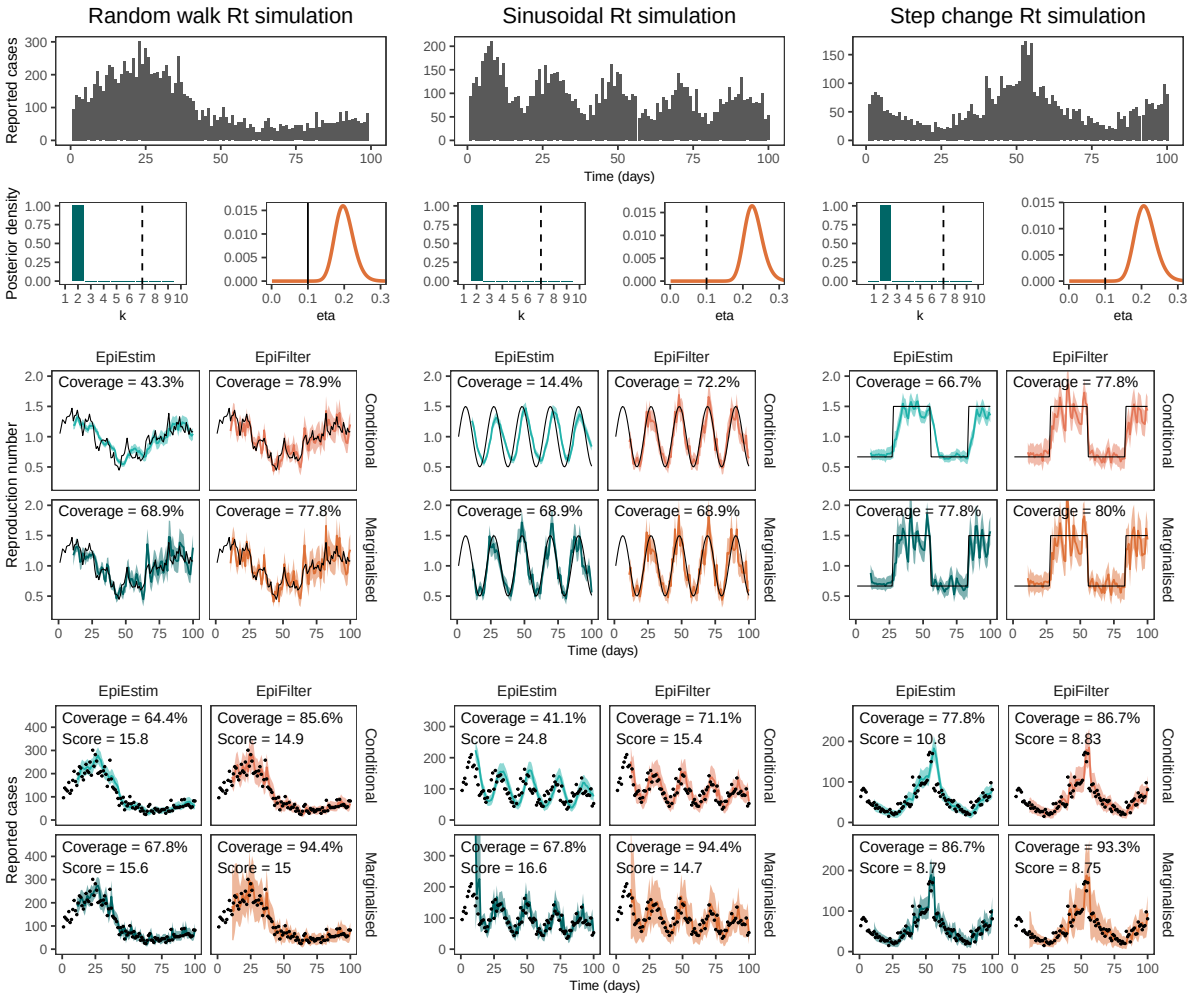

Figure A.7: Simulated case data (first row), posterior distributions for smoothing parameters at $t = 100$ (second row), estimates of R_t (third and fourth row), and estimates of predictive cases (fifth and sixth row) for three realisations of simulated epidemics, **with additional simulated observation noise**. The first column shows results for the Gaussian random walk simulation with $\eta = 0.1$, a dynamic model that precisely matches default EpiFilter. The second column shows results for the sinusoidal simulation, and the third column shows results for the step-change simulation. Vertical dotted lines in the parameter posterior distributions indicate the default parameter values, while the vertical solid line indicates both the default parameter value and the true value of the parameter (in EpiFilter when fit to the random walk simulation). Black lines (in R_t estimates) and black dots (in predictive C_t estimates) show the true values of R_t and C_t respectively. Predictive coverage of the 95% credible intervals (closer to 95% is better) and the CRPS scores (lower is better) are shown within each figure.

A.5.4 Alternative epidemic simulations

Thus far we have tested the models on data simulated using a Poisson renewal model. In this section we test the models on data generated using a simple stochastic susceptible-infectious-recovered (SIR) model.

Simulations are initialised with $S_0 = 9950$ susceptible individuals, $I_0 = 50$ infectious individuals, and 0 recovered individuals. On each time-step t, a Poisson-distributed number of new infections is generated with rate $\beta I_{t-1} S_{t-1}/N$ and a Poisson-distributed number of recoveries is generated simultaneously with rate γI_{t-1}. The theoretical instantaneous reproduction number is $R_t = \frac{\beta}{\gamma}\frac{S_{t-1}}{N}$.

We use $\beta = 1/4$ and $\gamma = 1/6.5$, implying a basic reproduction number of $R_0 = 1.625$ and a generation time distribution that follows a geometric distribution with mean 6.5 days. We use this generation time distribution when fitting our models (misspecification of the serial interval, and thus of the generation time distribution, is considered in Supplementary Section A.4).

We fit each model twice: first on a realisation from this SIR model directly, and secondly on the same realisation with binomial noise (simulated by sampling $\tilde{C}_t \sim \text{Binomial}(2 * C_t, 0.5)$).

Results are presented in Figure A.8, which shows that marginalisation once again improves model results on simulations both with and without observation noise. This is unsurprising as the simple SIR model is a special case of the renewal model with an exponential generation time distribution (geometric in the discretised case presented here). One notable difference is that default EpiFilter *undersmooths* R_t, compared to the oversmoothing observed in other simulations. This is a result of more gradual changes in R_t (the reproduction number in this SIR model changes only as a result of accumulated immunity).

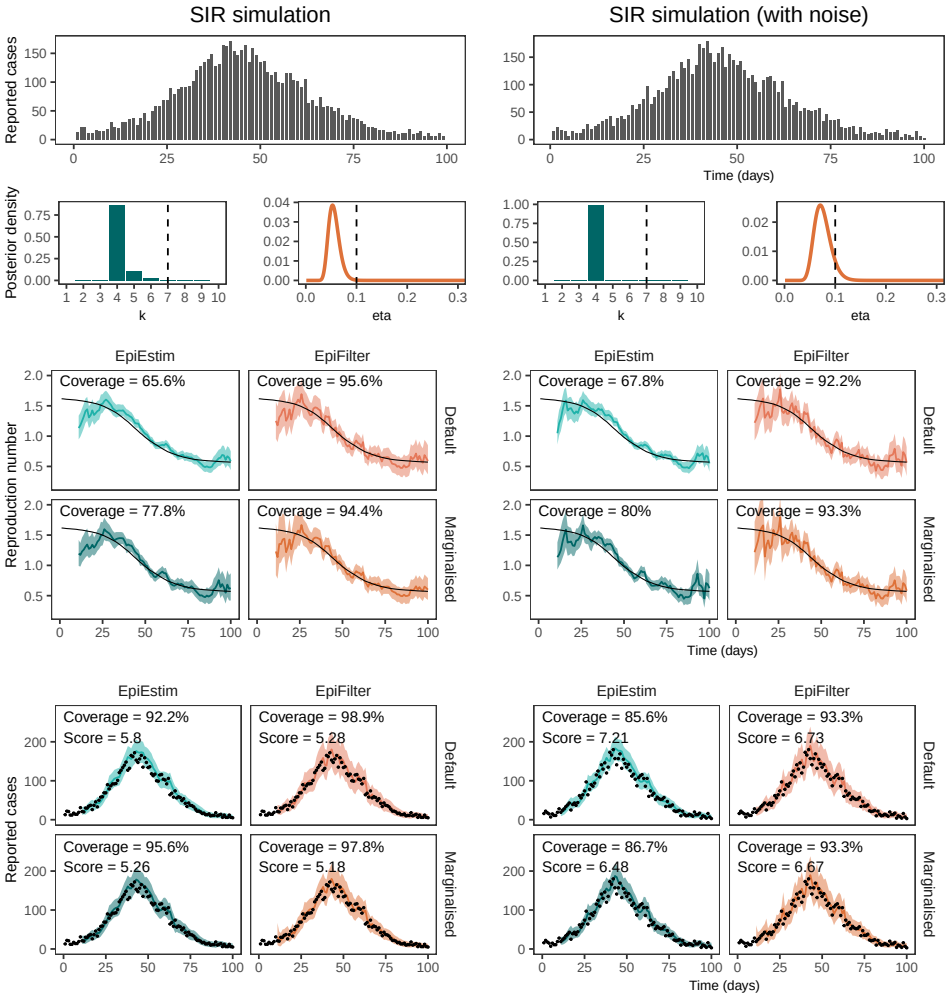

Figure A.8: Simulated case data (first row), posterior distributions for smoothing parameters at $t = 100$ (second row), estimates of R_t (third and fourth row), and estimates of predictive cases (fifth and sixth row) for two realisations of simulated epidemics from an SIR model. The first column shows results for the SIR simulation without observation noise, and the second column shows results for the SIR simulation with observation noise. Vertical dotted lines in the parameter posterior distributions indicate the default parameter values. Black lines (in R_t estimates) and black dots (in predictive C_t estimates) indicate the true values of R_t and C_t respectively. Predictive coverage of the 95% credible intervals (closer to 95% is better) and the CRPS scores (lower is better) are shown within each figure.

A.6 Additional real-world examples

A.6.1 New Zealand data

Neither EpiEstim nor EpiFilter allow for observation noise in their models, so all noise in the data is assumed to arise from the epidemic process. In practice, reporting noise can be a significant factor. We attempt to side-step this problem when fitting to New Zealand data by imposing a 5-day moving average on the raw case data. We test the impact of this on our conclusions here by re-fitting the model on the raw data, and on data smoothed using a 10-day moving average.

Figure A.9 shows that parameter estimates are highly dependent on any pre-smoothing performed on the data, with smoother estimates being produced on data that underwent more pre-smoothing. Despite this, marginalisation still consistently improves coverage of observed data and CRPS, and the models obtain good coverage of observables in general.

If observation noise is considered substantial, we encourage the reader to use a model that explicitly accounts for this. Examples in the literature that jointly handle observation noise and smoothing parameters are EpiNow2 and EpiLPS(MALA), both of which are outlined in Supplementary Section A.2.

A.6.2 Alternative datasets

Using the same data as [242], we present additional examples of the fitting of EpiEstim and EpiFilter to real-world data: the 1918 influenza outbreak in Baltimore and 2003 SARS outbreak in Hong Kong. The models are also fit to data that has been smoothed using a 7-day moving-average to eliminate some observation noise, which our models are unable to account for. Results are presented in Figure A.10 which shows that marginalisation almost universally improves both the coverage of predictive cases and the CRPS, with the sole exception of fitting EpiFilter to the smoothed 1918 influenza data.

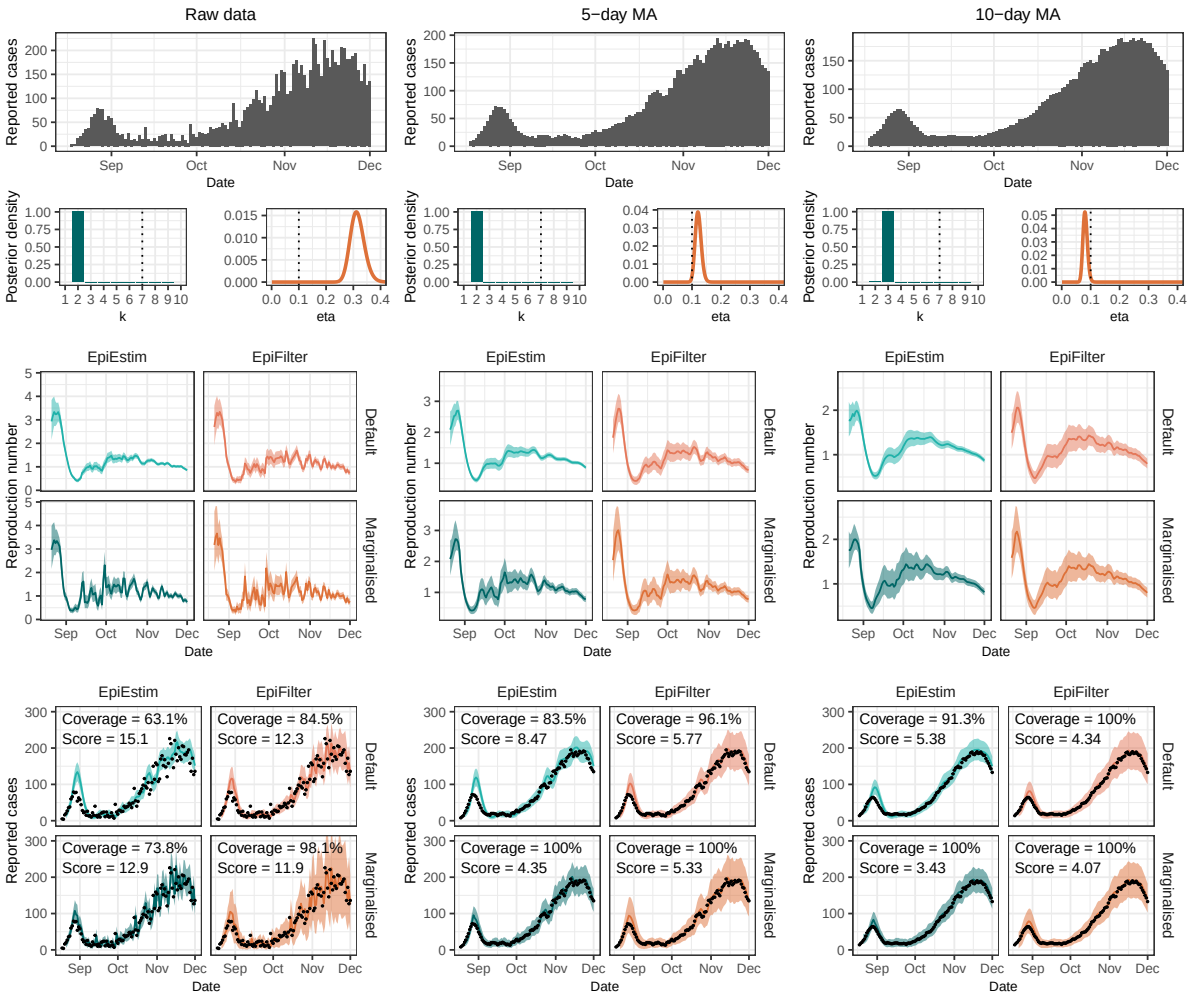

Figure A.9: Results from fitting all four models to raw New Zealand data (column 1), and New Zealand data smoothed using a 5-day and 10-day moving average (columns 2 and 3 respectively). In row 2, vertical dashed lines indicate default parameter values while coloured curves show parameter posterior distributions. Coloured lines in the remaining figures show central estimates (mean of the posterior distributions) and coloured bands show 95% credible intervals. Observed case data are shown in black points.

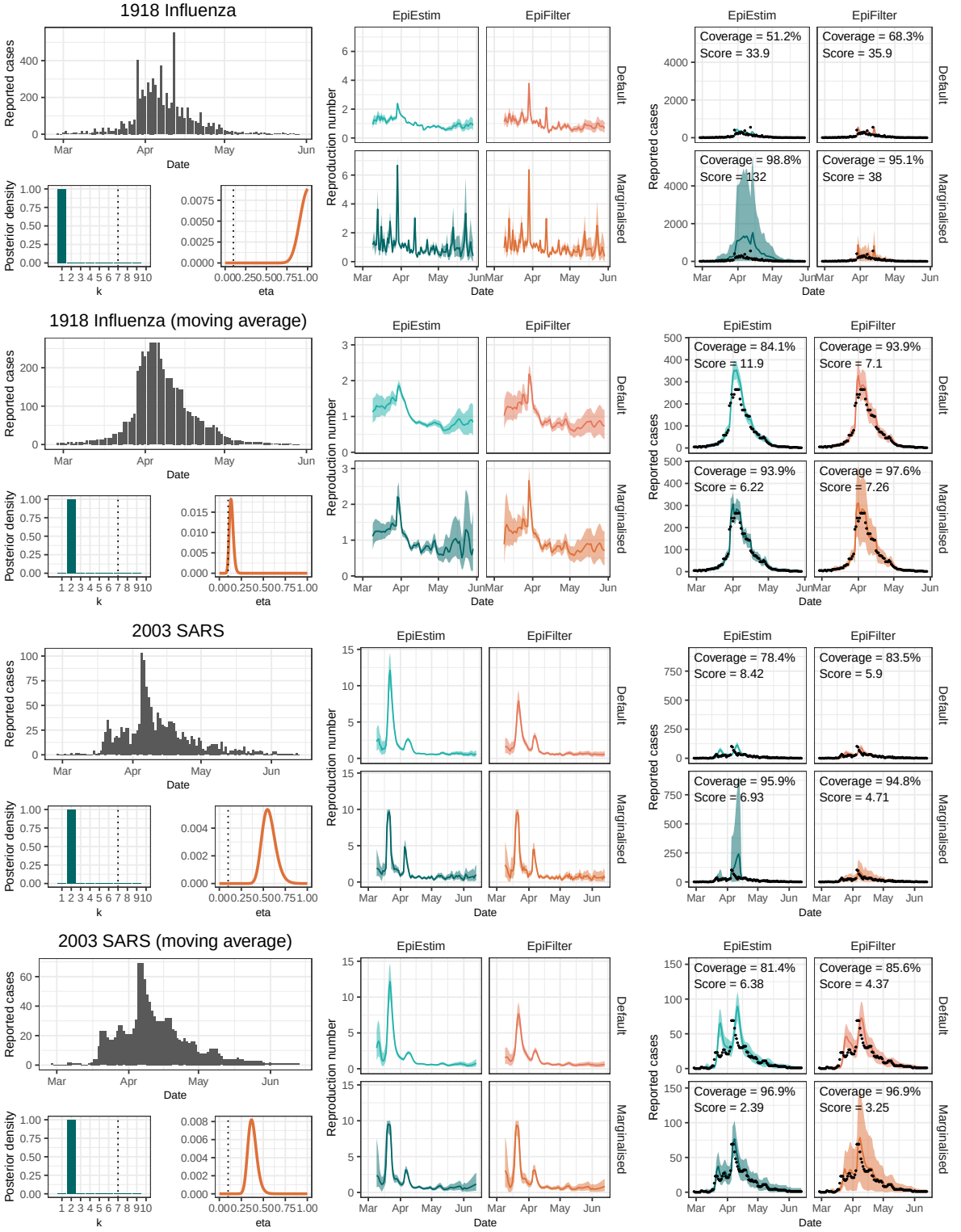

Figure A.10: Results from fitting EpiEstim and EpiFilter to two real-world time-series. The models are also fit on smoothed data (to eliminate some observation noise), demonstrating that this can impact parameter estimates.

A.7 Stepwise likelihoods

The model likelihoods are calculated by summing the log-likelihood of the one-step-ahead forecasts at each time-step. By plotting these one-step-ahead likelihoods for various values of k and/or η, we can identify the periods of the epidemic that are driving the observed parameter estimates (Figure A.11).

Most notably, we can see that, for most periods of time, smaller values of η are preferred to larger values. However, when R_t is changing rapidly (such as when R_t is crossing 1 in the sinusoidal simulation), these small values incur a large "penalty" in the likelihood, resulting in more posterior mass at larger values of η. The same effect also occurs in EpiEstim, although as most mass is concentrated around small values of k, the effect is less pronounced.

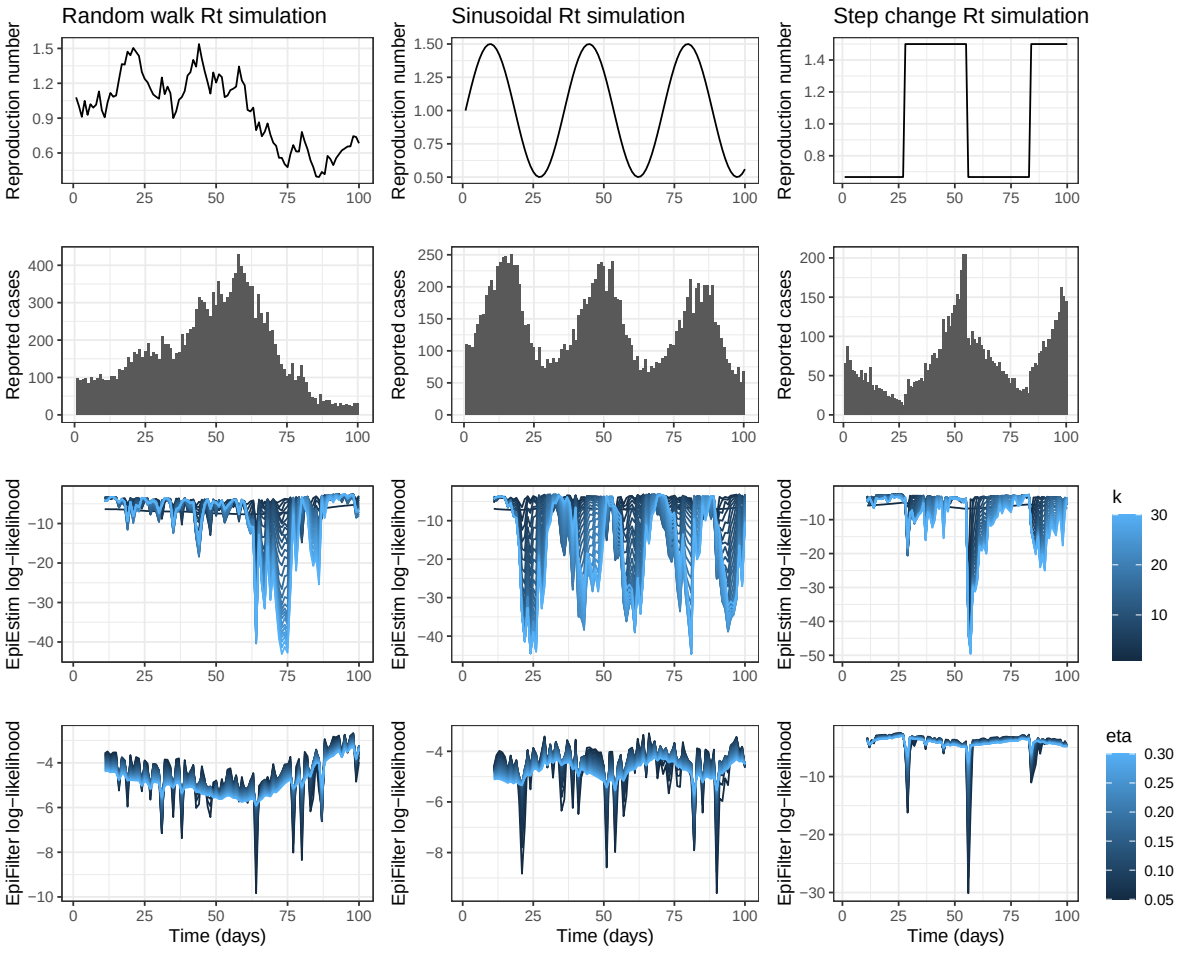

Figure A.11: Simulated R_t (first row), simulated reported cases (second row), stepwise log-likelihoods for EpiEstim (third row) and stepwise log-likelihoods for EpiFilter (fourth row) for the sinusoidal simulation. Darker lines show stepwise log-likelihoods associated with smaller values of η (increased smoothing) and smaller values of k (increased smoothing).

A.8 APE comparison

Parag & Donnelly [242] treat the selection of EpiEstim's k as a model-selection problem and use information theoretic arguments to suggest that the optimal value of k is that which minimises the accumulated prediction error (APE). The APE is defined (in our notation) as:

$$APE_k = \sum_{s<t} -\log P_{APE}(C_{s+1}|C_{1:s}, k). \tag{A.28}$$

Noting the following equivalence to the predictive decomposition of the likelihood (Equation A.2):

$$\ell(k|C_{1:t}) = \sum_{s=1}^{t} \log P(C_s|C_{1:s-1}, k) \;=\; -APE_k, \tag{A.29}$$

the k that minimises the APE should be the same k that maximises the log-likelihood. The two approaches differ, however, in their derivation of $P(C_s|C_{1:s-1}, k)$.

Parag & Donnelly derive the predictive distribution for C_s given $C_{1:s-1}$ by first noting that $R_{s-1}|C_{1:s-1}, k$ is Gamma-distributed with shape parameter $\alpha_{s-1,k}$ and rate parameter $\beta_{s-1,k}$, and then assuming that $C_s|R_{s-1}, \Lambda_s$ is Poisson distributed with rate $R_{s-1}\Lambda_s$. **We note that this introduces a new assumption in the model: specifically, that before we observe C_s, we assume R_s has been fixed for the preceding $k+1$ days.** The resulting predictive distribution for C_s is then a negative binomial distribution:

$$P_{APE}(C_s|C_{1:s-1}, k) \sim \text{NegativeBinomial}\left(r = \alpha_{s-1,k}, p = \frac{\beta_{s-1,k}}{\Lambda_s + \beta_{s-1,k}}\right). \tag{A.30}$$

This is slightly different to our approach (Equation A.7), where $r = \alpha_{s-1,k-1}$ and $p = \beta_{s-1,k-1}/(\Lambda_s + \beta_{s-1,k-1})$. The APE_k metric using the predictive distribution derived by Parag & Donnelly is precisely equal to $-\ell(k+1|C_{1:t})$. The additional assumption made by the APE approach results in a bias towards smaller values of k.

We present the APE-metric and negative-log-likelihood for values of k between 1 and 30 on the New Zealand COVID-19 dataset in Figure A.12. We note that the APE-metric is minimised at $k = 1$, whereas the negative-log-likelihood is minimised at $k = 2$.

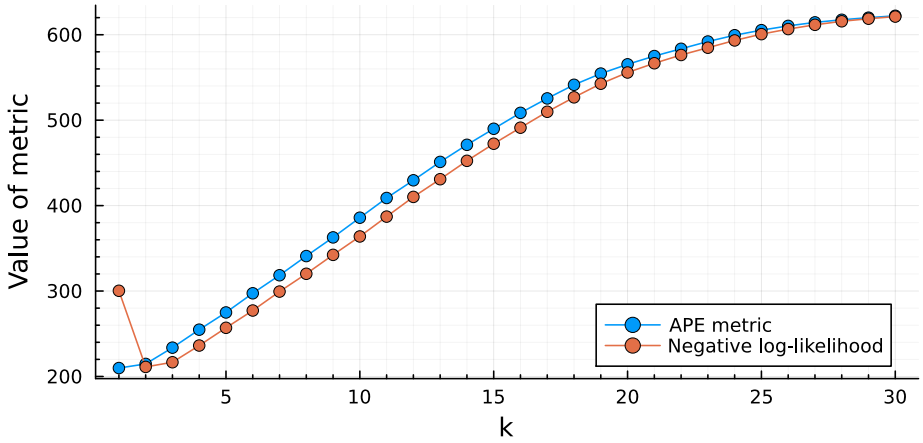

Figure A.12: Comparison of the APE metric from Parag & Donnelly [242] and the negative log-likelihood of k for EpiEstim on the NZ COVID-19 dataset. For $k \geq 1$, the value of APE_k is exactly the same as $-\ell(k{+}1|C_{1:t})$. The value of $\ell(1|C_{1:t})$ is the log-likelihood of $k = 1$ (where the predictive distribution for R_t is simply the prior distribution) and has no immediate analogue in the APE approach.

A.9 Testing grid-sizes

There are three instances in which we employ non-exact grid-based approximations to posterior distributions:

1. When marginalising out k in EpiEstim (grid-approximation for R_t)
2. Both default and marginal EpiFilter (grid-approximation for R_t)
3. When fitting η in EpiFilter (grid-approximation for η)

By default, these grids are chosen to be of size 1000 (Supplementary Section A.1.1). We test the sensitivity of our results to grids of size 500 (a coarser grid) and 2000 (a finer grid) and find that these grids have very little impact on our results. We further test EpiEstim with an R_t grid-size of just 100, at which size there is evidence of some degeneracy in the posterior distribution for R_t. Full results are discussed in the captions of Figures A.13, A.14, and A.15.

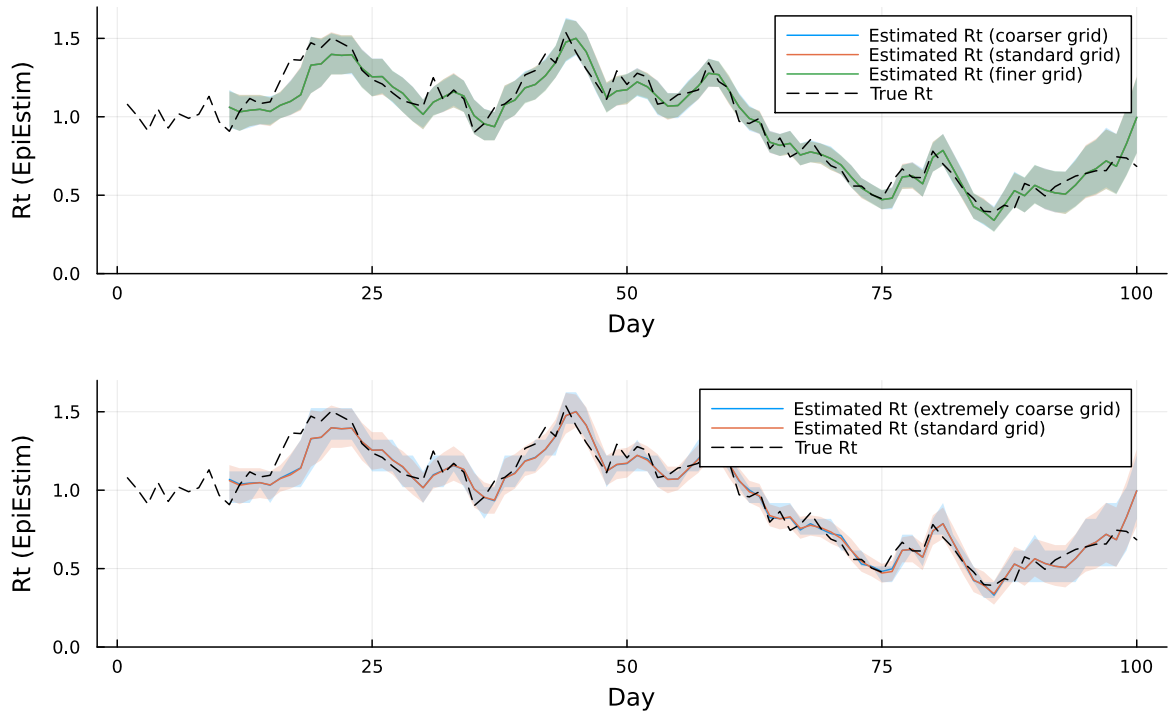

Figure A.13: Results from fitting (marginalised) EpiEstim to simulated data using three different grid-sizes for R_t: 500 (coarser), 1000 (standard), 2000 (finer). We also test the model with a grid-size of 100 (very coarse grid). Only in the case of the very coarse grid do we see evidence of degeneracy in the posterior distribution for R_t, suggesting that a grid size of 500 or more is sufficient.

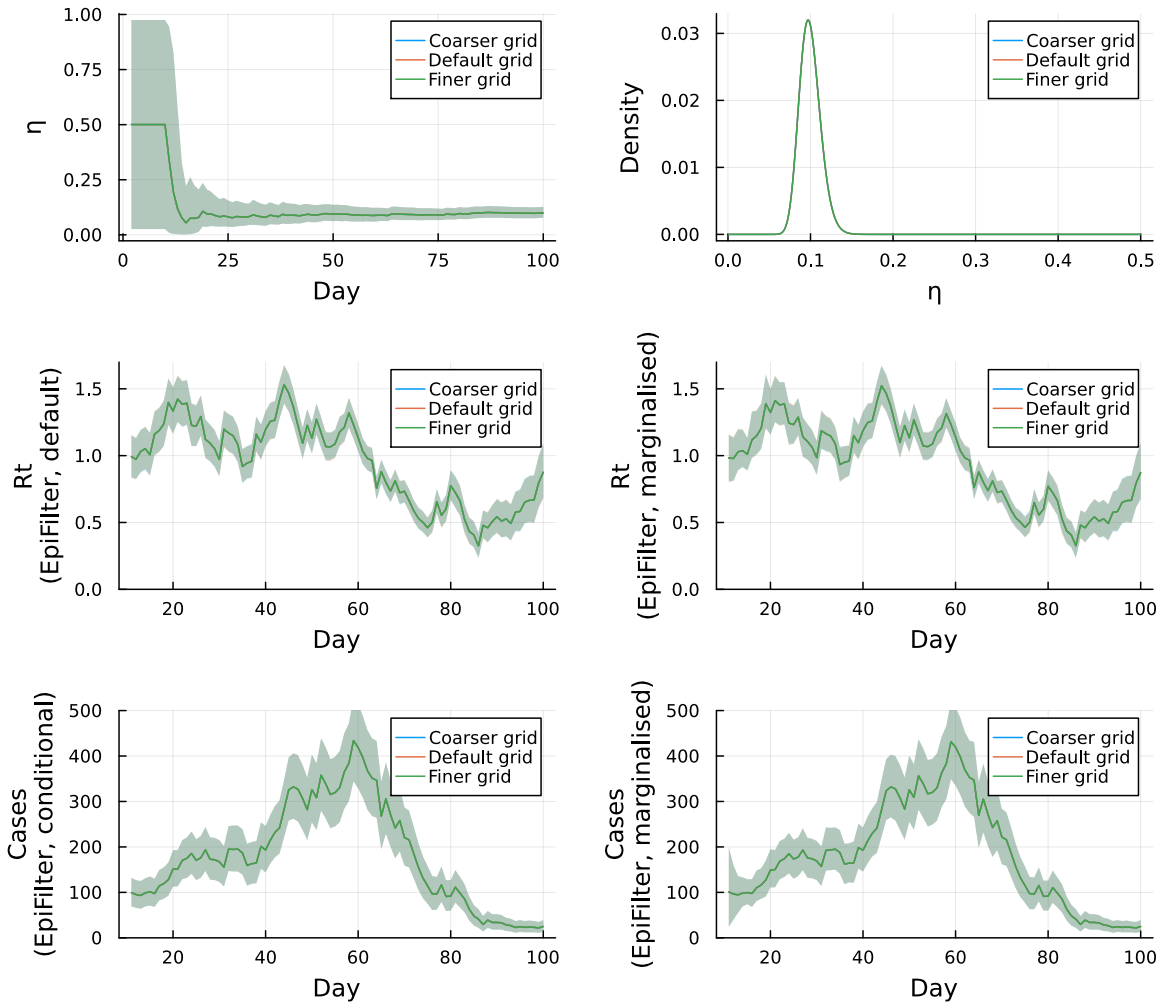

Figure A.14: Results from fitting EpiFilter (default and marginalised) to simulated data using three different grid-sizes for R_t: 500 (coarser), 1000 (standard), 2000 (finer). We find that the choice of grid-size has very little impact on the results. *The density plot for η is the posterior distribution of this parameter at the final time-step (i.e. conditional on all data).

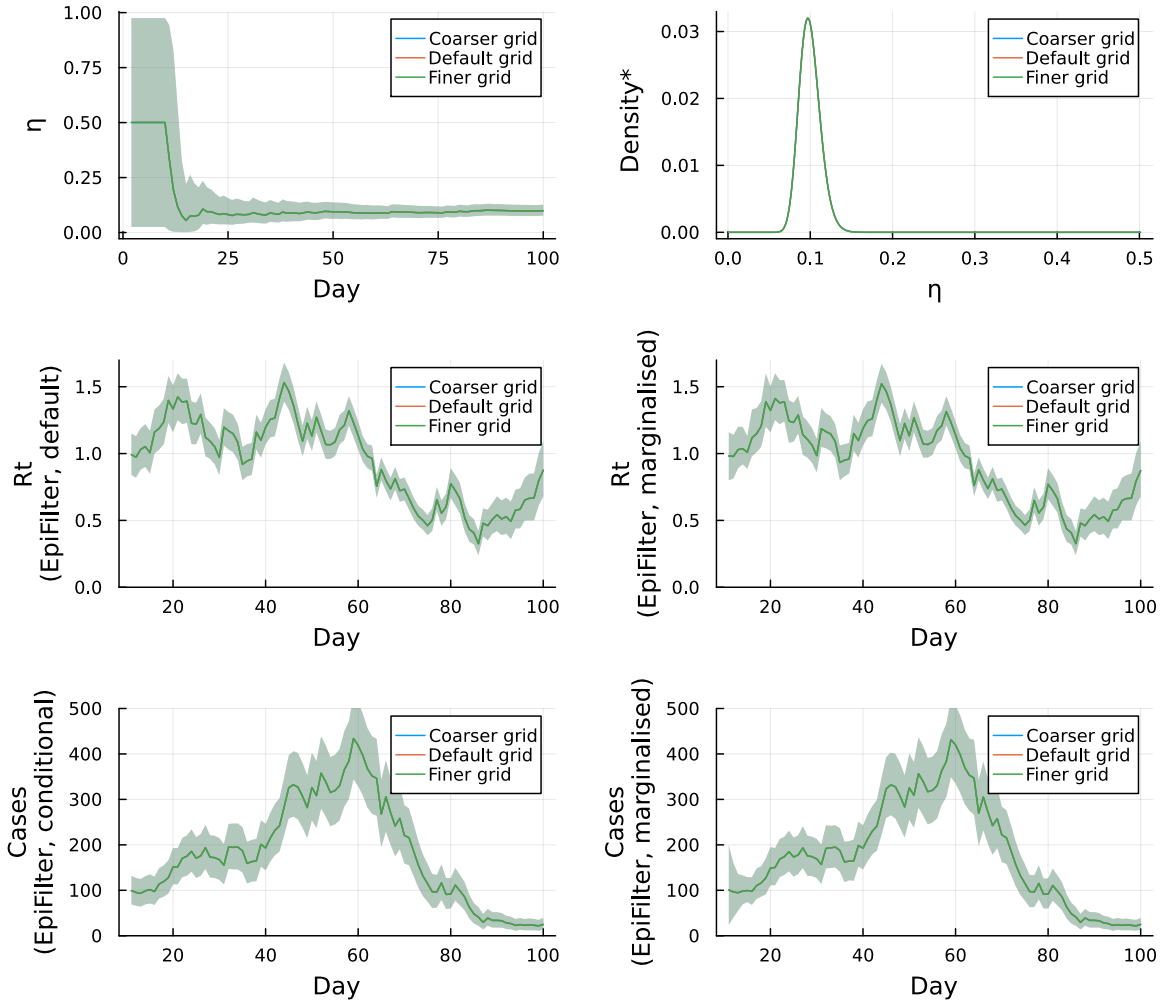

Figure A.15: Results from fitting EpiFilter to simulated data using three different grid-sizes for η: 500 (coarser), 1000 (standard), 2000 (finer). We find that the choice of grid-size has very little impact on the results, such that the resulting posterior distributions are indistinguishable. *The density plot for η is the posterior distribution of this parameter at the final time-step (i.e. conditional on all data).

Appendix B

Supplementary material for Chapter 6: A Bayesian model for repeated cross-sectional epidemic prevalence survey data

B.1 Sensitivity of the SIMPLE approach to choice of prior distributions

B.1.1 Invariance to reasonable prior distributions

In the main paper, we claim that results from the SIMPLE approach are largely invariant to the reasonable choice of initial distribution for the growth rate r_0, initial distribution for prevalence P_0, and prior distribution for σ. We demonstrate this here by refitting the basic model (growth rate epidemic model and basic observation model) to a single realisation of simulated data using different prior distributions. Figure B.1 supports our claim. Small deviations in r_t estimates at early time steps are observed when varying the prior for r_0, and similarly for P_t when varying the prior for P_0. The fixed-lag resampling ensures that future data still influences these estimates, hence they do not vary as much as might be expected. The remaining results are nearly identical regardless of prior distribution. These results hold over repeated simulations.

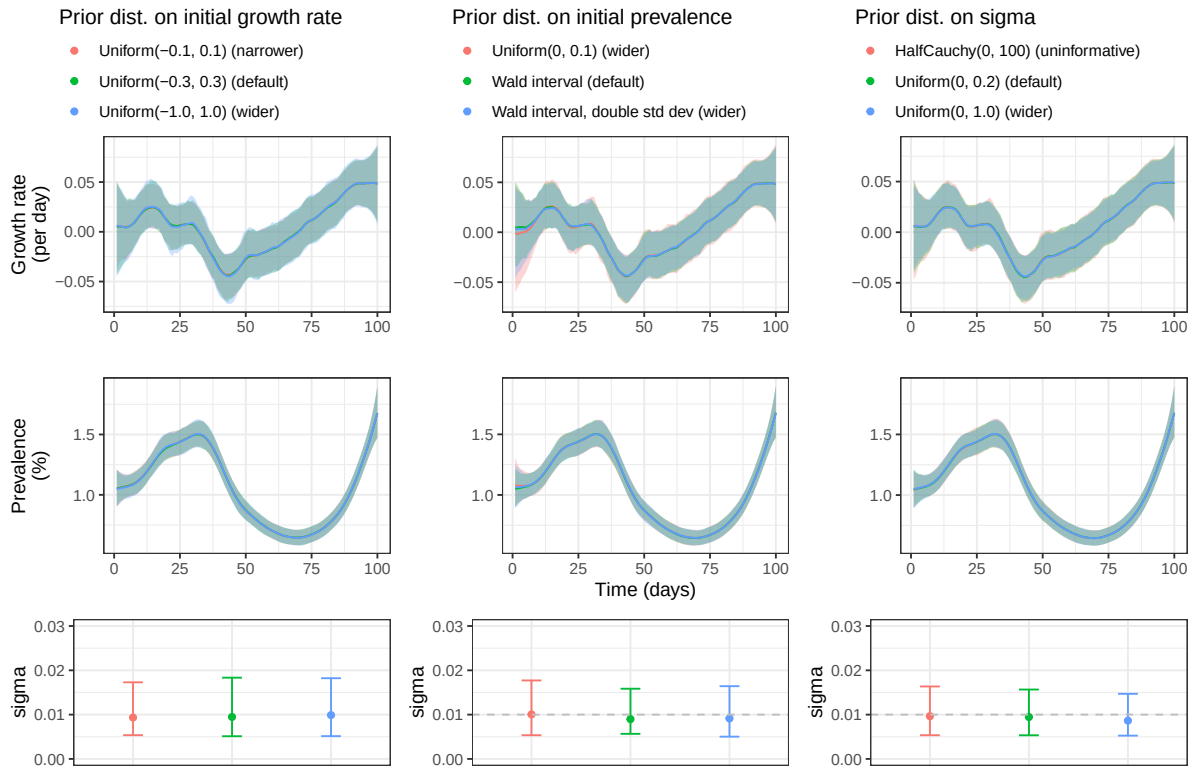

Figure B.1: Estimates of the growth rate and prevalence over time (central lines show the posterior mean and shading shows the 95% credible interval), and posterior mean and 95% credible intervals on parameter σ when fitting the SIMPLE approach (growth rate epidemic model and binomial observation model) to simulated data. Column 1 shows results under varying prior distributions on the initial growth rate, column 2 shows results under varying prior distributions for the initial prevalence, and column 3 shows results under varying prior distributions for the smoothing parameter σ.

B.1.2 Informative prior distributions

In the main paper, we claim that "using more informative prior distributions that place less mass on implausibly large values (of σ and ρ) can help to reduce the width of the credible intervals" on the hidden states. We demonstrate this here by fitting the beta-binomial model to 10 realisations of simulated data (with $\sigma = 0.008$ and $\rho = 2 \times 10^{-4}$) at various values of n_t. First, we fit the model using the default prior distributions ($P(\sigma) \sim U(0, 0.2)$ and $P(\rho) \sim U(0, 0.01)$). Then, we refit the model (to the same simulations) using tighter prior distributions $P(\sigma) \sim U(0.004, 0.016)$ and $P(\rho) \sim U(1 \times 10^{-4}, 4 \times 10^{-4})$, chosen to reflect typical lower and upper 95% credible intervals of the posterior distributions when the default model is fit to simulated data with large $n_t = 10,000$. Finally, we refit all models with σ and ρ fixed at their true values (i.e., $P(\sigma) \sim \delta(0.008)$ and $P(\rho) \sim \delta(2 \times 10^{-4})$ where δ is the Dirac

delta function).

Figure B.2 presents the coverage and average width of the 95% credible intervals for the growth rate r_t and prevalence P_t (in %) for each of the three scenarios. The effect is most pronounced for growth rate estimates, where credible intervals produced using the tighter prior distribution are substantially narrower than those produced using the default prior distribution, particularly at smaller values of n_t (where there are fewer data to inform the posterior distribution). Using known parameters results in a further reduction in the width of the credible intervals. This effect primarily occurs due to eliminating the possibility of higher values of σ. The coverage of the credible intervals also improves (becomes closer to 0.95 on average) as removing the possibility of higher values of σ also reduces the chance of overcovering the true value (due to wider credible intervals). The effect on prevalence estimates is less pronounced but still noticeable. These improvements are only expected when the prior distributions contain the true values of the parameters.

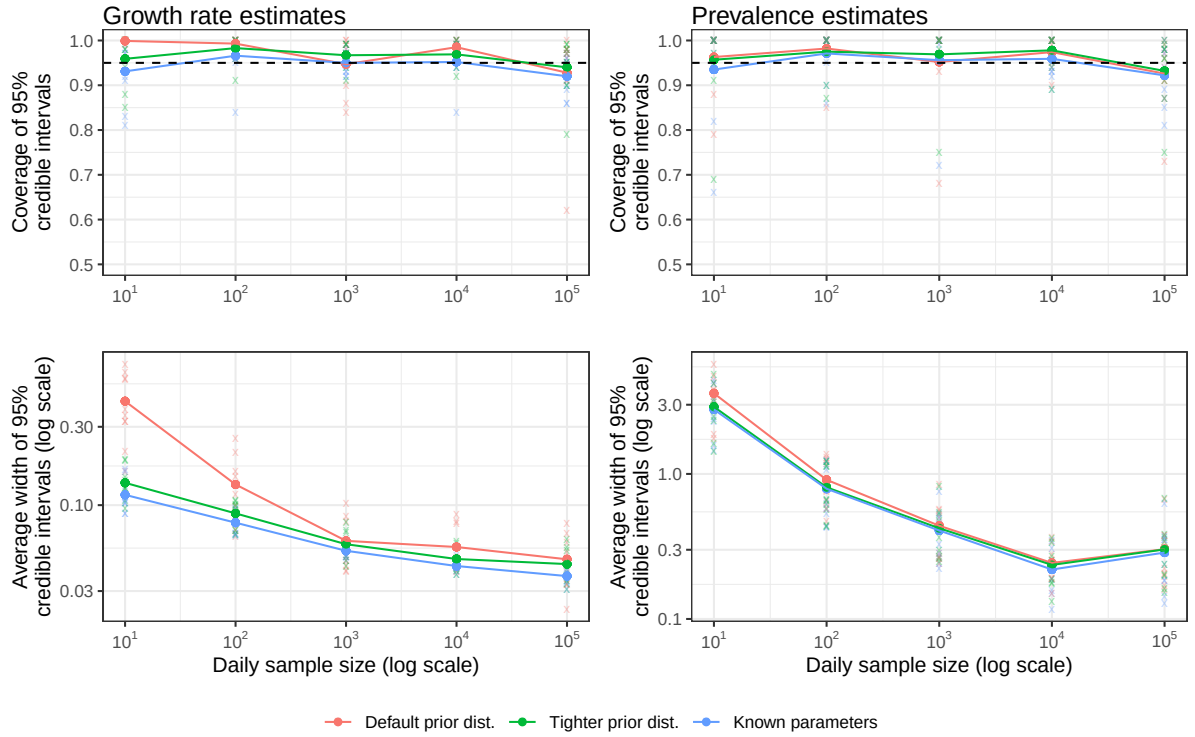

Figure B.2: Coverage and average width of 95% credible intervals for the growth rate r_t and prevalence P_t (%) from fitting the beta-binomial model with the default prior distribution (orange), with a tighter prior distribution (green), and with known parameter values (blue). Results from individual simulations are shown as transparent crosses, with averages over 10 simulations shown as points connected by solid lines. A range of assumed daily sample sizes n_t are considered (x-axis). The horizontal dashed line shows the target coverage of 95%.

B.2 Algorithms

We provide here a concise description of the algorithms used in the SIMPLE approach. We direct readers new to SMC methods to [40], which provides a comprehensive overview of the methodology in an epidemiological context.

For generality, we denote the hidden states at time step t by X_t (e.g., $X_t = (r_t, P_t)$ in the basic model), observed data by y_t (e.g., $y_t = (n_t, n_t^+)$), and model parameters by θ (e.g., $\theta = (\sigma, \rho)$ in the overdispersed model).

B.2.1 The bootstrap filter

Given a fixed lag L, we approximate the marginal posterior distribution of the hidden states at time t given past and future observed data $P(X_t|y_{1:t+L}, \theta)$ using a set of N_x "particles" $\{x_t^{(i)}\}_{i=1}^{N_x}$, where each x_t represents a sample from $P(X_t|y_{1:t+L}, \theta)$. The lag L is chosen to be large enough such that $P(X_t|y_{1:t+L}) \approx P(X_t|y_{1:T})$, but small enough to prevent particle degeneracy.

The particles $x_t^{(i)}$ are constructed using a fixed-lag bootstrap filter, described in Algorithm 4. Starting with a set of particles sampled from an initial state distribution $\{x_0^{(i)}\}_{i=1}^{N_x} \sim P(X_0)$, we iteratively propagate these particles by first sampling proposed particles from the state-space transition distribution $\{\tilde{x}_t^{(i)}\}_{i=1}^{N_x} \sim P(X_t|x_{t-1}^{(i)}, \theta)$, weighting them according to the observation distribution $w_t^{(i)} = P(y_t|\tilde{x}_t^{(i)}, \theta)$, and then resampling the proposed particles with replacement according to these weights. This process is repeated for each t until the final time step T is reached. The resampling step also resamples particles in the preceding L time steps, ensuring that the particles represent samples from the "smoothing posterior" $\{x_t^{(i)}\}_{i=1}^{N_x} \sim P(X_t|y_{1:t+L}, \theta)$.

In addition to producing the smoothing distribution (conditional on parameters θ), the bootstrap filter also produces estimates of the model likelihood $P(y_{1:T}|\theta)$ as a by-product. Specifically, note that the likelihood of the one-step-ahead forecast can be written as, and approximated by:

$$P(y_t|y_{1:t-1}, \theta) = \int P(y_t|X_t, \theta) P(X_t|y_{1:t-1}, \theta) dX_t \approx \frac{1}{N_x} \sum_{i=1}^{N_x} w_t^{(i)}.$$

The model log-likelihood can therefore be written as and then approximated by:

$$\ell(\theta|y_{1:t}) = \sum_{t=1}^{T} \log P(y_t|y_{1:t-1}, \theta) \approx \sum_{t=1}^{T} \log \left(\frac{1}{N_x} \sum_{i=1}^{N_x} w_t^{(i)} \right). \tag{B.1}$$

Algorithm 4 Fixed-lag bootstrap filter

1. Input:
   - N_x - number of particles
   - L - fixed lag
   - θ - parameter vector
   - $P(X_0)$ - initial state distribution
   - $P(X_t|X_{1:t-1}, \theta)$ - state-space transition distribution
   - $P(y_t|X_t, \theta)$ - observation distribution
2. Initialise:
   (a) For $i = 1, ..., N_x$ sample $x_0^{(i)} \sim P(X_0)$
   (b) Set $t = 1$
3. Project and filter:
   (a) For $i = 1, ..., N$ sample $\tilde{x}_t^{(i)} \sim P(X_t|x_{t-1}^{(i)}, \theta, \cdots)$.
   (b) Calculate weights $w_t^{(i)} = P(y_t|\tilde{x}_t, \theta, \cdots)$
4. Resample:
   (a) Resample $(x_{t-L:t-1}^{(i)}, x_t^{(i)})$ with replacement from $(x_{t-L:t-1}^{(j)}, \tilde{x}_t^{(j)})$ with probability proportional to $w_t^{(j)}$.
5. If $t < T$, increment t and go to step 3. Otherwise, end.

The quality of the approximation improves with larger N_x.

B.2.2 The particle marginal Metropolis-Hastings algorithm

PMMH is a simple Markov chain Monte Carlo (MCMC) algorithm that allows for the estimation of model parameters θ given observed data $y_{1:T}$. The algorithm uses the bootstrap filter to estimate the model likelihood, and incorporates these estimates into a Metropolis-Hastings step to determine whether to accept or reject proposed values of θ. We outline this in Algorithm 5.

Algorithm 5 Particle Marginal Metropolis-Hastings

1. Input:
   - A fixed-lag bootstrap filter (and corresponding parameters)
   - $P(\theta)$ - prior distribution for parameters
   - $q(\theta'|\theta)$ - proposal distribution for parameters
   - ESS - target effective sample size
   - $\hat{R}$ - target R-hat statistic
2. Initialise:
   (a) Sample $\theta_0 \sim P(\theta)$
   (b) Run Algorithm 4 to obtain estimated log-likelihood $\hat{\ell}(\theta_0|y_{1:T})$
   (c) Set counter $i = 1$
3. Metropolis-Hastings step:
   (a) Sample $\theta' \sim q(\theta'|\theta_{i-1})$
   (b) Run Algorithm 4 to obtain estimated log-likelihood $\hat{\ell}(\theta'|y_{1:T})$
   (c) Calculate acceptance probability α_i (Equation B.2)
   (d) Set $\theta_i = \theta'$ with probability α_i, otherwise set $\theta_i = \theta_{i-1}$
4. If $\text{mod}(i, 100) = 0$ AND $ESS_i > ESS$ AND $\hat{R}_i < \hat{R}$, return $\{\theta_i\}$, else GOTO 3.

The acceptance probability α_i at step i is calculated as:

$$\alpha_i = \min\left(1, \frac{\hat{P}(y_{1:T}|\theta')P(\theta')q(\theta_{i-1}|\theta')}{\hat{P}(y_{1:T}|\theta_{i-1})P(\theta_{i-1})q(\theta'|\theta_{i-1})}\right), \tag{B.2}$$

where θ_{i-1} is the previously accepted value of θ, θ' is the proposed value of θ, and $\hat{P}(y_{1:T}|\theta)$ is the estimated likelihood of the model given the data, calculated using Equation B.1. $q(\cdot|\cdot)$ denotes the proposal distribution (we use the adaptive multivariate normal approach from [40]) and $P(\theta)$ denotes the prior distribution for θ.

When using the fixed-lag bootstrap filter to estimate the model likelihood, we typically use $N_x = 1000$ particles, chosen to keep the standard deviation of log-likelihood estimates below approximately 1.2 [289]. When fitting to longer time series, such as the full REACT-1 dataset, we increase this to $N_x = 2000$. The algorithm is halted when the effective sample size (ESS) exceeds 100 and the Gelman-Rubin statistic ($\hat{R}$) is below 1.05. To reduce computation time, these criteria are checked every 100 iterations. We typically run three chains in parallel.

B.2.3 Marginal posterior distributions

Algorithm 4 produces samples from the conditional posterior distribution $P(X_t|y_{1:t+L}, \theta)$ while Algorithm 5 produces samples from the marginal posterior distribution for the model parameters $P(\theta|y_{1:T})$. Particles representing the marginal posterior distribution $P(X_t|y_{1:t+L})$ are obtained by repeatedly sampling $\theta^* \sim P(\theta|y_{1:T})$ and then running Algorithm 4 with this value of θ^*. This process is repeated, and the resulting particles are combined to approximate the marginal posterior distribution $P(X_t|y_{1:T})$. By default, we use 100 samples of θ and $N_x = 2000$ particles per filter, resulting in 2×10^5 particles representing the marginal posterior distribution.

The choice of fixed-lag L also depends on the application. When fitting a Markovian epidemic model (such as the model for the growth rate r_t), we can disable fixed-lag resampling while estimating the model likelihood to speed up computation. When estimating the marginal posterior distribution we use $L = 50$ by default, increasing to $L = 70$ when fitting to the full REACT-1 dataset to allow the model to "bridge the gap" between the end of study round 13 and the start of study round 14. When fitting a non-Markovian epidemic model (such as the model for R_t) we must choose L to include the majority of the mass of the generation time distribution and test-sensitivity curve - we use $L = 30$ in PMMH when fitting to REACT-1 data, again increasing this to $L = 50$ or $L = 70$ when estimating the marginal posterior distribution of the hidden states.

B.3 Eales approach knot spacing and equivalence with the SIMPLE approach

In this section we fit five models to REACT-1 data from rounds 14-to-19 (134 observations over 204 days), showing step-by-step the transition from the Eales approach to a slightly modified version of the SIMPLE approach.

All four implementations of the Eales approach use 200 warm-up iterations and 300 sampling iterations. As the data length is not divisible by 5, a target knot spacing of 5 days results in an effective spacing of 4.95 days. All other parameters are set to their default values.

For the Eales approach without splines, we fit the following model using Stan:

$$b_t \sim \text{Normal}\left(2b_{t-1} - b_{t-2}, \sigma_{Eales}\right),$$

$$P_t = \text{logit}^{-1}(b_t),$$

$$n_t^+ \sim \text{Beta-binomial}\left(n_t, \alpha_t, \beta_t\right),$$

where $\alpha_t = P_t(1/\rho - 1)$ and $\beta_t = (1 - P_t)(1/\rho - 1)$. Uniform prior distributions are used for b_1, b_2, and ρ, and a weakly informative inverse-gamma prior distribution with shape and scale 0.0001 is used for σ_{Eales}. All source code is available on GitHub.

For the SIMPLE approach with a logit link function, we implement the following epidemic model:

$$\tilde{r}_t = \tilde{r}_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{Normal}\left(0, \sigma\right),$$

$$\text{logit}P_t = \text{logit}P_{t-1} + \tilde{r}_t,$$

and retain the typical beta-binomial observation model. This model is equivalent to the Eales model without splines, expressed in a form better suited to SMC methods. To match the Eales model, we use the same inverse-gamma prior distribution for σ and artificially extend the data by 3 days at the start and finish, assuming $n_t = 0$ on these days. Minor differences remain due to how initial values are handled.

The Eales approach with a binomial observation model and spline knots every 5 days took 1m 59s to fit. Replacing the binomial distribution with a beta-binomial distribution increased the fitting time to 5m 13s due to the additional computation required, although the convergence

diagnostics (Rhat and ESS) improved as the beta-binomial model fits the data better. Decreasing the knot spacing from 5 days to 1 day further increased the fitting time to 8m 45s. Finally, fitting the Eales model without splines took a similar amount of time and produced identical results (allowing for Monte Carlo error). Fitting the equivalent model using SMC took 50s and produced nearly identical results. These results are summarised in Table B.1.

The sole reason for placing knots more than one day apart is to reduce computational burden. As the splines are unnecessary when fitting on the same timescale as our observations (unless we wish to interpolate at a higher resolution), and the SIMPLE approach (on $\text{logit}P_t$) is equivalent to the daily Eales approach without splines, the "ideal" Eales approach is equivalent to the SIMPLE approach on $\text{logit}P_t$. As the SIMPLE approach is approximately 9x faster to fit (on this dataset), we recommend using the SIMPLE approach over the Eales approach in this scenario.

This difference is expected to be greater on longer datasets, as the SIMPLE approach scales better with the number of observations, although for slowly-varying epidemics where knots can be placed farther apart without biasing results, there may be utility in using the Eales approach with splines.

Table B.1: Fitting diagnostics and parameter estimates for the binomial Eales approach with spline knots every 5 days, the beta-binomial Eales approach with spline knots every 5 days, the beta-binomial Eales approach with daily spline knots, the Eales approach without splines (a random walk on model coefficients), and a modified version of the SIMPLE approach modelling $\text{logit}P_t$.

Approach	Eales	Eales	Eales	Eales (no splines)	SIMPLE (logit)
Obs. model	Binomial	Beta-binom	Beta-binom	Beta-binom	Beta-binom
Knot spacing (days)	5	5	1	1	NA
Duration	1m 59s	5m 13s	8m 45s	8m 57s	50s
Max $\hat{R}$	1.06	1.02	1.04	1.03	1.01
Min ESS	59	230	273	210	143
$\hat{\sigma}$ (5-day)	0.22	0.16	NA	NA	NA
	(0.14, 0.39)	(0.098, 0.25)	NA	NA	NA
$\hat{\sigma}$ (1-day)	NA	NA	0.014	0.014	0.015
	NA	NA	(0.0089, 0.023)	(0.0087, 0.023)	(0.0093, 0.022)
$\hat{\rho}$ $\times 10^4$	NA	1.9	1.8	1.8	1.8
	NA	(1.0, 3.2)	(0.83, 3.0)	(0.90, 3.2)	(0.71, 3.1)

B.4 Comparisons with the original Eales approach

In this section we compare the original Eales approach with our modified version on the REACT-1 dataset.

The "original" approach in this section assumes a binomial observation distribution, sets the "max_treedepth" HMC hyperparameter to 10, and uses 1,000 warm-up iterations and 19,000 sampling iterations, with the warm-up and sampling iterations increased as needed to achieve convergence (Table B.2).

The "modified" approach in this section matches the approach described in the main paper: a beta-binomial observation distribution is used, the "max_treedepth" HMC hyperparameter is increased to 15, and the number of warm-up and sampling iterations is decreased to 200 and 300, respectively.

Both the original and modified approaches presented here have been updated from the source code to use new Stan syntax and cmdstanr (also necessitating updated processing code in R) and are initialised at plausible parameter values to reduce computation during initialisation. Our implementations of both the "original" and modified approaches are available on GitHub.

Runtime (measured on a 2021 M1 MacBook Pro) and convergence diagnostics are provided in Table B.2. On shorter periods with lower overdispersion (study rounds 8-to-13 and 14-to-16), the original approach converged within the default number of iterations, taking 1m 51s to 2m 36s to fit, while the modified approach took 35s to 56s. On periods where overdispersion was greater (study rounds 1-to-7 and 17-to-19), additional iterations were required for the original approach to converge, taking up to 17m 2s to fit, while the modified approach took up to 2m 58s to fit (study rounds 1-to-7). We were unable to obtain convergence of the original approach on the full dataset, even after 250,000 iterations and 31 hours of computation, while the modified approach completed in 37m 42s.

Figure B.3 presents results from fitting both models to the REACT-1 data by (grouped) study round. While both approaches produce similar results during rounds 8-to-13 and 14-to-16 (when overdispersion is estimated to be lower), there are substantial differences during rounds 1-to-7 and 17-to-19. The original approach produces estimates of r_t and P_t that are more variable,

as any extra-binomial noise must be absorbed by variation in r_t. Furthermore, the posterior predictive credible intervals for observed swab positivity are wider under the modified approach (see Figure B.4 for a closer look at rounds 1-to-7), due to the allowance for overdispersion.

Table B.2: Runtime, convergence diagnostics, and coverage of 95% credible intervals for n_t^+ for the original and modified Eales approaches. In some cases, additional iterations were required to obtain convergence of the original model. 50,000 iterations (2,000 warm-up + 48,000 sampling) were used for rounds 17-to-19 and 100,000 iterations (10,000 warm-up + 90,000 sampling) were used for rounds 1-to-7. We attempted up to 250,000 iterations on the full dataset, however convergence was not achieved.

Round	**All rounds**	**1-to-7**	**8-to-13**	**14-to-16**	**17-to-19**
Observations	400	147	119	67	67
Duration (days)	700	217	195	100	86
Original approach (1,000 warm-up + 19,000 sampling)					
Runtime	13m 12s	3m 27s	2m 36s	1m 51s	1m 49s
Max $\hat{R}$	3.02	1.39	1.01	1.00	1.12
Min ESS	3	8	577	648	18
Converged?	No	No	Yes	Yes	No
Coverage	-	-	94.1%	91.0%	-
Original approach (extra samples, see caption)					
Runtime	31h 4m 55s*	17m 2s	-	-	4m 29s
Max $\hat{R}$	3.62	1.05	-	-	1.02
Min ESS	2	74	-	-	122
Converged?	No	Weakly	-	-	Yes
Coverage	-	89.1%	-	-	94.0%
Modified approach (200 warm-up + 300 sampling)					
Runtime	38m 54s**	2m 58s	1m 14s	43s	35s
Max $\hat{R}$	1.04**	1.02	1.02	1.01	1.01
Min ESS	152**	380	259	245	260
Converged?	Yes	Yes	Yes	Yes	Yes
Coverage	96.5%	96.6%	97.5%	97.0%	98.5%

*This was timed on the Imperial College HPC cluster so the runtime is not directly comparable with the other approaches. **In order to obtain convergence of the modified approach on the full REACT-1 dataset, the "maximum tree-depth" HMC hyperparameter was increased from 15 to 16.

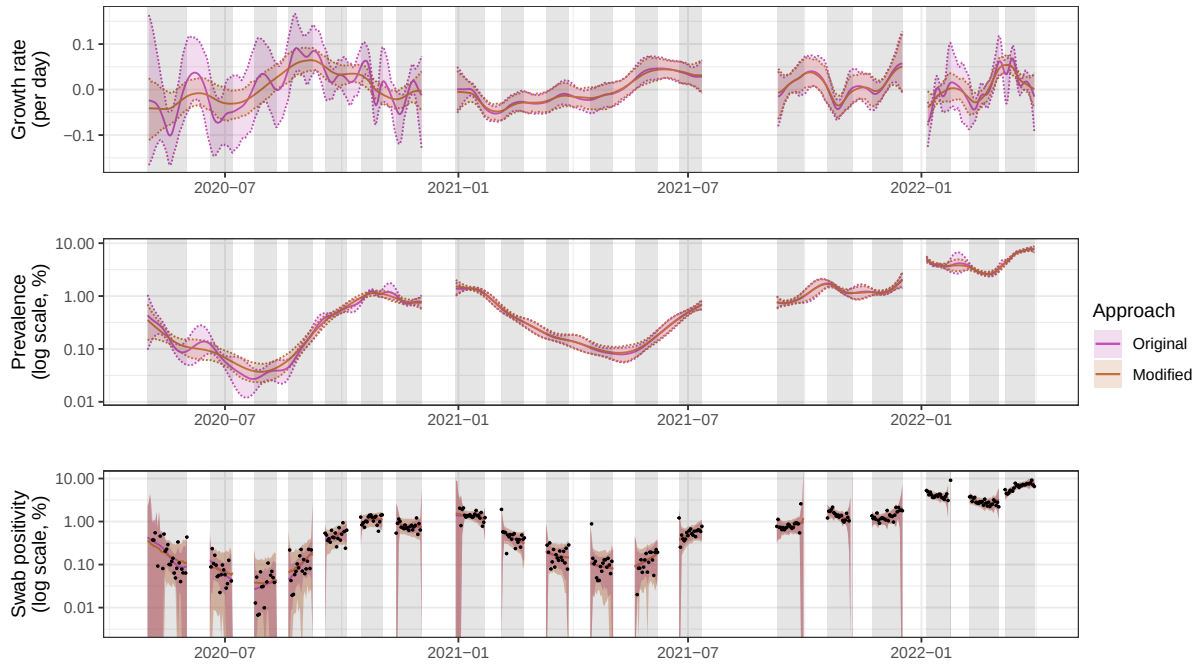

Figure B.3: Model results from fitting the original and modified Eales approaches to the REACT-1 dataset, grouped by study rounds 1-to-7, 8-to-13, 14-to-16, and 17-to-19. Solid lines show the posterior mean of each quantity, while shaded regions and dashed lines show 95% credible intervals. Grey shading indicates the periods when the data were collected.

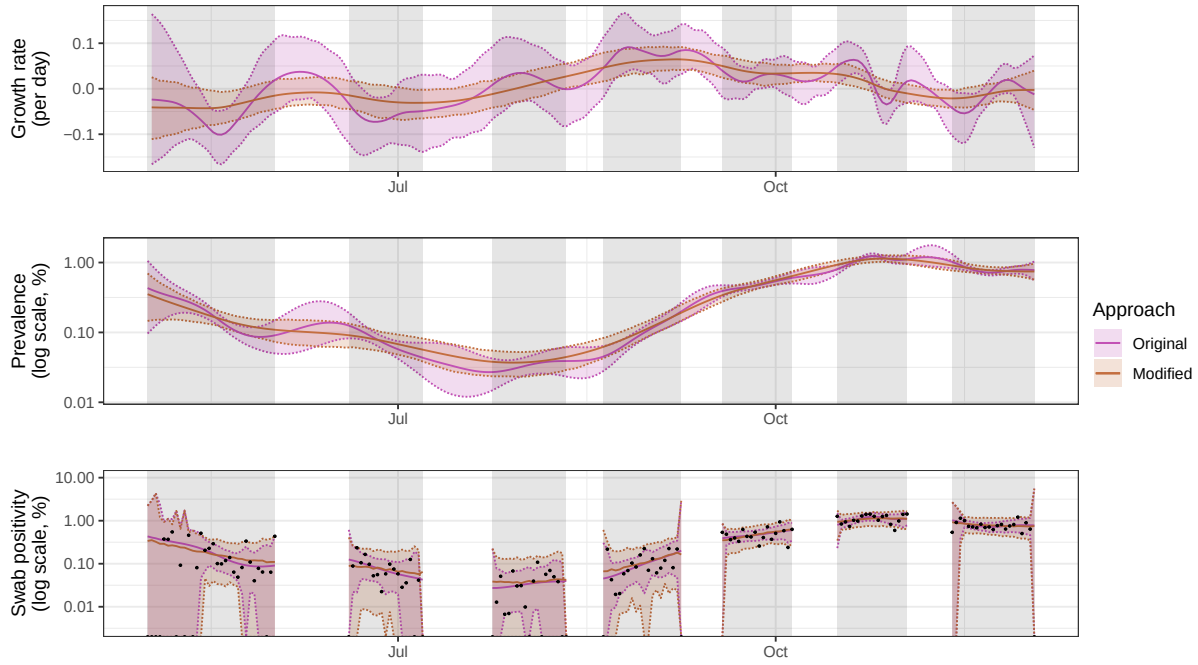

Figure B.4: Model results from fitting the original and modified Eales approaches to rounds 1-to-7 of the REACT-1 dataset. Solid lines show the posterior mean of each quantity, while shaded regions and dashed lines show 95% credible intervals. Grey shading indicates the periods when the data were collected.

B.5 RT-PCR test-sensitivity curves for SARS-CoV-2

The Abbott approach relies upon an externally-estimated infection-to-swab-positivity curve that describes how likely a given individual is to return a positive test result at a given time from infection. The use of this curve allows us to estimate infection *incidence* (and thus quantities such as the reproduction number) from prevalence data while accounting for the sensitivity of the test. By modelling swab positivity as a convolution of past incidence with this curve, we are also imposing additional smoothness assumptions on the observed data: even if the underlying incidence is not smooth, the observed prevalence will be.

By default, the Abbott approach uses a sensitivity curve estimated by Hellewell et al. [327], which reflects the sensitivity of self-administered nasopharyngeal RT-PCR tests by healthcare workers for SARS-CoV-2 in England in early 2020. A total of 241 PCR tests from 27 individuals were used to estimate this curve. Their median curve suggests a peak sensitivity of 78.5% at 4 days from infection and declines to a sensitivity of less than 10% at 20 days from infection.

Other estimates of time-varying RT-PCR sensitivity for SARS-CoV-2 are available and highlight considerable uncertainty in the literature. Binny et al. [46] used a total of 12,501 tests from 4,196 individuals that were tested using healthcare professional-administered nasopharyngeal RT-PCR tests for SARS-CoV-2 between June 2020 and November 2021. They found a peak sensitivity of 92.7% at 5 days from infection, declining to a sensitivity of less than 10% after 38 days. Kucirka et al. [342] pooled data from seven studies (a total of 1,330 swabs) to estimate a sensitivity curve peaking at 80.9% after 8 days, although their curve stops on day 21 at a sensitivity of 37%. These curves are presented in Figure B.5.

Substituting the Hellewell et al. (2021) curve for the Binny et al. (2023) curve and refitting the model to the REACT-1 dataset produces Figure B.6. While swab positivity estimates are largely similar, there are differences in the estimated growth rates. We also show the estimated infection incidence, demonstrating the substantial effect that the choice of sensitivity curve has on this quantity.

In the main paper, we hypothesised that the Abbott approach occasionally undercovers the growth rate due to additional smoothing assumptions imposed by the PCR sensitivity curve. As the Binny et al. (2023) curve represents a convolution over a longer time period, it enforces

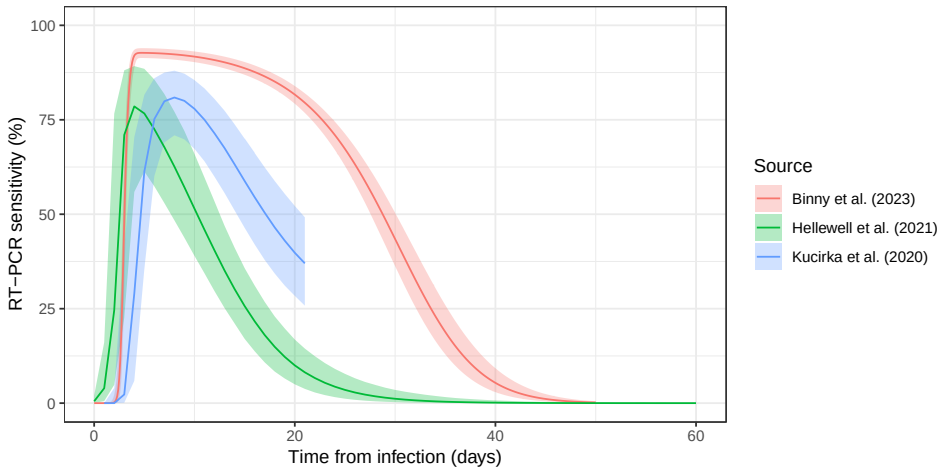

Figure B.5: Median estimated RT-PCR sensitivity for SARS-CoV-2 and corresponding 95% credible intervals from three different sources, highlighting substantial differences in the literature. The Abbott approach uses curves from Hellewell et al. (2021) [327].

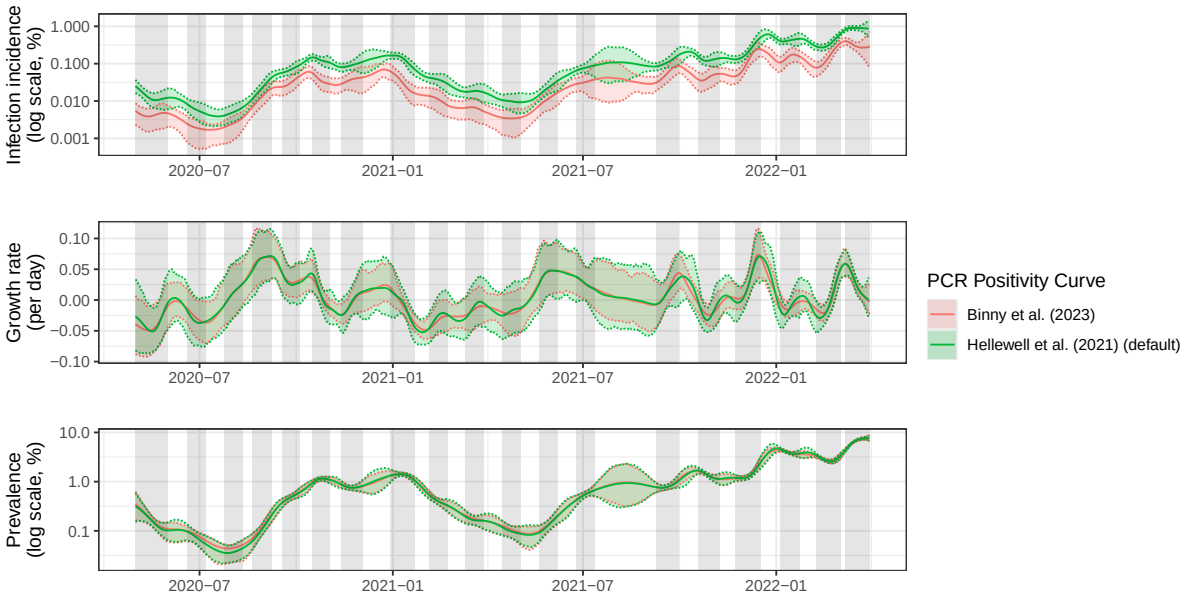

Figure B.6: Results from fitting the Abbott approach to REACT-1 data using two different PCR positivity curves.

additional smoothing on this quantity. We reproduce Figure 6.4 from the main paper in Figure B.7, additionally fitting the Abbott approach with the Binny et al. (2023) curve, and find that this hypothesis is supported. The Binny et al. (2023) curve results in even more smoothing of the growth rate estimates and also results in a decrease in accuracy of the estimated prevalence (when fit to simulations from the SIMPLE, Eales, and Abbott (Hellewell) approaches).

The SIMPLE approach also relies upon an externally-estimated infection-to-swab-positivity

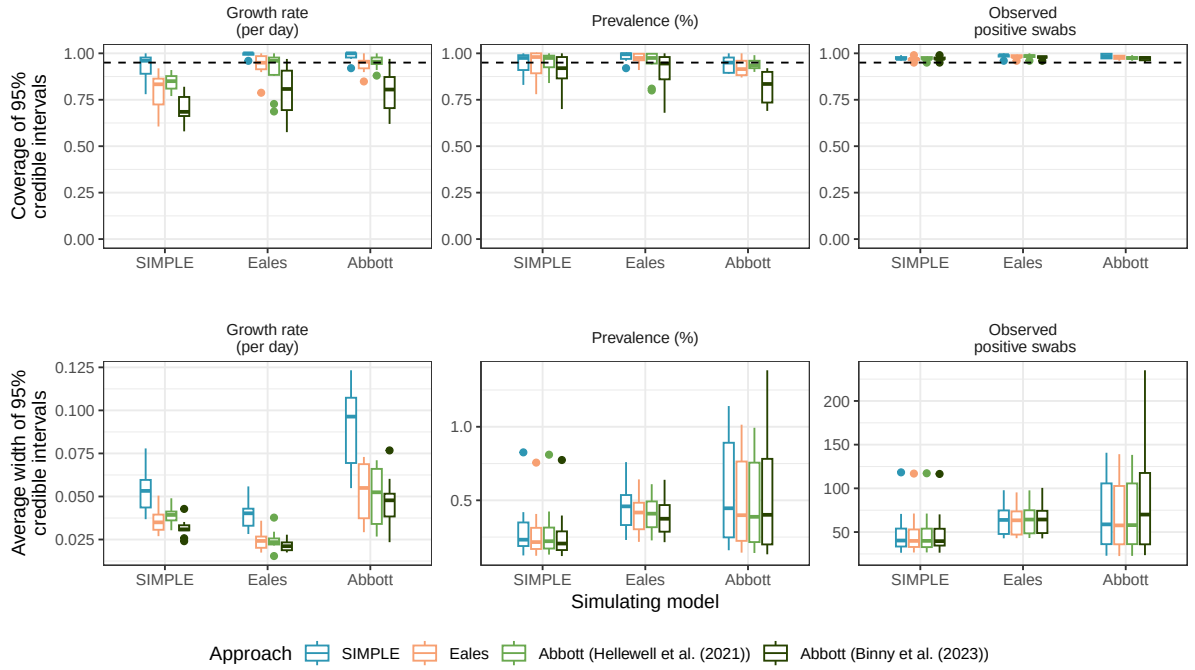

Figure B.7: Coverage and average width of 95% credible intervals for the growth rate, prevalence, and observed swab positivity from fitting the SIMPLE (extra-binomial), Eales, and Abbott approaches (with two different PCR positivity curves) to 10 simulated datasets from each model. Boxes present the interquartile range of the results with the median shown as a horizontal line. Whiskers extend to the most extreme data point within 1.5 times the interquartile range from the box. Outliers are shown as points.

curve when estimating the reproduction number R_t. In Figure B.8, we compare estimates by applying this approach to the REACT-1 dataset using the default Hellewell et al. (2021) curve alongside the Binny et al. (2023) curve. The latter curve results in a greater estimated value of σ_R of 0.10 $(0.068, 0.15)$ (the central estimate using the Hellewell curve is 0.069 $(0.050, 0.097)$), which is a result of the additional smoothing of incidence imposed by the Binny curve (thus R_t must become more variable to fit the same data). This greater value of σ_R, alongside the additional smoothing of incidence imposed by the Binny curve, means R_t fluctuates more, resulting in substantially wider credible intervals for R_t.

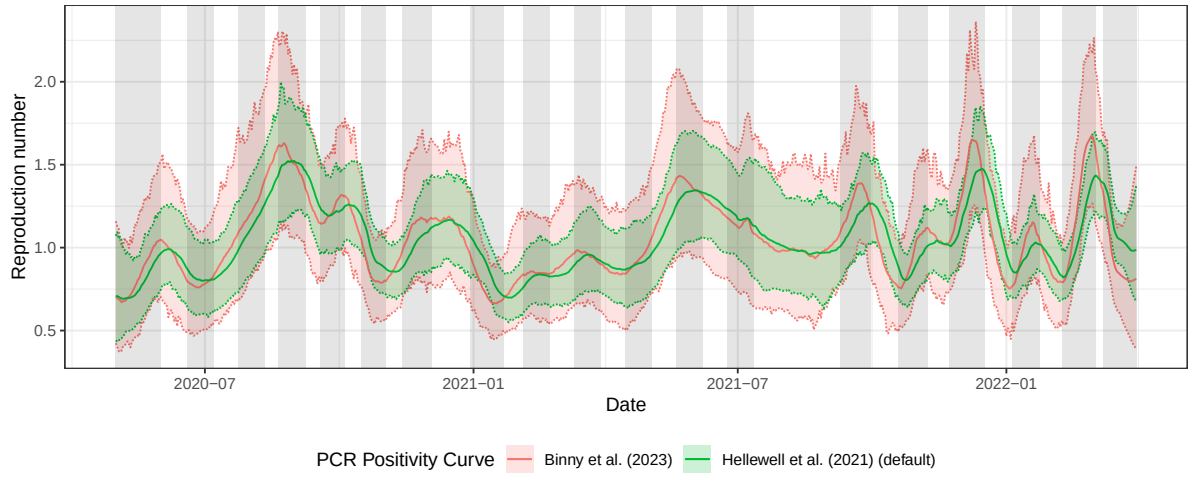

Figure B.8: Results from fitting the SIMPLE approach for R_t to REACT-1 data using two different PCR positivity curves. Solid lines show the posterior mean of each quantity, while shaded regions and dashed lines show 95% credible intervals. Grey shading indicates the periods when the data were collected.

B.6 REACT-1 survey weights

B.6.1 Observed survey weights

We claim in the main text that REACT-1 survey weights are well modelled by log-normal distributions with scale parameters ranging between 0.392 and 0.645. We present histograms of the observed survey weights from each study round, along with fitted log-normal distributions (using maximum likelihood estimation) in Figure B.9. The fitted distributions are generally in agreement with the data, although the observed weights are more centred around the mode than suggested by the fitted distributions. The "sdlog" parameter included in sub-figure headings is the scale parameter of the fitted log-normal distribution, which is equivalent to the parameter ξ that we use when generating synthetic weighted data.

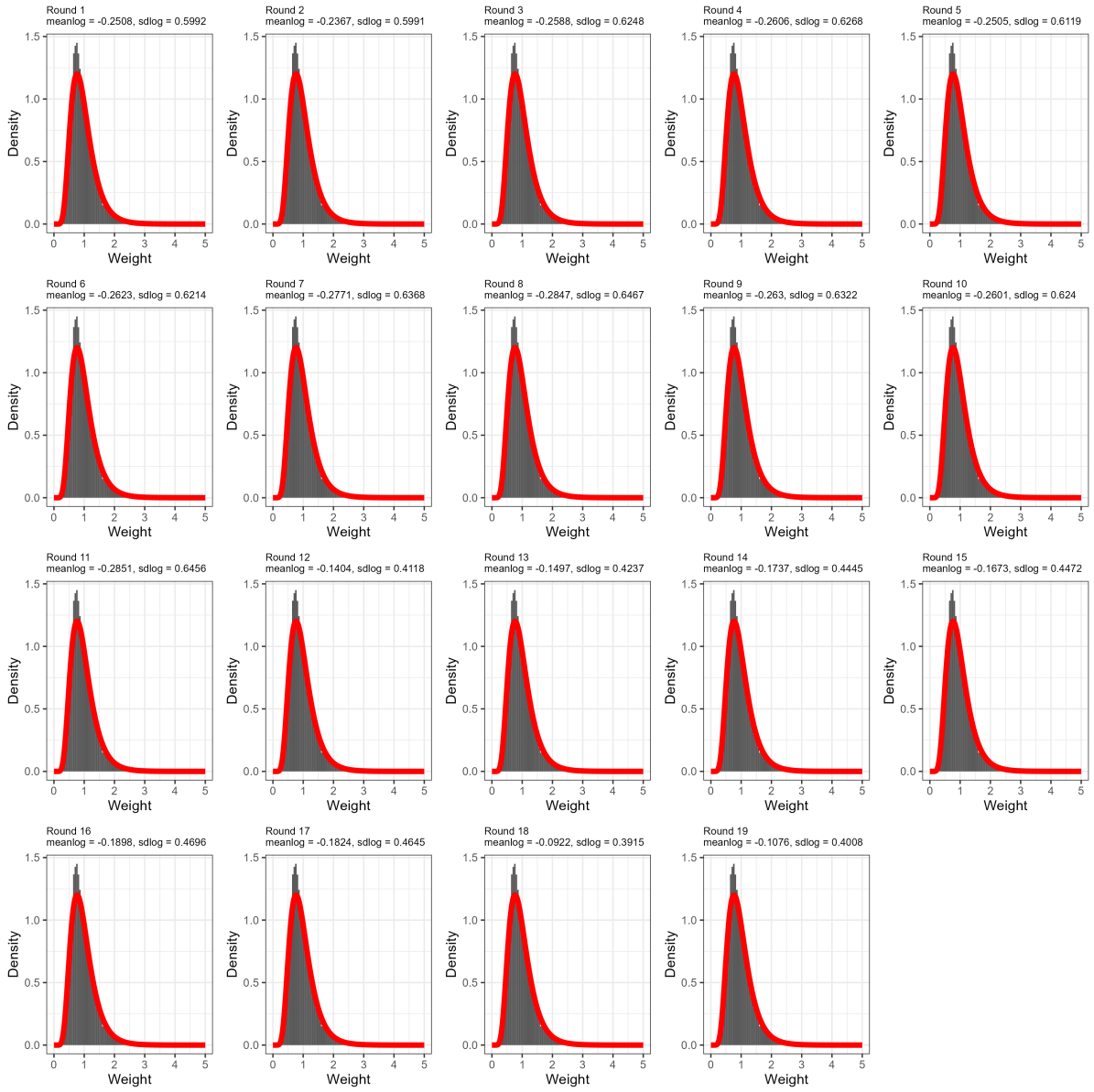

Figure B.9: Histogram of REACT-1 survey weights by study round (grey bars) and fitted (by maximum likelihood estimation) log-normal distributions (red lines). The parameters of the fitted distributions are given in the subheadings of each panel.

B.6.2 Survey weights and observed swab positivity

While we do not explicitly model the relationship between survey weights and observed swab positivity when fitting the model to data, it is informative to investigate this relationship empirically. Specifically, we fit the following logistic regression model to REACT-1 data using the *glm* function in R:

$$\log \frac{P_i}{1-P_i} = \beta_0 + \beta_1 \log w_i,$$

where P_i is the probability that the i^{th} swab, with survey weight w_i, is positive.

Very crudely, if we assume that $P_i = kw_i^q$ and that P_i is small, then we can show that $\beta_0 \approx q \log k$ and $\beta_1 \approx q$. Thus, the β_1 coefficient tells us about the relationship between survey weights and observed swab positivity. Specifically, $\beta_1 > 0$ implies those with greater weights (lower response probability) are more likely to test positive, and $\beta_1 = 1$ implies a linear relationship (like the one we assume for simulated weighted data).

Table B.3 presents summary statistics of observed REACT-1 survey weights by round, and the β_1 coefficient (and associated p-value) from this logistic regression model. In 14 out of 19 study rounds, the β_1 coefficient is statistically significantly greater than 0 at the $\alpha = 0.05$ significance level, suggesting that those with greater survey weights (i.e., people with characteristics associated with lower response probabilities) are more likely to test positive, although the estimated value is always less than 1, suggesting a sublinear relationship. This relationship also appears to break down in study rounds 18 and 19, rounds in which the population prevalence was greatest and incentives to improve response rates were introduced.

Table B.3: Summary statistics and logistic regression coefficient (see text) for observed survey weights in the REACT-1 study.

Round	Mean weight	Median weight (2.5th quantile, 97.5th quantile)	Coefficient (p-value)
1	0.953	0.724 (0.284, 2.96)	0.307 (p = 0.017)
2	0.964	0.736 (0.287, 2.88)	0.246 (p = 0.089)
3	0.96	0.718 (0.269, 3.03)	0.372 (p = 0.069)
4	0.961	0.717 (0.271, 3.07)	0.610 (p <0.001)
5	0.96	0.723 (0.282, 3.01)	0.408 (p <0.001)
6	0.958	0.714 (0.276, 3.09)	0.298 (p <0.001)
7	0.956	0.697 (0.273, 3.16)	0.361 (p <0.001)
8	0.954	0.693 (0.266, 3.19)	0.299 (p <0.001)
9	0.962	0.711 (0.266, 3.08)	0.324 (p <0.001)
10	0.959	0.719 (0.269, 3.02)	0.438 (p <0.001)
11	0.949	0.702 (0.253, 3.09)	0.221 (p = 0.112)
12	0.952	0.830 (0.440, 2.19)	0.762 (p <0.001)
13	0.949	0.820 (0.430, 2.23)	0.696 (p <0.001)
14	0.938	0.794 (0.408, 2.32)	0.446 (p <0.001)
15	0.944	0.797 (0.416, 2.32)	0.567 (p <0.001)
16	0.937	0.777 (0.391, 2.42)	0.618 (p <0.001)
17	0.939	0.786 (0.394, 2.39)	0.400 (p <0.001)
18	0.987	0.899 (0.426, 2.09)	-0.0357 (p = 0.473)
19	0.978	0.869 (0.425, 2.19)	0.0199 (p = 0.520)

B.7 Supplementary results (simulated)

B.7.1 Survey design and epidemic dynamics: hidden-state estimation

In the main paper, we considered the coverage and width of the 95% credible intervals for prevalence P_t when fit to data from each model. We reproduce these results for the estimated growth rate r_t in Figure B.10. Unlike P_t, fitting non-weighted models to weighted data does not impact estimates of the growth rate r_t, although this assumes there are no temporal biases in survey weights. The basic model still struggles to estimate r_t when the data are overdispersed, however, as the estimated r_t must fluctuate significantly to account for observed noise in the data.

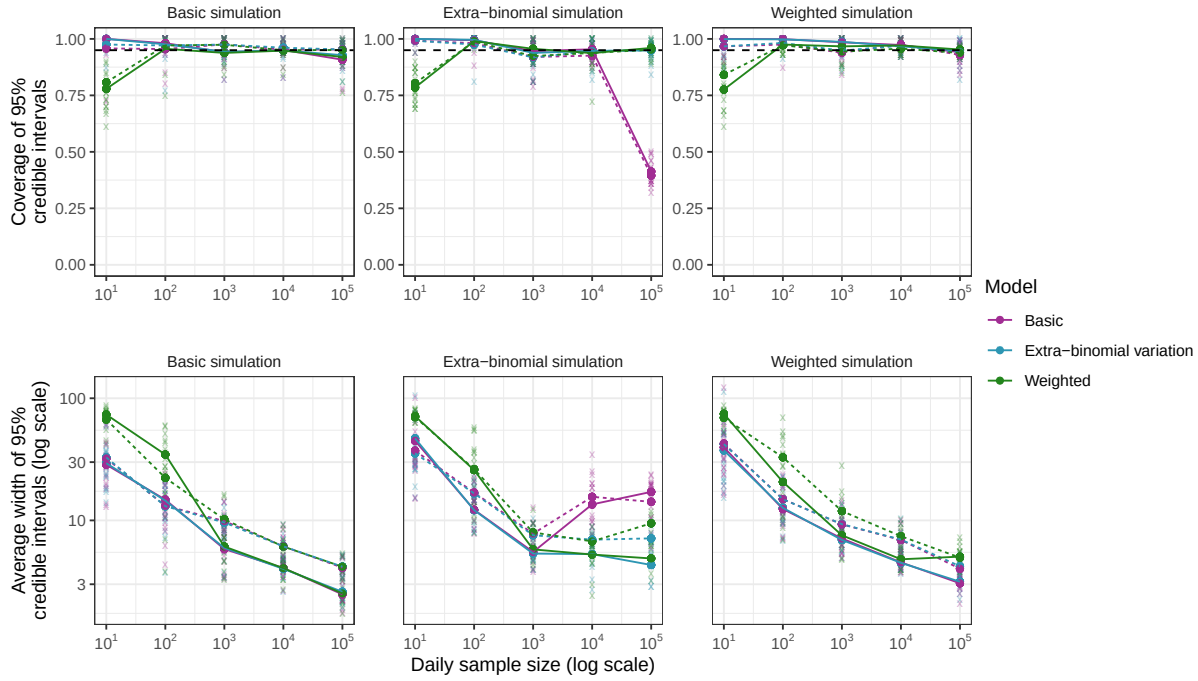

Figure B.10: Coverage and average width of 95% credible intervals for the estimated growth rate r_t from fitting all three models (basic: purple, extra-binomial: blue, weighted: green) to simulated data from each model (basic: column 1, extra-binomial: column 2, weighted: column 3). Results from individual simulations are shown as transparent crosses, with averages over 10 simulations shown as points connected by solid lines (for assumed $\sigma = 0.008$) and dashed lines (for assumed $\sigma = 0.016$). A range of assumed daily sample sizes n_t are considered (x-axis). The horizontal black dashed line indicates the target coverage of 95%.

B.7.2 Survey design and epidemic dynamics: parameter estimation

In addition to coverage and credible interval width for the hidden states, we also consider how the estimation of model parameters varies with survey design and epidemic dynamics. Figure

B.11 presents the estimated value of σ, ρ (where applicable) and c (where applicable). As expected, higher n_t results in more precise estimates of model parameters. All three models produce good estimates of σ, even when there is no extra-binomial variation or survey weight bias, suggesting there is little cost to including the additional parameter. When fitting the basic model to data featuring extra-binomial variation, the estimated value of σ increases to compensate for the additional noise, with a greater effect observed at higher values of n_t. When fitting the weighted model to data featuring extra-binomial variation, the variance adjustment factor c increases to account for this, so the weighted model still produces good estimates of σ.

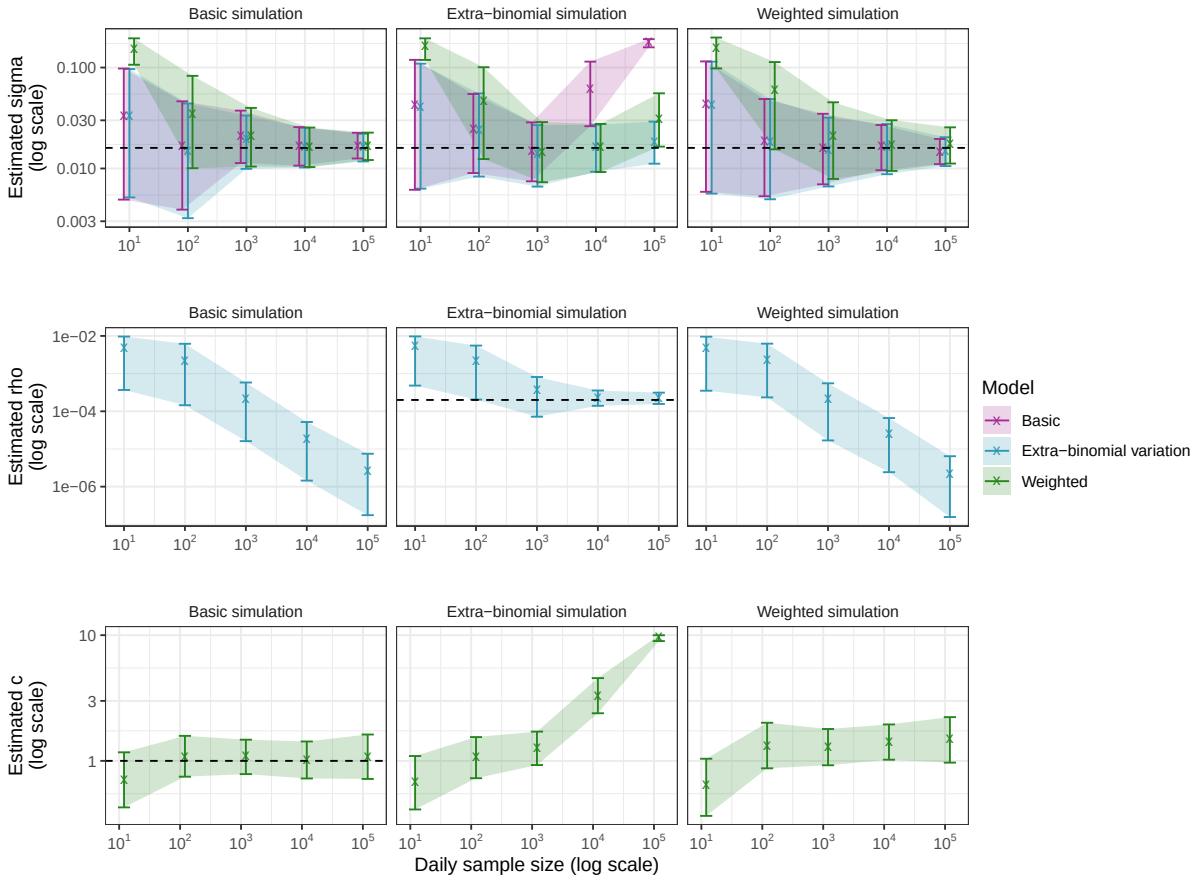

Figure B.11: 95% credible intervals for model parameters from fitting all three models (basic: purple, extra-binomial: blue, weighted: green) to simulated data from each model (basic: column 1, extra-binomial, column 2, weighted: column 3) at varying values of n_t and $\sigma = 0.016$. Black dashed lines denote true parameter values where applicable.

B.8 Comparing model runtimes

Runtimes reported in the main paper reflect a single successful run of each model. However, runtimes can vary substantially across runs, even for the same model and dataset. To assess relative performance more robustly, we repeat each fitting procedure multiple times.

Observed runtimes depend on many factors, including dataset length, model complexity, parameterisation, model fit, and the system used to fit the model. The results in this section are based on subsets of the REACT-1 dataset and may not generalise to other settings. We use real data to avoid artefacts from fitting models to their own simulations, but caution that these comparisons remain context-specific.

We test each model on multiple subsets of the REACT-1 dataset of increasing length, always starting from study round 1 (1 May 2020). Each subset is defined by its final study round—rounds 1 through 13 in sequence, followed by rounds 15, 17, and 19 to reduce computation. Most models are fit 10 times per subset, except for the SIMPLE approach (reproduction number epidemic model) and the Eales approach, which are run 5 times each for rounds 13, 15, and 17, and 3 times for round 19.

All models use their default parameterisations, except for the Eales approach on all 19 rounds, where the HMC "max_treedepth" is increased to 16 to ensure convergence. For the SIMPLE approach with the reproduction number epidemic model, we use $N = 2000$ particles per PMMH likelihood evaluation for datasets ending at round 13 or later.

We present two runtime metrics: the time taken to complete a single successful fit (Figure B.12, upper panel), and the average time taken to achieve convergence (Figure B.12, lower panel). The latter accounts for failed fits, so is always equal to or greater than the former. As convergence is built into the stopping criteria for the SIMPLE approaches, only the Eales and Abbott approaches are affected by failed fits.

Stan models may fail to converge for two main reasons: (1) too few iterations, and (2) divergent chains. The first occurs when the data are too short (providing limited information) or too long (introducing many parameters), and affects both Stan-based models. The second (divergent chains) can occur regardless of dataset length and primarily affects the Abbott model. Improved

parameter initialisation may reduce the frequency of these divergent chains, but is context-dependent and not explored further here.

The SIMPLE approach (growth rate epidemic model) and Abbott approach have comparable runtimes when successful, though the non-zero failure rate of the Abbott approach increases its average time per successful fit. The SIMPLE approach (reproduction number epidemic model) and the Eales approach are substantially slower overall. While runtimes for these two approaches are similar on shorter datasets, the SIMPLE approach (reproduction number epidemic model) is faster than the Eales approach on the longest dataset.

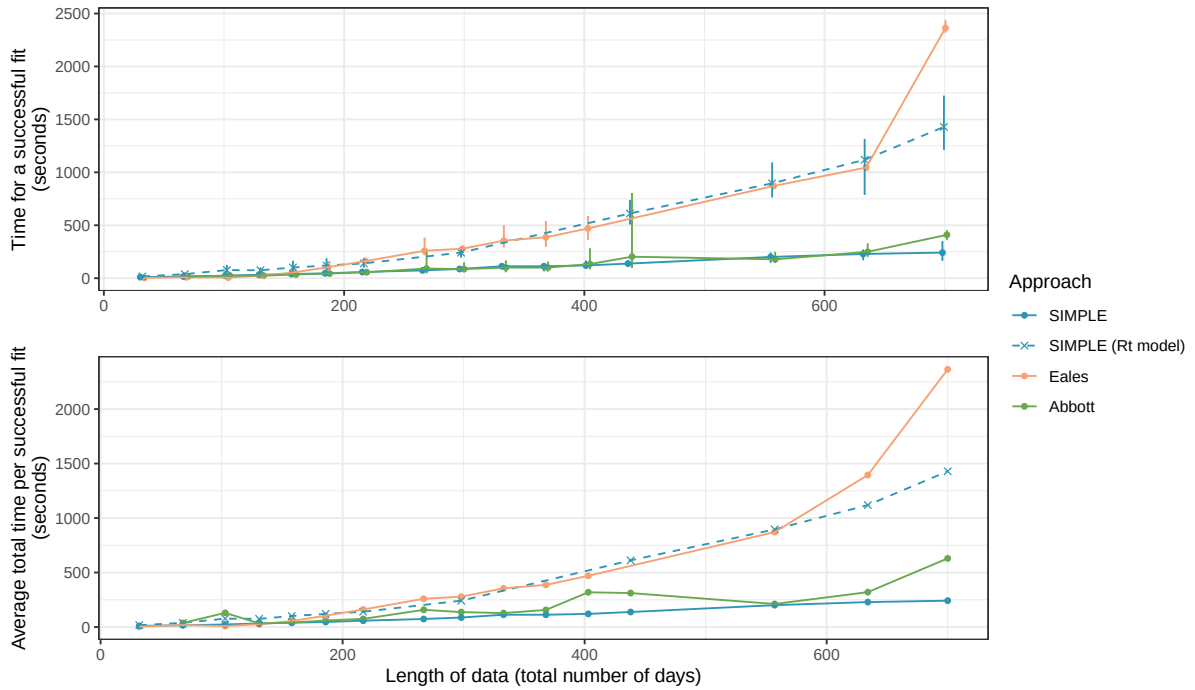

Figure B.12: Runtimes for fitting each model to subsets of the REACT-1 dataset of increasing length (always starting from study round 1). The upper panel shows the average time taken to complete a single successful fit (points), with error bars showing the observed range (minimum to maximum). The lower panel shows the average time taken to achieve convergence, which is equal to or greater than the time for a single fit due to occasional failures. The Abbott approach failed to converge on the shortest dataset (round 1 only, 32 days), though this could likely be resolved by adjusting hyperparameters.

B.9 Supplementary results (REACT-1)

B.9.1 Full study versus separate models

In the main paper, we present estimates of r_t, P_t, and n_t^+/n_t from fitting to all 19 REACT-1 study rounds simultaneously (Figure 6.5), despite substantial variation in parameter estimates when fitting to shorter time periods (Table 6.2). In Figure B.13, we compare hidden-state estimates from the full model with those obtained by fitting to shorter periods. Estimates are similar across rounds 1-to-7 and 8-to-13, although slightly wider credible intervals for r_t are observed in rounds 14-to-16 and 17-to-19. These reflect greater uncertainty about model parameters and higher estimated values of σ.

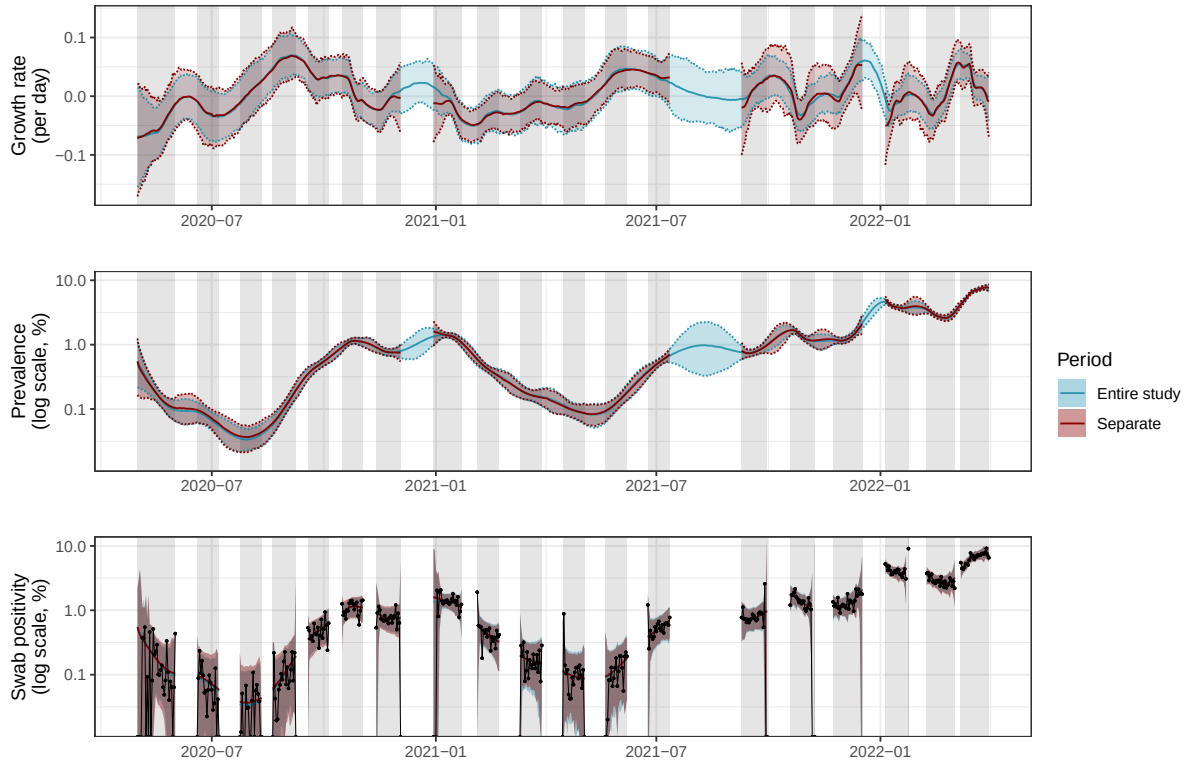

Figure B.13: Estimates of the growth rate r_t, prevalence P_t, and observed swab positivity n_t^+/n_t for SARS-CoV-2 in England between 1 May 2020 and 31 March 2022, based on data from the REACT-1 study. Blue curves present estimates from fitting to all 19 study rounds simultaneously while dark red curves present estimates from fitting to four shorter periods. Solid coloured lines show central estimates while shading and dashed lines show 95% credible intervals. Daily true observed swab positivity n_t^+/n_t is shown in black points.

B.9.2 Overdispersed versus basic model

We present estimates of r_t, P_t, and n_t^+/n_t from the REACT-1 study using the beta-binomial model, allowing for extra-binomial variation in the data. We compare these estimates to those from the basic model in Figure B.14. While estimates in rounds 8-to-13 and 14-to-16 are similar between models, estimates in rounds 1-to-7 and 17-to-19 are noticeably different - these are the periods in which the estimated value of ρ is higher (Table 6.2).

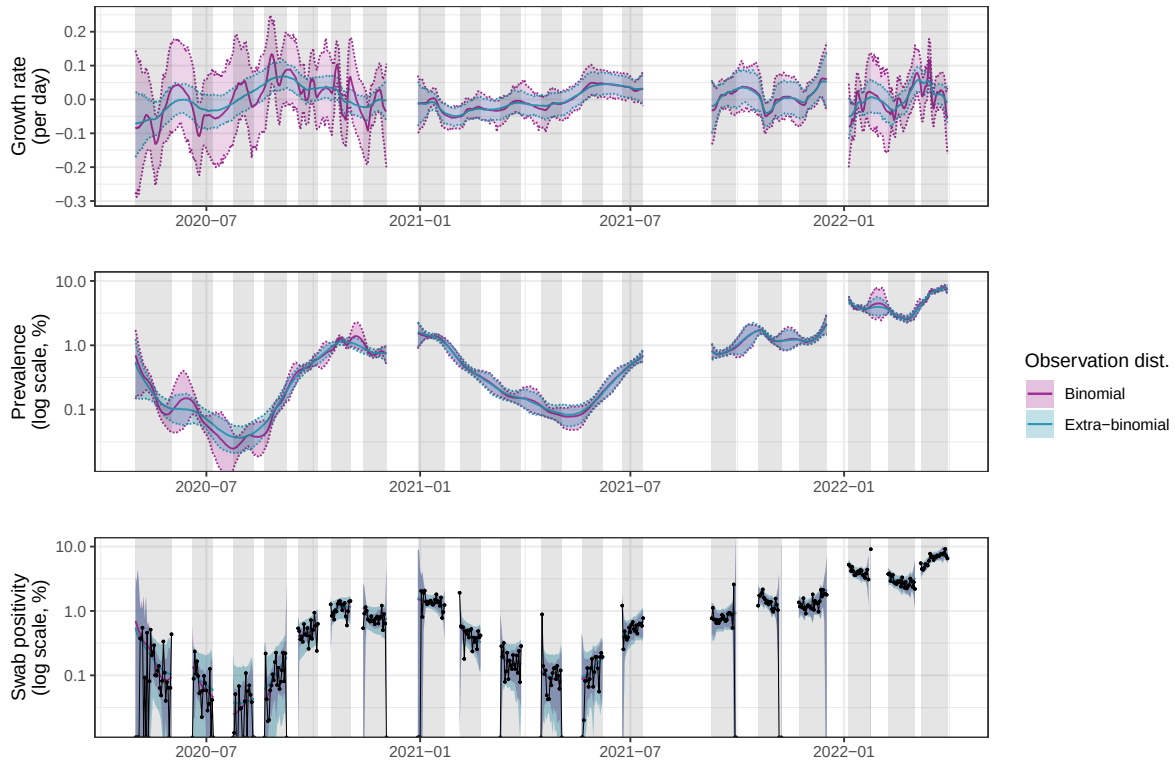

Figure B.14: Estimates of the growth rate r_t, prevalence P_t, and observed swab positivity n_t^+/n_t for SARS-CoV-2 in England between 1 May 2020 and 31 March 2022 using data from the REACT-1 study. Blue curves present estimates from the model allowing for extra-binomial variation while pink curves present estimates from the basic model. Solid coloured lines show central estimates while shading and dashed lines show 95% credible intervals. Daily true observed swab positivity n_t^+/n_t is shown in black points.

B.9.3 Weighted data

As survey weights applicable on a daily basis were not available for the REACT-1 study, we have thus far not fit the weighted model to these data. In Figure B.15 we compare the extra-binomial model with the weighted model when fit to REACT-1 data, assuming that the provided survey weights (which are applicable on a round-by-round basis) are valid for use with

daily data. Growth rate estimates are largely unchanged, although there are periods where population swab positivity is estimated to be greater in the weighted model.

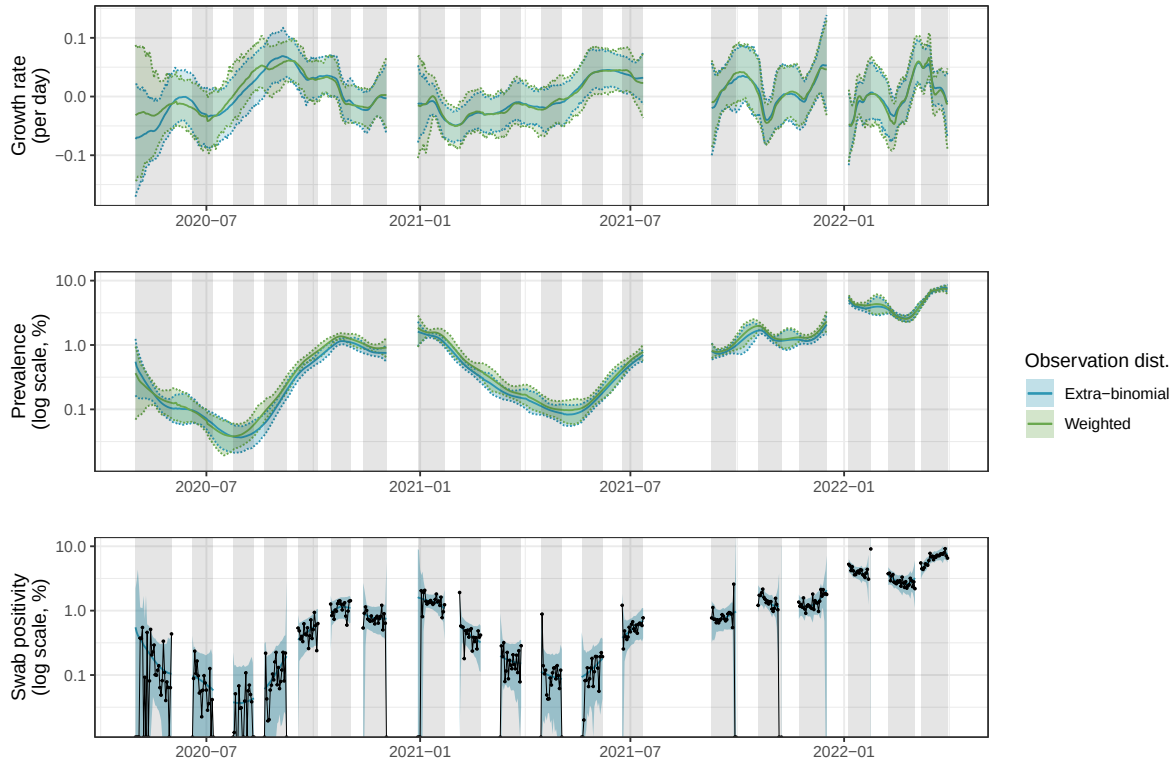

Figure B.15: Estimates of the growth rate in swab positivity r_t, population swab positivity P_t, and observed swab positivity n_t^+/n_t for SARS-CoV-2 in England between 1 May 2020 and 31 March 2022 using data from the REACT-1 study. Blue curves present estimates from the model allowing for extra-binomial variation while green curves present estimates from the weighted model, assuming that survey weights are valid on a day-by-day basis. Solid coloured lines show central estimates while shading and dashed lines show 95% credible intervals. Daily true observed swab positivity n_t^+/n_t is shown in black points.

B.10 Accounting for multiple variants

Genomic sequencing has become an increasingly important tool for monitoring the spread of infectious diseases. This was most notable during the COVID-19 pandemic, where the repeated emergence of new variants of concern necessitated the monitoring of variant-specific growth rates. Genomic sequencing of survey samples is time-consuming and expensive, so the estimation of variant-specific growth rates using sequencing from only a subset of samples is a key problem. We show how the SIMPLE approach can be extended to allow for this.

Specifically, using superscript $v = 1, \ldots, V$ to denote variant, we define variant-specific growth rates and swab positivity:

$$r_t^{(v)} = r_{t-1}^{(v)} + \epsilon_t^{(v)}, \quad \epsilon_t^{(v)} \sim N(0, \sigma_r^{(v)}),$$

$$P_t^{(v)} = P_{t-1}^{(v)} e^{r_t^{(v)}}.$$

We retain the standard beta-binomial observation distribution for the number of total positive swabs, with overall prevalence defined as $P_t = \sum_v P_t^{(v)}$. We further assume that n_t^s of the positive swabs are sent for genomic sequencing and that number of these swabs that are positive for each variant considered follows a multinomial distribution. Specifically:

$$n_t^+, \left\{n_t^{(v)}\right\} \sim \text{Beta-binomial}(n_t, \alpha_t, \beta_t) \times \text{Multinomial}\left(n_t^s, \left\{\frac{P_t^{(v)}}{P_t}\right\}\right),$$

where the "$\times$" operator denotes the product of two probability density functions. $\alpha_t = P_t(1/\rho - 1)$ and $\beta_t = (1 - P_t)(1/\rho - 1)$. We now have $V + 1$ parameters: $\sigma_r^{(v)}$, the standard deviation of the daily growth rate for each variant, and ρ, the overdispersion parameter.

The hidden states associated with a given variant are initialised on the first day the variant is detected. We stop estimating the associated hidden states at the end of the final round in which a given variant is detected.

We fit this model to REACT-1 data in two groups: rounds 8-to-13 (for Wildtype (WT), Alpha (AL), and Delta (DE) variants) and rounds 14-to-19 (for Delta, Omicron (OM), and Omicron BA.2 (BA) variants). Table B.4 presents parameter estimates and 95% credible intervals. Figure B.16 shows the estimated growth rates and prevalence for each variant. Crucially these

estimates leverage data from both sequenced and unsequenced swabs to inform estimates of variant-specific growth rates.

Table B.4: Parameter estimates for the variant model fit to REACT-1 data.

Rounds	$\sigma^{(WT)}$	$\sigma^{(AL)}$	$\sigma^{(DE)}$	$\sigma^{(OM)}$	$\sigma^{(BA)}$	ρ ($\times 10^4$)
8-to-13	0.039 (0.0014, 0.13)	0.0077 (0.0013, 0.018)	0.014 (0.00089, 0.051)	- -	- -	1.3 (0.53, 2.3)
14-to-19	- -	- -	0.016 (0.0084, 0.029)	0.035 (0.020, 0.062)	0.020 (0.007, 0.048)	1.7 (0.54, 3.3)

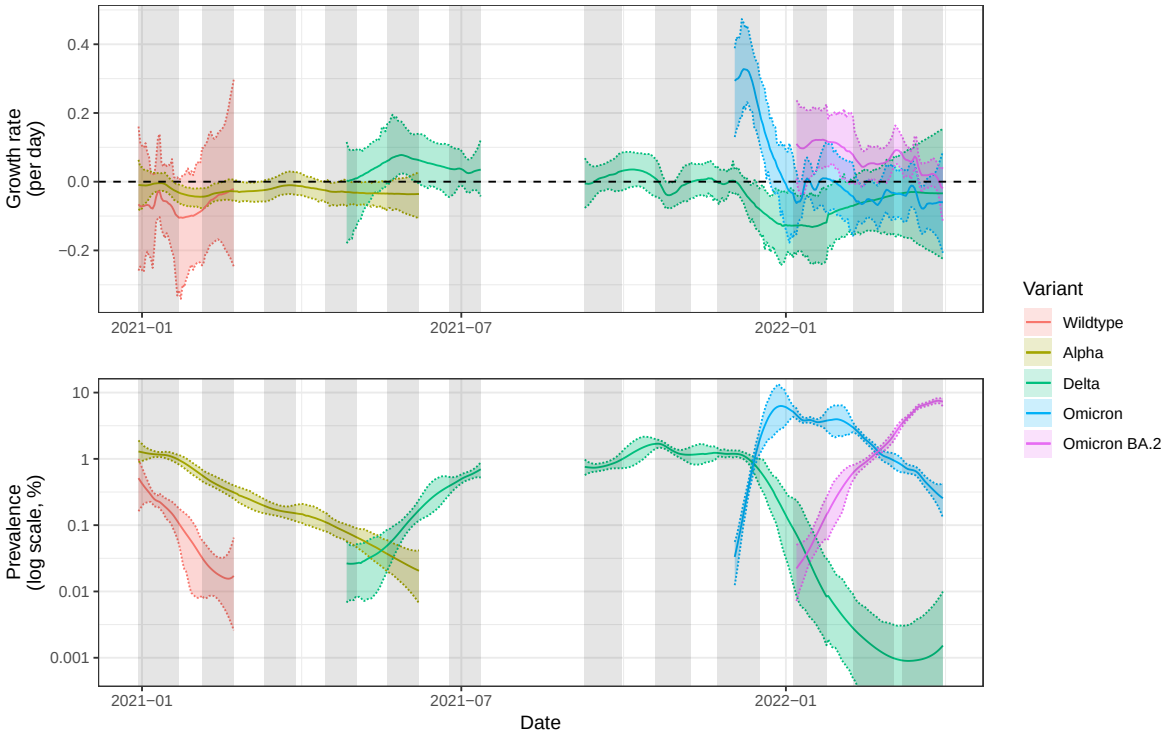

Figure B.16: Estimated growth rates and prevalence of SARS-CoV-2 by variant between 30 December 2020 and 31 March 2022 using data collected in the REACT-1 study. Solid lines show the posterior mean of each quantity and shaded regions/coloured dashed lines show 95% credible intervals. Shaded regions indicate the periods when the data were collected. The horizontal black dashed line shows zero growth rate.

The observation model leveraged above can be derived by assuming that the probability of sequencing a given swab is independent of the variant, and that the total number of sequenced swabs is deterministic given the total number of positive swabs. Thus:

$$P\left(\{n_t^{(v)}\}, n_t^+, n_t^s | n_t, \cdots\right) = P\left(\{n_t^{(v)}\} | n_t^s, \cdots\right) P(n_t^s | n_t^+, \cdots) P(n_t^+ | n_t, \cdots),$$

where the first term is the multinomial distribution, the middle term is a Dirac delta at the observed n_t^s, and the final term is the beta-binomial distribution.

This framework readily allows for further analysis. Samples of growth rate advantages ($r_t^{v_a}/r_t^{v_b}$) can be obtained by comparing $r_t^{v,i}$ values across variants v, where i indexes the particle number. Forecasts of variant prevalence can be obtained by projecting the model forward in time.

B.11 Comparisons with UKHSA consensus estimates

During the COVID-19 pandemic, the UK Health Security Agency (UKHSA) produced consensus estimates of the growth rate and the reproduction number in England. These estimates were produced by combining estimates from a range of models that were fit to a variety of data sources (not including data from REACT-1) [192, 340]. The UKHSA estimates are the most official estimates of the growth rate and the reproduction number in England, so it is informative to compare our estimates with these.

UKHSA consensus estimates were reported, and are reported here, as the lower and upper bounds of the 90% credible intervals for the growth rate and the reproduction number. No central estimates were provided. Estimates were produced weekly and reflect data up to a few days before the estimate was produced (the exact gap varies by week). Figure B.17 presents our estimates of the growth rate and the reproduction number from applying all approaches to the REACT-1 dataset, alongside the UKHSA estimates (shown in black).

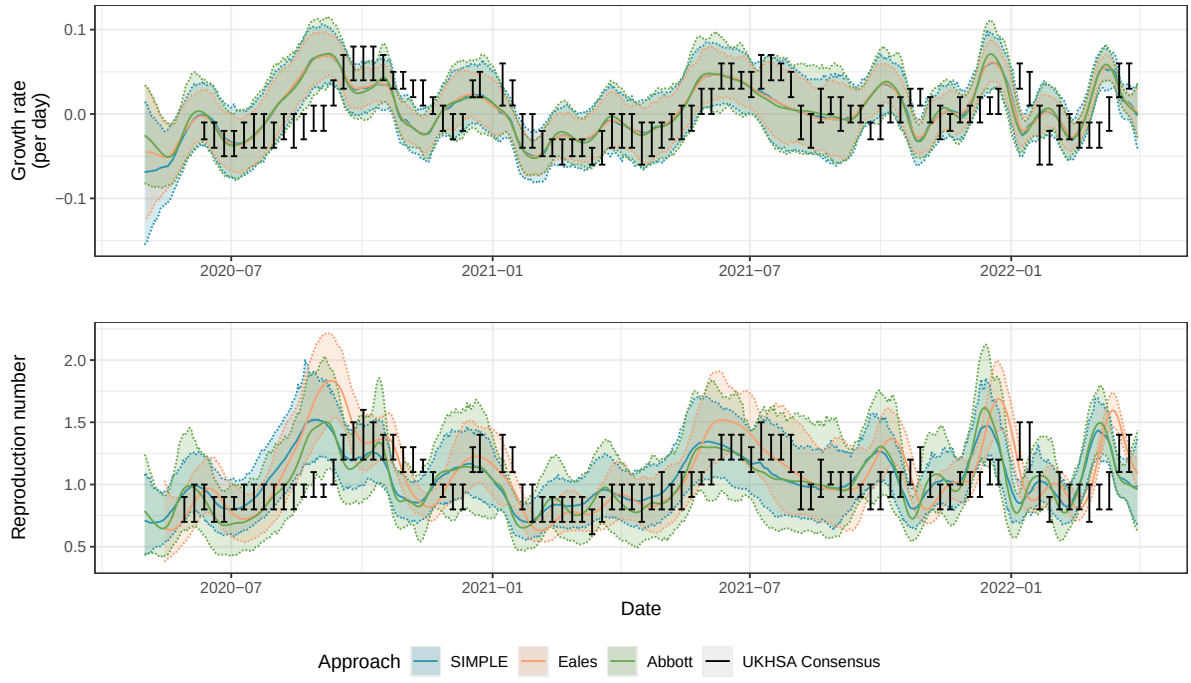

Figure B.17: Comparison of our estimates of the growth rate and the reproduction number for SARS-CoV-2 in England with UKHSA consensus estimates. Our estimates are shown in blue (SIMPLE approach), orange (Eales approach), and green (Abbott approach). The UKHSA estimates are shown in black, with the lower and upper bounds of the 90% credible intervals shown as horizontal lines.

There are many reasons why our estimates differ from the UKHSA estimates:

1. The UKHSA estimates are consensus estimates produced by combining estimates from a range of methods. This generally results in narrower credible intervals than those produced by any individual method, particularly where the individual models contributing to the consensus estimate agree. Consensus estimates can also result in more conservative estimates (closer to 0 for the growth rate and 1 for the reproduction number), as individual model variance is averaged out.

2. The UKHSA estimates are produced in real-time and are not updated retrospectively, whereas our approaches all use the full dataset to produce estimates. This explains why the UKHSA estimates lag behind our estimates: an increase in cases the following week can be explained by an increase in r_t and R_t today, but the UKHSA estimates will not reflect this until the increase is observed.

3. The UKHSA estimates are produced from different data sources to our estimates. The REACT-1 study did not contribute to the UKHSA estimates, which leveraged data such as reported cases and hospitalisations (among other, individual model-dependent, data sources).

4. All models make their own set of assumptions. In addition to serial interval and smoothing differences discussed in the main paper, the model structure can also impact estimates. Not all of the individual models in the UKHSA consensus estimates used the renewal model, for example.

5. The UKHSA report 90% credible intervals, while we report 95% credible intervals.

While consensus estimates like those produced by the UKHSA are generally more robust than estimates from any individual model, we are hesitant to use these comparisons as evidence that our approaches are producing inaccurate estimates. Points (2) and (3) above are particularly important: our estimates are produced retrospectively using high-quality data and are not subject to the same real-time constraints as the UKHSA estimates. Data that are available in real-time (such as reported cases and hospitalisations) are noisy and subject to biases which will impact all models that use them (and thus the consensus estimates). The instances where credible intervals do not overlap in Figure B.17 largely occur due to delays in the UKHSA

estimates, which would likely not exist if the estimates were produced using retrospective data.

Appendix C

Supplementary material for Chapter 7: Jointly estimating epidemiological dynamics of COVID-19 from case and wastewater data in Aotearoa New Zealand

C.1 Supplementary methods

C.1.1 Data

Wastewater sampling

Composite samples were collected by an autosampler, which collects a small volume of wastewater at regular intervals over the course of a 24-hour period. When composite samples were not available, 'grab' samples were collected and ranged from a sample being taken at a single point in time to three samples taken over 30 minutes. Grab samples represent only the composition of the source at the time of collection and may not be as representative as a 24-hour composite sample collected by an autosampler. Following collection, samples were couriered overnight to the Institute of Environmental Science and Research (ESR) for processing.

Virus concentration was performed by SARS-CoV-2 detection and quantitation by RT-PCR of the N-gene as described in [157]. Wastewater (250 mL) was concentrated to 1.25 mL, of which 0.2 mL was used for nucleic extraction. Six RT-PCR replicates were performed for each sample. SARS-CoV-2 RNA was considered detected if any of the RT-PCR replicates were positive. A

result of 'not detected' meant that SARS-CoV-2 RNA was either absent from the sample or at a level too low to be reliably reported.

RT-PCR data (quantification cycle values) were converted to genome copies per reaction using a standard curve and then to genome copies per litre of wastewater. Each sample is multiplied by the estimated volume of wastewater entering the wastewater treatment plant to estimate the total genome copies per day. These are summed over all sites sampled on that day and divided by the total volume of wastewater entering all sampled wastewater treatment plants to give daily estimates of genome copies per litre of wastewater. We also calculate the daily total population across sampled catchment sites.

The data and comprehensive details about collection and processing are available at https://github.com/ESR-NZ/covid_in_wastewater [349].

Reported cases

National daily reported cases of COVID-19 were obtained from the New Zealand Ministry of Health and are available at https://github.com/minhealthnz/nz-covid-data [348]. Reported case data exhibit a clear day-of-the-week effect, which we remove in pre-processing using a simple linear regression model. The log-transformed daily case count was regressed against the day of the week and the data were then divided by the exponential of the regression coefficient for each day of the week. Adjusted daily case counts were then scaled so the total case count remains consistent. We perform this day-of-the-week adjustment in the following consecutive time windows: before 1 January 2022, 1 January to 28 February 2022, 1 March to 30 June 2022, 1 July to 30 September 2022, 1 October to 31 December 2022, and 1 January 2023 to 31 March 2023. These are mostly three months in duration, except that the boundary between the second and third windows was selected to coincide with the change from PCR-only testing to widespread availability and use of RATs [352], as this impacted the weekly reporting pattern.

Algorithmically, this seasonal adjustment is performed when the data are loaded. As it results in non-integer daily case counts, we round the outputs to the nearest integer. The functions to do this are included in the "loadNZData.jl" function at https://github.com/nicsteyn2/NZWastewaterModelling/tree/main/src.

C.1.2 Model derivation and algorithms

Fixed-lag bootstrap filter

We have hidden states CAR_t, R_t, I_t, observed data W_t, C_t, and fixed parameter vector θ. We are interested in learning the joint posterior distribution $P(CAR_{1:T}, R_{1:T}, I_{1:T}, \theta | W_{1:T}, C_{1:T})$. The goal of Algorithm 6 (below) is to construct a set of particles $\left\{\left(CAR_k^{(i)}, R_k^{(i)}, I_k^{(i)}\right) : i = 1, ..., N\right\}$ that approximate this distribution.

For simplicity, let our hidden states be collectively denoted $X_t = (CAR_t, R_t, I_t)$ and our observed data $y_t = (W_t, C_t)$. We start with the filtering distribution at time t which we denote q_t (defined below). We decompose this filtering distribution into the following recursion:

$$\begin{aligned} q_t &= P_\theta(X_{1:t}|y_{1:t}) && \text{(C.1)} \\ &\propto P_\theta(y_t|X_{1:t}, y_{1:t-1})P_\theta(X_{1:t}|y_{1:t-1}) && \text{(C.2)} \\ &= P_\theta(y_t|X_{1:t})P_\theta(X_t|X_{1:t-1}, y_{1:t-1})P_\theta(X_{1:t-1}|y_{1:t-1}) && \text{(C.3)} \\ &= P_\theta(y_t|X_{1:t})P_\theta(X_t|X_{1:t-1})q_{t-1}, && \text{(C.4)} \end{aligned}$$

where $P_\theta(y_t|X_{1:t}) = P_\theta(W_t|I_{1:t-1})P_\theta(C_t|CAR_t, I_{1:t-1})$ is our joint observation distribution and $P_\theta(X_t|X_{1:t-1}) = P_\theta(CAR_t|CAR_{t-1})P_\theta(R_t|R_{t-1})P_\theta(I_t|I_{1:t-1})$ is our joint state-space transition distribution.

The decomposition makes two assumptions: (1) y_t is conditionally independent of $y_{1:t-1}$ given X_t, and (2) X_t is conditionally independent of $y_{1:t-1}$ given $X_{1:t-1}$. The definition of the observation and state-space transition distributions require further assumptions that are clear from their definition.

We define our importance sampling distribution π in a similar recursive fashion:

$$\pi(X_{1:t}|y_{1:t}) = \pi(X_t|X_{1:t-1}, y_{1:t})\pi(X_{1:t-1}|y_{1:t-1}), \tag{C.5}$$

requiring the non-restrictive assumption that our importance distribution on $X_{1:t-1}$ does not depend on y_t. This further allows us to define the importance sampling weights recursively:

$$\begin{aligned} w_k^{(i)} &\propto \frac{P_\theta(y_t|X_{1:t}^{(i)})P_\theta(X_t^{(i)}|X_{1:t-1}^{(i)})q_{t-1}^{(i)}}{\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, y_{1:t})\pi(X_{1:t-1}^{(i)}|y_{1:t-1})} && \text{(C.6)} \\ &= \frac{P_\theta(y_t|X_{1:t}^{(i)})P_\theta(X_t^{(i)}|X_{1:t-1}^{(i)})}{\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, y_{1:t})} w_{k-1}^{(i)}. && \text{(C.7)} \end{aligned}$$

Finally, we choose our importance sampling distribution $\pi(X_t^{(i)}|X_{1:t-1}^{(i)}, y_{1:t})$ to be the state-space transition distribution $P_\theta(X_t^{(i)}|X_{1:t-1}^{(i)})$, allowing these terms to be cancelled, and giving the recursion for the particle weights:

$$w_k^{(i)} = P_\theta(y_t|X_{1:t}^{(i)})w_{k-1}^{(i)}. \tag{C.8}$$

At each time step, we re-sample with replacement from our current particles according to $w_{k-1}^{(i)}$, thus the selected particles at time step t can be viewed as direct draws from $P_\theta(X_t|y_{1:t})$.

Additionally re-sampling the h most recent time steps re-weights past particles according to $P_\theta(y_t|X_{1:t}^{(i)})$, thus they can be viewed as samples from $P_\theta(X_t|y_{1:t+L})$. We note that in doing this resampling, we break the particle ancestry, so only have samples from the marginal distributions $P_\theta(X_t|y_{1:t+L}) \approx P_\theta(X_t|y_{1:T})$ and not the full joint distribution over state-trajectories $P_\theta(X_{1:t}|y_{1:t+L})$ (except for the final L states).

Conditional on a fixed value of θ, Algorithm 6 presents our fixed-lag bootstrap filter. This assumes a default wind-in period equal to h (30-days), although this can be changed as necessary.

Practical considerations for the bootstrap filter

Choosing the Fixed-Lag h

In an ideal world we would resample the entire state-history at every step, producing direct samples from our desired posterior distribution $P(X_t|y_{1:T})$. In fact, this also admits samples from the joint posterior distribution across all time steps $P(X_{1:T}|y_{1:T})$, which would allow us to sample individual trajectories. However, as T increases, the number of individual unique particles that remain in the earlier time steps (most obviously at $t = 1$) gets increasingly small. Eventually only a handful, or even a single, unique particle will remain for X_1, which provides a very poor approximation of $P(X_1|y_{1:T})$. This can be somewhat overcome by increasing the number of particles N_x, but given present computing power this quickly becomes impractical. Instead, we re-sample only the most recent h time steps, which means our particles at time step t are samples from $P(X_t|y_{1:t+h})$ whenever $t + h < T$. The value of h needs to be large enough that $P(X_t|y_{1:t+h}) \approx P(X_t|y_{1:T})$ while being small enough that we avoid particle degeneracy for reasonable values of N_x.

Calculating Cumulative Infections

In Figure 4 of the main text we present estimates of cumulative infections CI_t. If we were resampling entire particle histories at each time step (rather than fixed-lag resampling) we would be able to set $CI = \sum_{u=1}^{t} I_u^{(i)}$ for our i^{th} sample of cumulative infections at time t. However, our fixed-lag resampling breaks the state-histories, so this approach is invalid. Instead, we augment cumulative infections as an additional hidden state, at each time step setting:

$$CI_t^{(i)} = CI_{t-1}^{(i)} + I_t^{(i)}, \tag{C.9}$$

and resampling as usual. As the particle filter produces samples from $P(CI_t|y_{1:t+h})$, this method produces valid estimates of the pointwise cumulative infections CI_t.

Algorithm 6 Fixed-lag bootstrap filter

Input: Parameter vector θ, data $W_{1:T}$ and $C_{1:T}$
Sample $CAR_{-h:0}^{(i)}$, $R_{-h:0}^{(i)}$, and $I_{-h:0}^{(i)}$ from some initial state distribution
for $t = 1, ..., T$ **do**
  Sample $CAR_t^{(i)} \sim P_\theta(CAR_t|CAR_{t-1}^{(i)})$
  Sample $R_t^{(i)} \sim P_\theta(R_t|R_{t-1}^{(i)})$
  Sample $I_t^{(i)} \sim P_\theta(I_t|R_t^{(i)}, I_{t-L:t}^{(i)})$
  Set $w_t^{(i)} = P_\theta(W_t|I_{t-L:t}^{(i)}, CAR_t^{(i)})P_\theta(C_t|CAR_t^{(i)}, I_{t-L:t}^{(i)})$
  Sample with replacement indices $\{x_j\}_{j=1}^N$ from $i = 1, 2, ..., N$ with probability $w_t^{(i)}$
  Update $(CAR_{t-L:t}^{(j)}, R_{t-L:t}^{(j)}, I_{t-L:t}^{(j)}) \leftarrow (CAR_{t-L:t}^{(x_j)}, R_{t-L:t}^{(x_j)}, I_{t-L:t}^{(x_j)})$
end for
Return $(CAR_{1:T}^{(i)}, R_{1:T}^{(i)}, I_{1:T}^{(i)})$

Wind-in period

In practice, we wind-in using two steps: (1) hidden states are randomly allocated values for $t = -h$ to $t = 0$, and (2) the filter is run for $t = 1$ to $t = k$ for some k. We present results and calculate likelihoods for $t \geq k + 1$. The first step is necessary for there to be sufficient state-history to calculate the expected values of I_t, C_t, and W_t, which all involve convolutions of past infections. However, only a few of these randomly allocated trajectories will be plausible, leading to considerable uncertainty in the initial estimates of our hidden states. Thus we use data to filter particles as described above in the period $0 \leq t \leq k$ and start the estimation

window at $t = k$, so that all particle chains have plausible past trajectories at this time. In general k should be chosen to be greater than h.

This second wind-in period means that the estimation window only begins k days after the start of the period for which data are available. It is possible to run the algorithm without the second wind-in period, which may be necessary when data are limited. However, this leads to greater uncertainty about estimated states in the early part of the time period and introduces substantial additional variation in the estimates of the model log-likelihood (Supplementary Section C.1.2). In practice, we used $k = 50$, except for model runs starting on 1 January 2022 where we use $k = 31$ as the earliest available data were 1 December 2021.

Likelihood estimation

Thus far we have focused on estimating the value of the hidden states given some known parameter vector θ. Our particle filtering algorithm also admits a tidy, albeit noisy, method for estimating the likelihood function $L(\theta|y_{1:T}) \propto P(y_{1:T}|\theta)$ [300]. First note that:

$$P(y_{1:T}|\theta) = \prod_{t=1}^{T} P_\theta(y_t|y_{1:t-1}). \tag{C.10}$$

We can write each term in the product as:

$$P_\theta(y_t|y_{1:t-1}) = \int P_\theta(y_t|X_{t-L:t}) P_\theta(X_{t-L:t}|y_{1:t-1})\, dX_{t-L:t}. \tag{C.11}$$

Note $P_\theta(y_t|X^{(i)}_{t-L:t}) = w^{(i)}_t$. Furthermore, our projected (but non-filtered) particles $\{\tilde{X}^{(i)}_{t-L:t}\}_{i=1}^N$ at time t provide an approximation to $P_\theta(X_{t-L:t}|y_{1:t-1})$. Together this conveniently allows us to approximate this integral using:

$$P_\theta(y_t|y_{1:t-1}) \approx \frac{1}{N} \sum_{i=1}^{N} w^{(i)}_t. \tag{C.12}$$

Taking logarithms gives our estimator of the log-likelihood:

$$\hat{\ell}(\theta|y_{1:T}) = \sum_{t=1}^{T} \log\left(\frac{1}{N} \sum_{i=1}^{N} w^{(i)}_t \right). \tag{C.13}$$

As each term inside the outer sum is an approximation, the noise of this estimator grows with the length of data. In general this noise increases linearly with time, as does the time it takes to run a single filter, thus the computational requirements approximately scale with $O(T^2)$ [289].

Particle marginal Metropolis-Hastings

Particle marginal Metropolis-Hastings (PMMH; Algorithm 7) is an established algorithm designed to estimate the joint posterior distribution $P(X_{1:T}, \theta|y_{1:T})$ of the hidden states and fixed parameter vector given our data, although in practice we use this method to estimate the marginal posterior distribution $P(\theta|y_{1:T})$. The algorithm uses the following proposal density:

$$q((X'_{1:T}, \theta')|(X_{1:T}, \theta)) = q(\theta'|\theta)P_{\theta'}(X'_{1:T}|y_{1:T}), \tag{C.14}$$

where $X'_{1:T}$ are generated by running a particle filter at θ'. This gives an acceptance probability of:

$$a = \min\left(1, \frac{\hat{P}(y_{1:T}|\theta')P(\theta')q(\theta|\theta')}{\hat{P}(y_{1:T}|\theta)P(\theta)q(\theta'|\theta)}\right), \tag{C.15}$$

where $q(\theta'|\theta)$ is our proposal density on our parameters and $\hat{P}(y_{1:T}|\theta)$ is an estimate of the model evidence at parameter vector θ' - this is the exponential of $\hat{\ell}(\theta|y_{1:T})$ described in Supplementary Section C.1.2. The validity of using an estimate of the likelihood rather than an exact calculation is confirmed in [300], which is a key difference between PMMH and standard Metropolis-Hastings.

Algorithm 7 Particle marginal Metropolis-Hastings

Require: Prior distribution $P(\theta)$ on θ, proposal density $q(\theta'|\theta)$, and number of MCMC steps N
  Initialise $\theta_0 \sim P(\theta)$ and run Algorithm 6 to estimate $\hat{P}_{\theta_0} = P(\theta_0|y_{1:T})$
  for $i = 1, ..., N$ **do**
    Sample $\theta' \sim q(\cdot|\theta_{i-1})$
    Run Algorithm 6 to estimate $\hat{P}_{\theta'} = P(\theta'|y_{1:T})$
    Calculate acceptance probability $a_i = \min\left(1, \frac{\hat{P}(y_{1:T}|\theta')P(\theta')q(\theta_{i-1}|\theta')}{\hat{P}(y_{1:T}|\theta_{i-1})P(\theta_{i-1})q(\theta'|\theta_{i-1})}\right)$
    Let $\theta_i = \theta'$ with probability a_i, else let $\theta_i = \theta_{i-1}$
  end for
  Return $(CAR^{(i)}_{1:T}, R^{(i)}_{1:T}, I^{(i)}_{1:T})$

Prior distributions and proposal variances

The parameters σ_R, σ_{CAR}, k_c, and k_w may change as the epidemiological landscape changes so we fitted them in five distinct time periods: (1) 1 January 2022 – 31 March 2022, (2) 1 April 2022 – 30 June 2022, (3) 1 July 2022 – 30 September 2022, (4) 1 October 2022 – 31 December 2022, and (5) 1 January 2023 – 31 March 2023. Choosing the duration of these

windows requires balancing changing epidemiological dynamics (we expect these parameters to somewhat change over time) with using more data to obtain more precise estimates.

We use wide independent uniform distributions for our prior distributions on σ_R, k_c, and k_w. A wide uniform prior distribution can also be placed on σ_{CAR}, however, this results in a relatively high-valued posterior estimate for this parameter as the model can choose values of CAR_t that closely fit the fluctuations in reported case data. We want our estimates of CAR_t to reflect an underlying reporting rate, rather than the daily noise in reporting, so use a prior distribution on σ_{CAR} to ensure this. For time periods encompassing 1 April 2022 – 31 March 2023 we use a normal distribution with mean 0.006 and standard deviation 0.00204, truncated on $(0, \infty)$, which has a 95^{th} quantile of 0.01. For the first time period, encompassing 1 January 2022 to 31 March 2022, we use a higher-mean normal distribution with mean 0.024 and standard deviation 0.00816, truncated on $(0, \infty)$, which has a 95^{th} quantile of 0.04. The use of a higher-mean prior distribution for σ_{CAR} in this first period allows the model to fit to the rapid change in CAR_t that is thought to have occurred when RATs were rolled out in February 2022 [352]. Table C.1 reports our choices for prior distributions.

We use independent normal proposal densities for each parameter. The chosen standard deviations of the proposal densities are given in Table C.2. We outline how we chose these in Supplementary Section C.1.2.

Table C.1: Prior distributions on parameters. All normal distributions (represented by N) are truncated on $(0, \infty)$. The continuous uniform distribution is represented by U. The period starting 1 March 2022 continues until the end of considered period on 31 March 2023.

Period starting	σ_R	σ_{CAR}	k_c	k_w
1 Jan 2022	N(0.024, 0.00816)	U(0, 0.1)	U(0, 400)	U(0, 0.02)
1 Mar 2022	N(0.006, 0.00204)	U(0, 0.1)	U(0, 400)	U(0, 0.02)

Table C.2: The chosen standard deviation for each independent normal proposal distribution.

Period starting	σ_R	σ_{CAR}	k_c	$k_w(\times 10^{-7})$
1 Jan 2022	0.024	0.004	4.3	1.4
1 Apr 2022	0.018	0.0014	21	5.5
1 Jul 2022	0.010	0.0015	30	6.6
1 Oct 2022	0.0073	0.001	22	7.4
1 Jan 2023	0.0089	0.0015	22	8.9

Practical considerations for particle marginal Metropolis-Hastings

In situations where the proposed particle is rejected, it is not necessary to re-estimate $\hat{P}_{\theta_{i-1}}$. One can typically let $\hat{P}_{\theta_i} = \hat{P}_{\theta_{i-1}}$. In some situations, particularly when the variance of the likelihood estimator is large, the Markov chain can get stuck on values of θ' where the estimate of $P_{\theta'}$ was unusually high, resulting in slower convergence. To avoid this we re-estimate $P_{\theta_{i-1}}$ if n consecutive proposals have been rejected. Choosing n requires balancing the computational cost of running additional particle filters with the cost of slower mixing chains. We use $n = 5$ in this work.

Theoretically, this algorithm works irrespective of the number of particles N_x used in each filter, however, this does have a substantial impact on the performance of the algorithm. A general heuristic is that N_x should be chosen such that the standard deviation of $\hat{\ell}(\theta|y_{1:T})$ is around 1.2-1.3 [289]. This standard deviation is a function of θ itself (generally speaking estimates of ℓ at more likely values of θ have lower standard deviations) so choosing the ideal N_x is not a simple task. We use $N_x = 10^5$ particles per filter when fitting to three-month time periods.

PMMH is a computationally expensive algorithm. Our results rely on 8 PMMH chains for each of the five time periods considered. Whilst running this on high-performance computing services can allow us to utilise the 40 cores required to run each chain simultaneously, it can still take multiple days to generate sufficient samples. We find that, with $N_x = 10^5$, it takes us approximately 20 hours to generate 2000 samples (although this can be faster if more modern central processing units are used). As seen in Supplementary Section C.2.5, this can be enough to meet certain convergence criteria even if this is fewer samples than most MCMC algorithms target. Practically we expect this to make little difference to our posterior estimates of θ, and even less of a difference to our posterior estimates of $X_{1:T}$. This final point can be seen by noting that the marginal posterior estimates of $X_{1:T}$ are very similar to the posterior estimates of $X_{1:T}$ conditional on any plausible value of θ - the majority of the uncertainty comes from the relationship between the hidden states and the observed data, rather than the hyperparameters that characterise this relationship.

Despite generally using wide uniform prior distributions, we want to ensure our chains start at plausible values of θ, otherwise considerable computation time must be spent on a wind-

in period. Therefore, we first computed approximate bivariate heatmaps of the estimated log-likelihood on a coarse parameter mesh (Supplementary Section C.2.2). We then initialised chains in the part of parameter space with relatively high likelihood values. As well as reducing convergence times, this technique also provides some reassurance that the PMMH algorithm is not missing any hidden modes, and that the posterior distribution found is similar to empirical results.

We ran the PMMH algorithm twice for each three-month block. First we did a training run, with seemingly plausible values for the parameter proposal variances, to provide a crude estimate of the posterior variance. Then a second, final, run was performed using the heuristic proposal variance of $2.38\sigma_{posterior}/n_{dim}$, where $\sigma_{posterior}$ is the estimated posterior standard deviation of the chosen parameter from the first run, and $n_{dim} = 4$ is the dimensionality of the parameter space. This heuristic could be replaced with adaptive MCMC methods - these are well known but slightly more complicated to implement from scratch.

Posterior distribution on hidden states

One way of estimating the marginal posterior distribution $P(X_{1:T}|y_{1:T})$ is to store one (or more) trajectories from each PMMH step. The resulting set of particle trajectories are samples from the pointwise marginal posterior distribution $P(X_t|Y_{1:T})$. However, as we fit the parameters in three-month windows, the PMMH method only outputs trajectories in three-month blocks, which cannot be easily joined together.

Instead, we first run Algorithm 7 to generate a set of fixed parameter values $\{\theta_i\}_{i=1}^{N_c} \sim P(\theta|y_{1:T})$. We then sample from $P(X_{1:T}|y_{1:T})$ by iteratively uniformly sampling θ^* from $\{\theta_i\}_{i=1}^{N_c}$, running Algorithm 6 with N_x particles at $\theta = \theta^*$, and keeping N_s trajectories (where $N_s \leq N_x$) from the output. Repeating this N_c times (once for each parameter sample) gives a set of $N_s N_c$ particle trajectories that approximate the pointwise posterior distribution $P(X_t|y_{1:t+h}) \approx P(X_t|y_{1:T})$. Note that each sample from $\{\theta_i\}_{i=1}^{N_c}$ consists of five independent sets of values for the inferred parameters (σ_{CAR}, σ_R, k_c and k_w), one for each of the five three-month periods. When we run Algorithm (7) for the whole 15-month period, we assume that the values of these four parameters change instantaneously from one three-month block to the next.

Typically, we want to run this with sufficient unique draws from $P(\theta|y_{1:T})$ (say $N_c \geq 100$) to appropriately account for uncertainty in θ. The number of samples retained from each iteration should be chosen so our overall number of trajectories is sufficiently large - we choose $N_c N_s = 2 \times 10^6$. Finally, as we are less concerned by minor degeneracy in individual particle filters, the number of particles used in each filter N_x can be smaller than if we were just running the filter once. For our results, $N_x = 10^5$ worked well, but this needs to be tailored to individual purposes.

When presenting results we calculate the mean of the samples as the central estimate and use the 2.5^{th} and 97.5^{th} quantiles to represent our 95% credible intervals (CrI).

Pre-determined parameters

In addition to the estimated parameters, there are others that we fix. These are the generation time distribution g_u, infection-to-reporting distribution L_u, the infection-to-shedding distribution ω_u, and the average total genome copies per infection α. We also pre-specify the fixed-lag resampling window $h = 30$.

The generation time is assumed to be a discretised Gamma random variable with mean 3.3 days and standard deviation 1.3 days [352, 374, 375, 415]. The infection-to-reporting and infection-to-shedding distributions are calculated as convolutions of an incubation period (infection-to-onset) distribution (Weibull with mean 2.9 days and standard deviation 2.0 days [373]) and an onset-to-reporting distribution (estimated from New Zealand data, mean 1.8 days and standard deviation 1.8 days) or an onset-to-shedding distribution (mean 0.7 and standard deviation 2.6 [157]). The Gamma and Weibull distributions were discretised by taking their value at integer times and normalising. All of these distributions are presented in Supplementary Figure C.1.

Estimating curvewise extrema

Our primary methods produce pointwise estimates of the hidden states at each time step (e.g., mean and quantiles of the samples at each fixed value of t). As the timing of peaks and troughs can be quite variable between particles, using pointwise statistics to quantify the heights and timings of peaks or troughs (e.g., local maxima in the median value across particles) can be misleading [416]. For this purpose, it is important to consider these on a curvewise basis (e.g.,

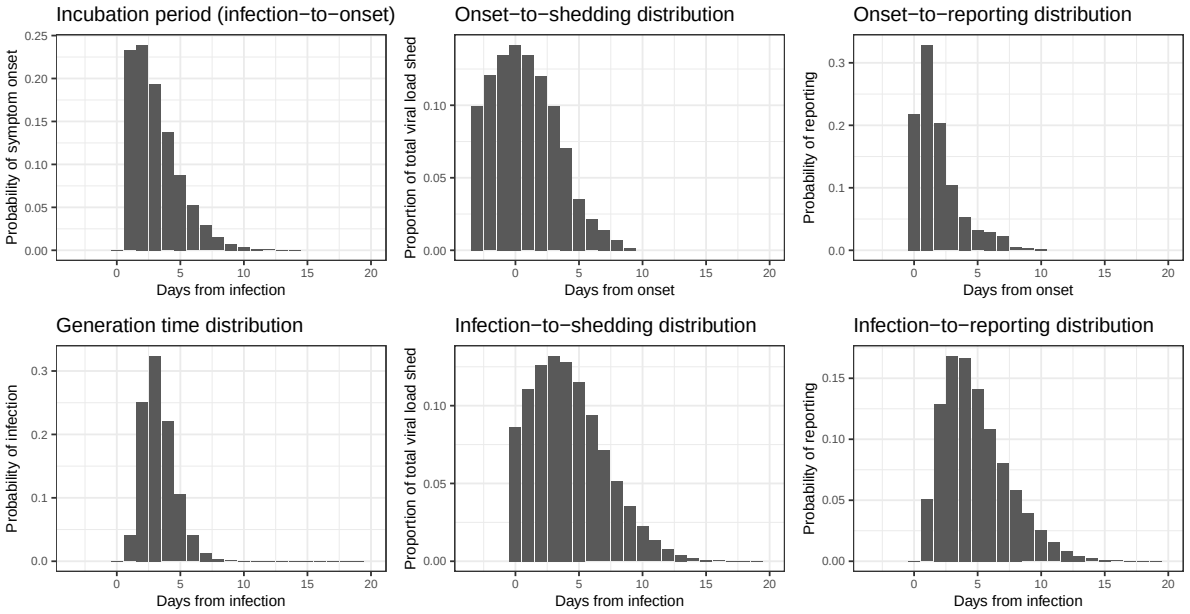

Figure C.1: Delay distributions. Distributions reported in the first row were used as inputs to create the infection-to-shedding and infection-to-reporting distributions reported in the second row.

median of the local maxima across particles). This requires samples from the joint posterior distribution $P(X_{s:t}|y_{1:T})$ over some fixed window $s:t$.

We achieve this by increasing our resampling length h and halting the algorithm when the period of interest is contained in $(T-h, T-30)$ where T represents the final stopping time. Limiting the lower window to $t \geq T-h$ ensures that complete trajectories are faithfully resampled over the time period of interest. Limiting the upper window to $t \leq T-30$ ensures we are using appropriately smoothed samples that are sufficiently informed by data.

C.2 Supplementary results

C.2.1 Synthetic verification of hidden state estimates

Before analysing real-world data, we performed synthetic tests to verify our model. We imposed a prescribed time-varying reproduction number and CAR and ran a forward simulation of our model to calculate the median number of infections, cases and wastewater data for a fixed value of α. We then used the simulated case and wastewater data as inputs to the particle filter to estimate R_t and CAR_t. To investigate how the addition of wastewater data affected model performance, we ran the particle filter with three different sets of inputs: only case data, only

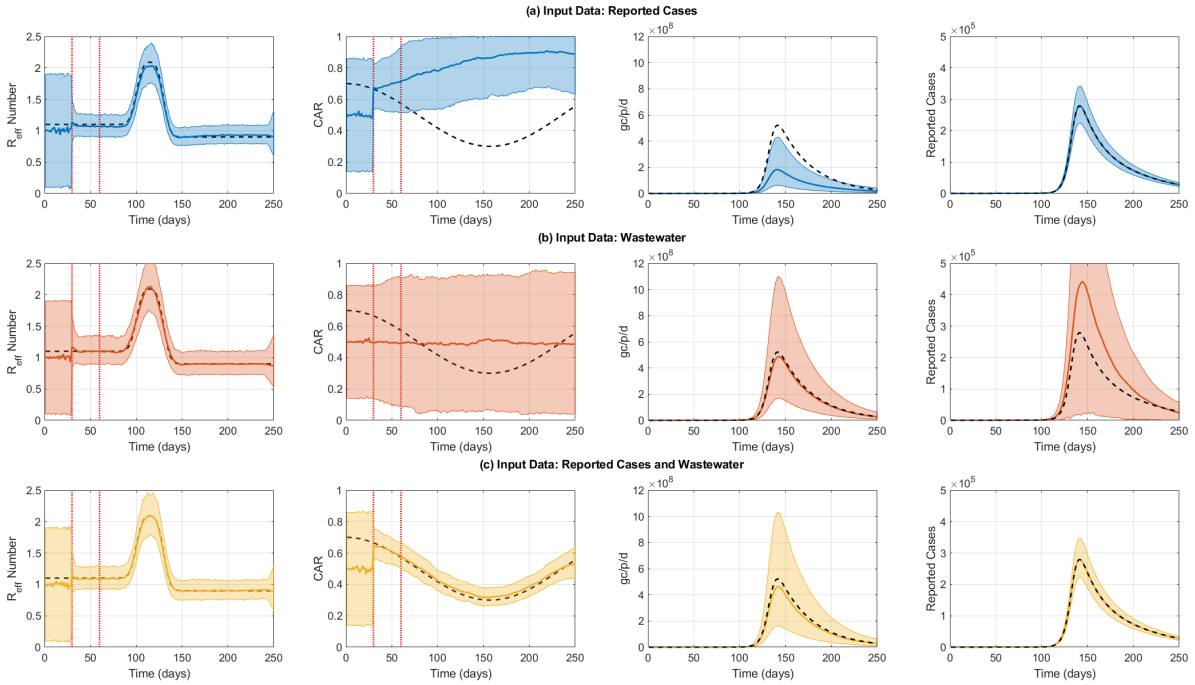

Figure C.2: Synthetic results. (left) Instantaneous reproduction number, (centre-left) case ascertainment rate, (centre-right) wastewater data in genome copies per person per day (gc/p/d) and (right) reported cases. Results are shown when using (a) only reported cases as input data for the particle filter, (b) only wastewater data, and (c) both reported cases and wastewater data. Solid lines present central estimates. Shaded regions show 95% CrIs on the value of the hidden states (left and centre-left columns) and 95% CrIs on the prediction distribution for wastewater data and reported cases (centre-right and right columns). Black dashed lines indicate the synthetic data. Vertical red lines in hidden state plots (left and centre-left columns) indicate the end of the two wind-in periods.

wastewater data, and both case and wastewater data. The results are shown in Figure C.2. The parameters used were $\sigma_R = 0.1$, $\sigma_{CAR} = 0.02$, $k_c = 100$, and $k_w = 1 \times 10^{-6}$ along with $\alpha = 3 \times 10^9$.

All three data combinations resulted in a reasonable estimate of R_t (Supplementary Figure C.2). The model error was smallest when using both reported cases and wastewater data as input (root mean square error between the true solution and the median of the particle filter output was 3.4 and 0.7 times larger when only using reported cases or wastewater data, respectively). Previous work has estimated R_t from reported case data [139]. The results presented in Supplementary Figure C.2 demonstrate that R_t can also be estimated from wastewater data independently from case data and that the most accurate result is achieved by combining reported case information with wastewater data.

For CAR, there were substantial differences between the three different sets of input data (Figure C.2). When using only reported cases or wastewater data separately, the model did not have sufficient information to inform estimates of CAR. As a consequence, estimates were either inaccurate or did not capture the temporal trend and had very wide credible intervals. When reported case information was combined with wastewater data, there was good agreement between the estimated CAR and the true solution, with a relatively narrow credible interval. This illustrates the value of combining wastewater data with reported case information to obtain reliable estimates of changes in CAR over time.

C.2.2 Visualising log-likelihood estimates

As discussed in Supplementary Section C.1.2, we simulated the model likelihood on a coarse grid of parameter values to get a preliminary estimate of the plausible range of parameter values. We present examples of two outputs in Figure C.3. The left-plot shows log-likelihood estimates for σ_R for the third estimation window (1 July 2022 to 30 September 2022) while the right-plot shows a bivariate heatmap of log-likelihood estimate for σ_{CAR} and k_c. Code to reproduce these figures is provided in the *usefulscripts* subfolder at https://github.com/nicsteyn2/NZWastewaterModelling.

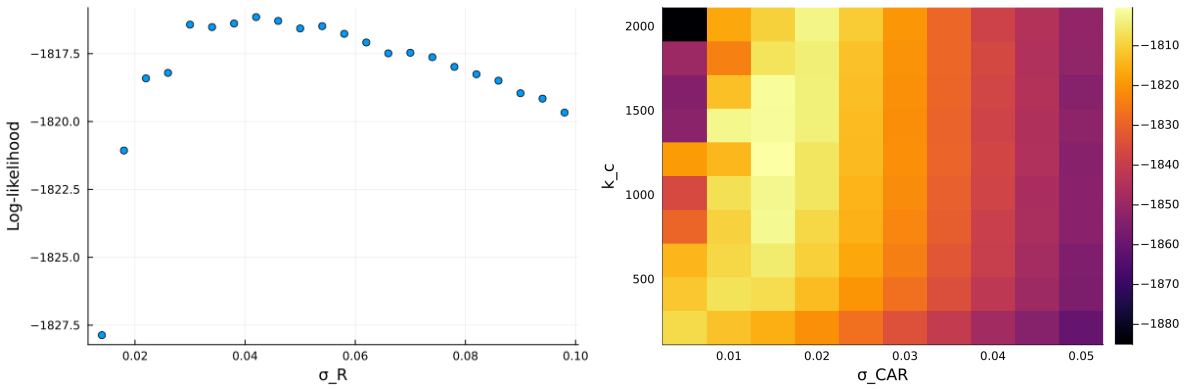

Figure C.3: Log-likelihood estimates for model parameters. Log-likelihood estimates for various values of σ_R (left) and various combinations of values of σ_{CAR} and k_c (right).

C.2.3 Sensitivity to delay distributions

The model relies upon three distributions describing the delay from infection to reporting, shedding detected viral genome copies, and infecting other people. We test the effect of shifting these distributions by one day backward or forward. We do this by appending a zero at the

start (to shift times backward), or by removing the first entry in d (to shift times forward), where d_i is the probability vector specifying the likelihood the delay takes i days. This has the effect of shifting the mean of the distributions by approximately one day in either direction.

Supplementary Figures C.4 to C.6 show the effect of these shifts to be minimal, with the exception of the shedding distribution, where shifts can substantially impact estimates of the absolute case ascertainment rate, even though the relative case ascertainment rate is similar despite the shifts.

Note, due to computational limitations we do not re-fit fixed parameters for each shift, although the effect of doing this is expected to be negligible.

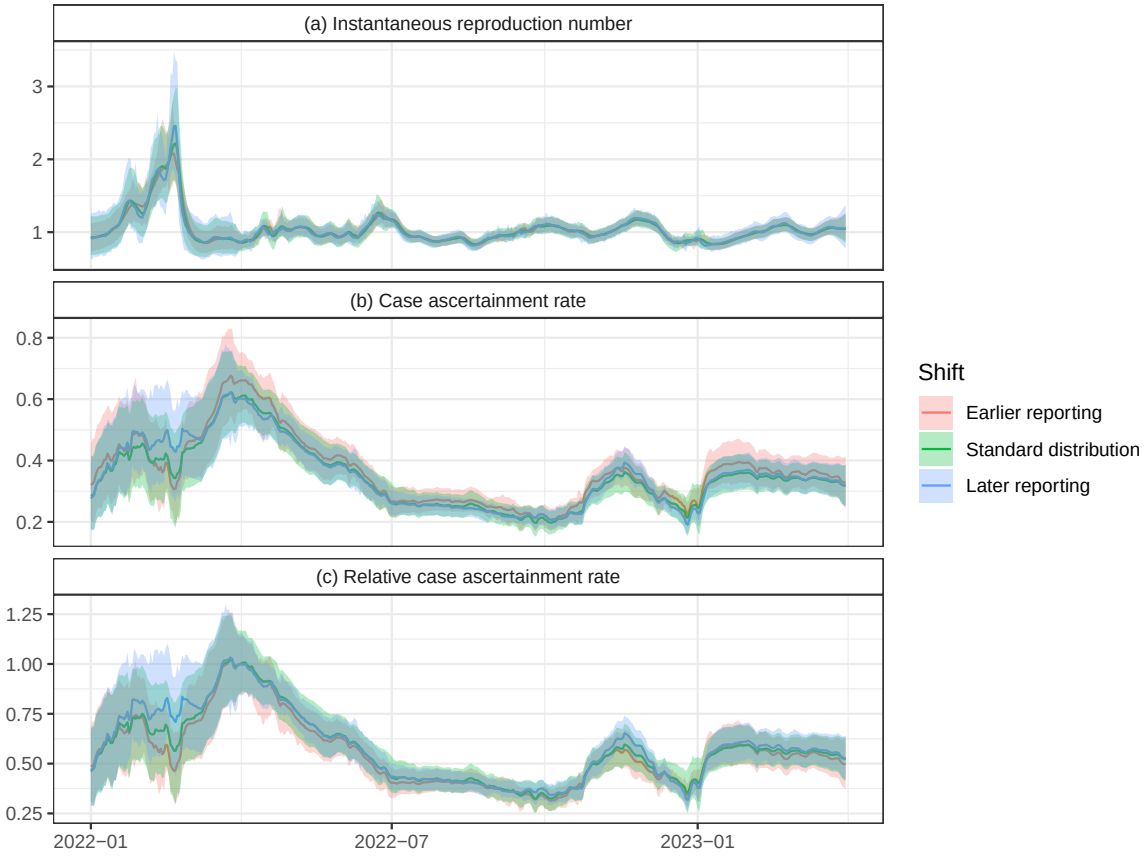

Figure C.4: The effect of shifting the reporting time distribution by one day forward or backward. The effect of shifting the reporting time distribution by one day forward or backward on (a) instantaneous reproduction number, (b) case ascertainment rate, and (c) relative case ascertainment rate.

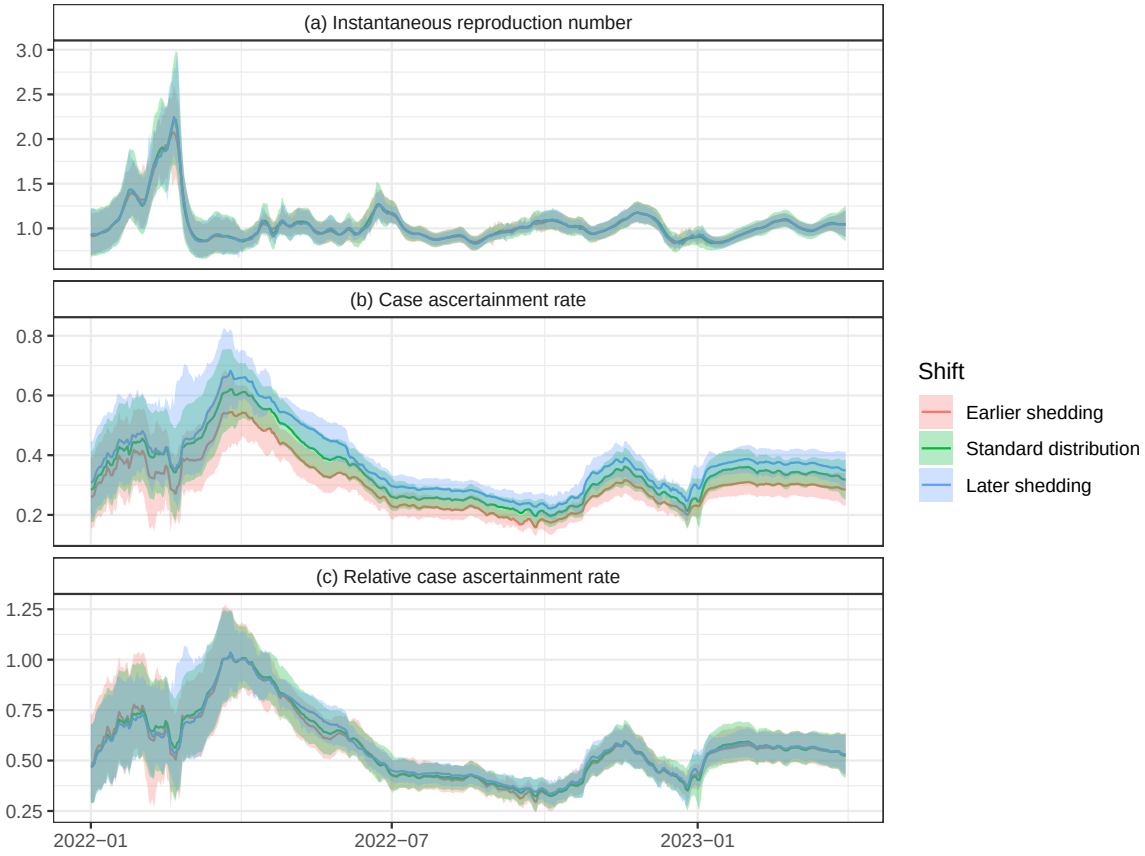

Figure C.5: The effect of shifting the shedding time distribution by one day forward or backward. The effect of shifting the shedding time distribution by one day forward or backward on (a) instantaneous reproduction number, (b) case ascertainment rate, and (c) relative case ascertainment rate.

C.2.4 Fitting to reported cases and wastewater data separately

The observation distribution is the product of the case observation distribution and the wastewater observation distribution. We can instead run the model with only one of these observation distributions. As there is no longer any information about the case ascertainment rate, we fix this at an arbitrary value of $CAR_t = 0.5$ with no fluctuation over time ($\sigma_{CAR} = 0$). We use PMMH to again fit the relevant parameters to each model (cases-only and wastewater-only). We report the posterior mean and 95% credible intervals, alongside estimates from the full model, in Supplementary Table C.3.

In general, estimates of σ_R were greater when fit to a single source of data. For reported cases, this was accompanied by slightly smaller estimates of k_c, suggesting that the assumption of a fixed CAR caused the attribution of some uncertainty to shift from the observation process to

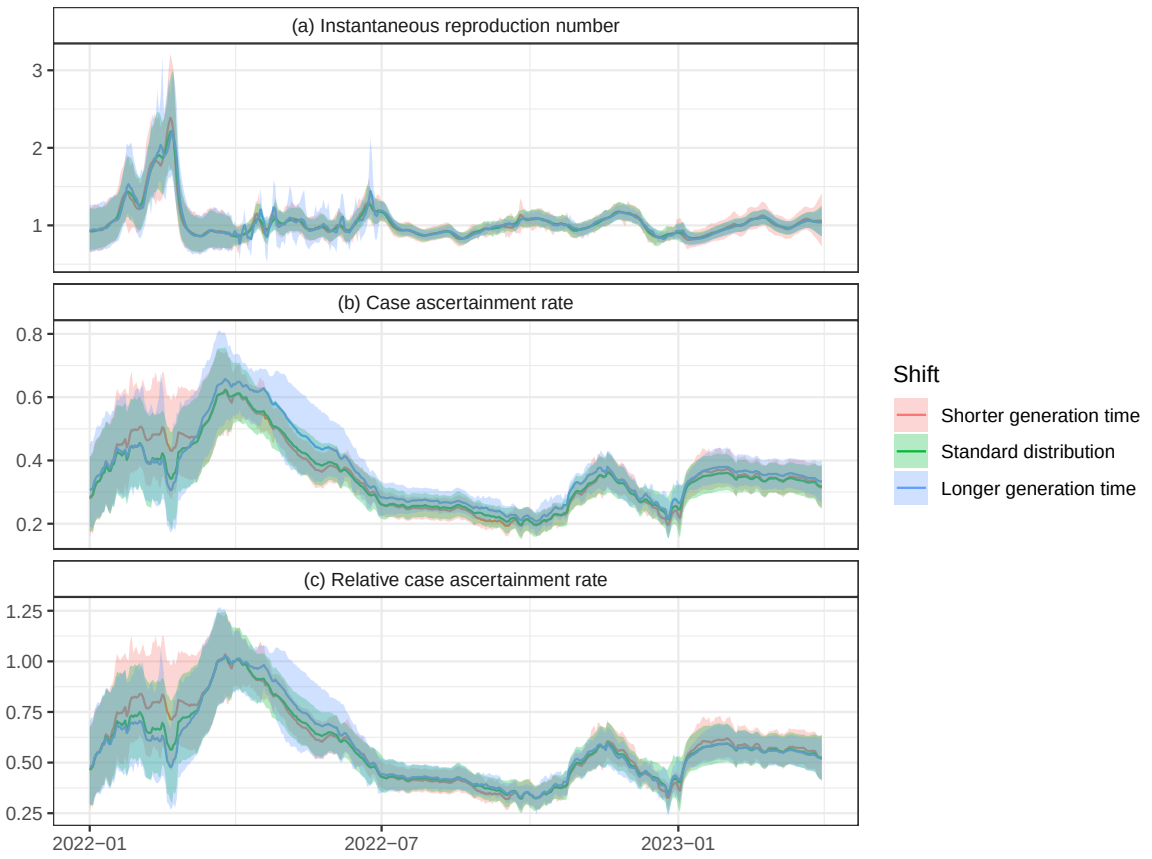

Figure C.6: The effect of shifting the generation time distribution by one day forward or backward. The effect of shifting the generation time distribution by one day forward or backward on (a) instantaneous reproduction number, (b) case ascertainment rate, and (c) relative case ascertainment rate.

fluctuations in the hidden state R_t.

Supplementary Figure C.7 shows reproduction number estimates from these separate models. The credible intervals resulting from fitting the model to wastewater data alone are substantially wider than when fitting to cases and wastewater. The reproduction number estimates from fitting the model to cases alone exhibit higher short-term fluctuations at times, which could be suggestive of overfitting.

Table C.3: Central estimates and 95% CrIs for estimated model parameters in each time period for the model when fit to both series, as well as separately to cases and wastewater data. Dates in the 'Period' column are the start date for the three-month period. All outputs presented to 2 s.f. Higher values of σ_R and σ_{CAR} suggest R_t and CAR_t vary faster. Higher values of k_c and k_w indicate a lower variance in the corresponding observation distribution. Note a different prior distribution was used for σ_{CAR} in the first period, which may also impact estimates of other parameters in this period.

Period	Model	σ_R	σ_{CAR}	k_c	$k_w(\times 10^{-6})$
1 Jan 2022	Joint	0.12 (0.069, 0.21)	0.03 (0.017, 0.043)	31 (20, 49)	1.5 (1.1, 2)
	Cases	0.14 (0.077, 0.23)	0 (assumed)	27 (18, 38)	-
	Wastewater	0.14 (0.070, 0.25)	0 (assumed)	-	1.5 (1.1, 2.1)
1 Apr 2022	Joint	0.069 (0.041, 0.12)	0.0099 (0.0053, 0.014)	170 (100, 250)	4.8 (3.2, 6.8)
	Cases	0.056 (0.030, 0.11)	0 (assumed)	150 (100, 210)	-
	Wastewater	0.057 (0.036, 0.10)	0 (assumed)	-	5.2 (3.6, 7.3)
1 Jul 2022	Joint	0.037 (0.020, 0.066)	0.0063 (0.0018, 0.01)	330 (220, 400)	4.8 (3.3, 6.5)
	Cases	0.053 (0.022, 0.13)	0 (assumed)	240 (140, 360)	-
	Wastewater	0.060 (0.024, 0.14)	0 (assumed)	-	5.2 (3.7, 7.1)
1 Oct 2022	Joint	0.038 (0.020, 0.068)	0.011 (0.0073, 0.014)	170 (110, 270)	7.2 (4.7, 10)
	Cases	0.048 (0.020, 0.098)	0 (assumed)	120 (64, 140)	-
	Wastewater	0.063 (0.032, 0.12)	0 (assumed)	-	8.5 (5.7, 12)
1 Jan 2023	Joint	0.038 (0.018, 0.073)	0.0093 (0.0041, 0.015)	150 (84, 330)	6.8 (4.4, 10)
	Cases	0.056 (0.023, 0.11)	0 (assumed)	140 (71, 270)	-
	Wastewater	0.059 (0.027, 0.12)	0 (assumed)	-	7.1 (4.9, 10)

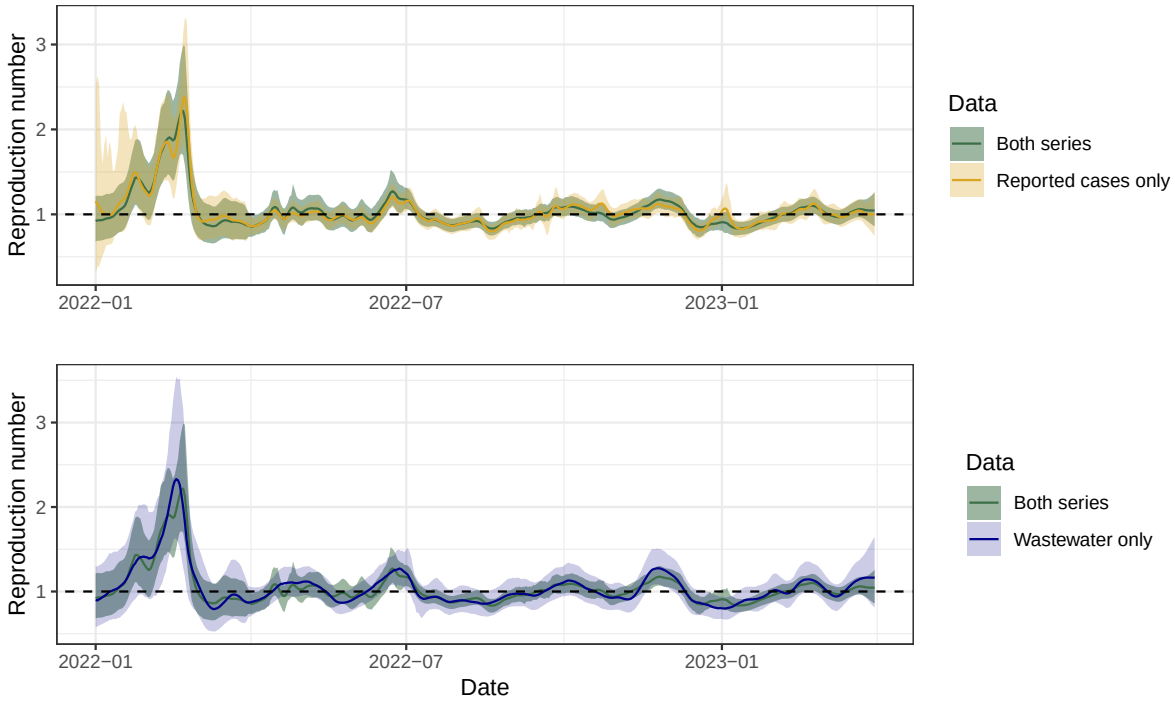

Figure C.7: Sensitivity of estimated reproduction number to input data. The estimated reproduction number when the model is fit to reported cases only (gold, upper panel) and wastewater data only (blue, lower panel). The original method fit to both series is included for comparison (green, both panels).

C.2.5 Particle marginal Metropolis-Hastings outputs

This section presents the trace plots, pairwise scatterplots of samples, pairwise correlations of samples, univariate kernel density plots, and the Gelman-Rubin diagnostic for each of the four parameters over the five three-month time periods in which they are estimated. Each of the eight chains were run for 5,000 iterations, with the initial 100 samples dropped as a wind-in period (although as discussed in Supplementary Section C.1.2, our initial choices of parameter values mean this is somewhat superfluous).

The Gelman-Rubin diagnostic tests for convergence by comparing intra-chain variance with inter-chain variance. We use the *coda* package in R [417] to calculate a point estimate and 95% upper confidence bound on the Gelman-Rubin diagnostic. It is generally accepted that values less than 1.1 imply convergence, although some use the more relaxed cutoff of 1.2. These are reported in Supplementary Table C.4.

Table C.4: Gelman-Rubin diagnostics and total sample sizes (for all eight chains) for each parameter and time period.

Period starting	σ_R	σ_{CAR}	k_c	k_w	Sample size
1 Jan 2022	1.01 (1.01)	1.01 (1.02)	1.01 (1.01)	1.00 (1.02)	39,200
1 Apr 2022	1.00 (1.01)	1.01 (1.06)	1.01 (1.02)	1.01 (1.02)	39,200
1 Jul 2022	1.01 (1.02)	1.01 (1.02)	1.01 (1.03)	1.01 (1.01)	39,200
1 Oct 2022	1.01 (1.02)	1.01 (1.04)	1.02 (1.04)	1.00 (1.01)	39,200
1 Jan 2023	1.02 (1.04)	1.02 (1.05)	1.04 (1.09)	1.01 (1.03)	39,200

To reduce the filesize of this thesis, the MCMC trace plots, scatter plots, kernel densities, and pairwise correlations are provided in the online supplementary material only. They can be accessed at https://doi.org/10.1038/s43856-024-00570-3.

Appendix D

Supplementary material for Chapter 8: Pandemic-risk-related behaviour change in England from June 2020 to March 2022: REACT-1 study among over 2 million people

D.1 Data

D.1.1 Overview of survey questions

We primarily focused on four survey questions:

1. Are you shielding and/or taking specific precautions because you are concerned that you/your child will become severely ill with COVID-19?
2. Did you/your child leave home for any reason in the last 7 days?
3. Not including members of your household, how many different people did you have contact with yesterday? By contact, we mean: any direct skin-to-skin physical contact (e.g., kiss/embrace/handshake), being less than 2 metres from another person for over 5 minutes.
4. Do you/does your child mainly wear any kind of face covering or mask when you/they are outside your/their home, because of COVID-19?

We also considered responses to additional questions:

a) In the last 7 days for what reasons have you left home?

b) Do you consider yourself/your child to be at risk of severe illness for COVID-19, for example due to an underlying health condition?

c) Do you think you have/your child has or have/has had COVID-19?

d) When were you told/did you think you/your child first had COVID-19?

e) Have you/has your child had a coronavirus vaccine?

f) If yes, how many doses (injections) have you/has your child had so far?

Variations in question wording and other nuances in interpretation are described in this section. These questions were answered by the survey respondent (if over the age of 18), by the respondent's parent (if aged between 5 and 12), and either by the parent or by the respondent themselves (if aged between 13 and 17).

Are you shielding and/or taking specific precautions because you are concerned that you/your child will become severely ill with COVID-19?

Variations of this question were asked from June 2020 to March 2022 (study rounds 2 to 19). Between June 2020 and November 2020 (study rounds 2 to 7) the question specifically asked about shielding, then between January 2021 and December 2021 (study rounds 8 to 16) the question was expanded to include "taking specific precautions", before the reference to shielding was removed between January 2022 and March 2022 (study rounds 17 to 19). Supplementary Table D.1 summarises these changes.

Table D.1: Changes to the wording of the shielding-related question in the REACT-1 study.

Period	Question
Jun 2020 – Nov 2020 (Rounds 2-to-7)	Are you **shielding** because you are concerned that you/your child will become severely ill with COVID-19?
Jan 2021 – Dec 2021 (Rounds 8-to-16)	Are you **shielding or taking specific precautions** because you are concerned that...
Jan 2022 – Mar 2022 (Rounds 17-to-19)	Are you **taking specific precautions** because you are concerned that...

Did you/your child leave home for any reason in the last 7 days, that is since <DATE/MONTH>?

This question was first asked in June/July 2020 (study round 2). Between July 2020 and September 2021 (study rounds 3-to-14, except round 7) additional wording was added to clarify that short trips such as shopping or exercise should be included. This question was not asked in the November 2020 survey (study round 7). Between October 2021 and March 2022 (study rounds 15-to-19) this question was replaced by a list of reasons for leaving the home, with one option being "I did not leave the home". Supplementary Table D.2 summarises these changes.

Table D.2: Changes to the wording of the leaving-home question in the REACT-1 study.

Period	Question
Jun 2020 – Jul 2020 (Round 2)	"Did you/your child leave home for any reason in the last 7 days, that is since <DATE/MONTH>?"
Jul 2020 – Nov 2020 (Rounds 3-to-6)	"Did you/your child leave home for any reason in the last 7 days, that is since <DATE/MONTH>? Please include even short trips outside the home, e.g., for shopping, exercise."
Nov 2020 – Dec 2020 (Round 7)	Question not asked
Jan 2021 – Sep 2021 (Rounds 8-to-14)	"Did you/your child leave home for any reason in the last 7 days, that is since <DATE/MONTH>? Please include even short trips outside the home, e.g., for shopping, exercise."
Oct 2021 – Mar 2022 (Rounds 15-to-19)	"In the last 7 days, for what reasons have you left home?"*

* There is an option to select *"I haven't left home"*.

Not including members of your household, how many different people did you have contact with yesterday? If you/they had contact with a person more than one time, please count them only once. By contact we mean: any direct skin-to-skin physical contact (e.g., kiss/embrace/handshake), being less than 2 metres from another person for over 5 minutes.

This question was asked between June 2020 and March 2022 (study rounds 2 to 19). Between November 2020 and April 2021 (study rounds 7 to 11), the question wording appeared differently for those that reported being in a support or childcare bubble, beginning with: "Not including members of your household or people in your support and childcare bubbles…".

Furthermore, between November 2020 and March 2022 (study rounds 7 to 19), those aged 17 or less were asked not to include contacts at school.

Respondents were asked to enter 0 if they had no contacts yesterday, and to give their best guess if they were not sure. While we focused on the proportion of people that reported having no contacts in the main paper, we also presented the mean number of self-reported contacts in Supplementary Section D.2.2.

Do you/does your child mainly wear any kind of face covering or mask when you/they are outside your/their home, because of COVID-19?

This question was asked between October 2020 and March 2022 (study rounds 6 to 19). There are five options: (1) No, (2) Yes, at work/school only, (3) Yes, in other situations only (including public transport and shops), (4) Yes, usually both at work/school and in other situations, and (5) My/their face is already covered for other reasons (e.g., religious or cultural reasons). We calculate the proportion of people that responded with any yes (options 2, 3, and 4) given that the individual does not already cover their face for other reasons (option 5).

In the last 7 days for what reasons have you left home? Please select all that apply.

Between June 2020 and September 2021 (study rounds 2 to 14) participants were only asked this question if they responded yes when asked if they had left the home in the past 7 days. Between October 2021 and March 2022 (study rounds 15-to-19) all participants were asked this question. The available options varied by round (Supplementary Table D.3).

Table D.3: Available options for reasons for leaving home by round.

Reason	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
For work	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
To volunteer	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
For medical or dental appointments	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
To care for someone else*	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
To socialise with people in a public place	Y	Y	Y	Y	Y		Y											
To socialise with people in a personal place	Y	Y	Y	Y	Y		Y											
To meet with someone outside								Y	Y	Y								
To meet with people in your childcare bubble								Y	Y	Y								
To meet with people in your support bubble									Y	Y								
To socialise with people outside											Y	Y	Y	Y	Y	Y	Y	Y
To socialise with people inside											Y	Y	Y	Y	Y	Y	Y	Y
For outdoor exercise	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
To go shopping	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
For errands	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
To take a child to school or childcare								Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
To get a vaccination									Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
To walk a dog/other pet care									Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
To go to school/college/university										Y	Y	Y	Y	Y	Y	Y	Y	Y
To go on holiday (in the UK or abroad)											Y	Y	Y	Y	Y	Y	Y	Y
Other reasons	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
I have not left home in the past 7 days**														Y	Y	Y	Y	Y

* The wording of these questions sometimes changes. ** This option replaced the original question about whether or not the participant had left home in the last 7 days; from round 15 onwards participants were simply presented with this list.

Do you consider yourself/your child to be at risk of severe illness for COVID-19, for example due to an underlying health condition?

A variation of this question was asked throughout the study. Table D.4 shows changes to question wording and the available answers.

Do you think you have/your child has or have/has had COVID-19?

This question was asked in all rounds of the study. The wording did not change. Four answers were available:

1. Yes, confirmed by a positive test
2. Yes, suspected by a doctor but not tested
3. Yes, my own suspicions
4. No

If the individual responded with (1), (2), or (3), they were then asked: "*When were you told/did you think you/your child first had COVID-19? If you are not sure, please give an estimate*".

Table D.4: Question wording and available answers for the question relating to whether an individual believes they are at risk of severe illness for COVID-19.

Period	Question	Available answers
May 2020 (Round 1)	"Have you/has your child been contacted by letter or text message to say you/they are at severe risk from COVID-19 due to an underlying health condition and should be shielding?"	1. Yes 2. No
Jun 2020 – Nov 2020 (Rounds 2-to-7)	"Do you consider yourself/your child to be at risk for severe illness for COVID-19, for example due to an underlying health condition?"	1. Yes 2. No
Jan 2021 – Dec 2021 (Rounds 8-to-16)	"Do you consider yourself/your child to be at risk for severe illness for COVID-19, for example due to an underlying health condition **or because you/they are clinically extremely vulnerable?**"	1. Yes 2. No 3. Don't know
Jan 2022 – Mar 2022 (Rounds 17-to-19)	"Do you consider yourself/your child to be at risk for severe illness for COVID-19, for example due to an underlying health condition?"	1. Yes 2. No 3. Don't know

We used responses to this question to derive three variables:

- **Reports a current suspected/confirmed infection**: when the individual responds with option (1), (2), or (3), and their estimated date of first infection is within 13 days prior to completing the questionnaire.
- **Reports a previous suspected/confirmed infection**: when the individual responds with option (1), (2), or (3), and their estimated date of first infection is at least 14 days prior to completing the questionnaire. We also include those who do not report an estimated date of first infection in this group.
- **Reports a previous confirmed infection**: when the individual responds with option (1) and their estimated date of first infection is at least 14 days prior to completing the questionnaire. We also include those who do not report an estimated date of first infection in this group.

Vaccination-related questions

In round 8, all vaccination-related questions were asked during the follow-up survey (the survey featuring behavioural questions, typically completed when self-administering the swab, as

opposed to the short registration survey). These questions were:

1. Have you/has your child had a coronavirus vaccine? (Yes, No, Don't know).
2. If yes, how many doses (injections) have you/has your child had so far? (One, Two, More than two).

In round 9 onwards, the above questions were asked during the registration survey (the survey asked when signing up to participate in the study). The follow-up survey then asked:

1. We asked you this when you registered to receive a swab test kit but we would just like to double check, have you/has your child ever had a coronavirus vaccine? (Yes, No, Don't know).
2. If yes, can we just check, have you/has your child had a vaccination since you/they registered to take part in this study? (Yes, No, Don't know).
3. If yes, how many doses (injections) have you/has your child had so far? (One, Two, More than two).

Between rounds 8 and 13, these questions were asked of all respondents (or parents of respondents) aged 16 and over. From round 14 onwards these questions were asked of all respondents (or parents of respondents) aged 12 and over. From rounds 15 onwards, questions asking about the number of doses received also included the options for "Three" and "More than three". When processing the data, we treat anyone that reports at least one dose in either the registration survey or follow-up survey as being "vaccinated", and anyone that reports at least two doses in either survey as being "double vaccinated".

D.1.2 Non-response rates and inclusion/exclusion criteria

There are a total of 3,401,391 participants listed across the 19 rounds of REACT-1. Of these, 1023 did not have valid survey weights, these were removed prior to analysis, leaving a potential sample size of 3,400,368.

When considering behavioural questions, we do not explicitly exclude those with invalid swab results, although we omitted round 1 of the study from our analysis as only a small subset of behavioural questions were asked. Of the 3,238,601 participants with valid weights in study

rounds 2-to-19, a total of 2,177,657 participants at least partially responded to the follow-up survey. Individuals were not required to answer all questions in the follow-up survey, so sample sizes for individual questions may be smaller than this total. Missing responses to individual questions were ignored in the analysis. Sample sizes by study round are provided in Supplementary Table D.5.

When considering behavioural variables on a daily basis, the date the survey was last accessed was used. If this was missing, the date the swab was taken was used. If both dates were missing, and the individual is still reported as responding to the follow-up survey, then their records were ignored. There are 10,502 individuals for which the date is missing while they still responded to the follow-up survey, and this only occurred in Round 12 (n = 5559) and Round 13 (n = 4943).

Comparisons with mobility and stringency data leverage daily aggregations of individual responses. In these cases, we only consider days on which at least 200 individuals responded to the survey.

Table D.5: Total participants and number of participants with and without valid survey weights by round. Also, the number of participants that returned a valid RT-PCR swab, the number with a valid survey date, the number that (at least partially) completed the follow-up survey and had a valid survey date, and the number that satisfy all three criteria. Percentages in parentheses are of those with valid survey weights, as we filter out invalid weights before performing any analysis.

R	Total	Invalid weights	Valid weights	Valid swabs	Valid survey date	Responded to survey & valid survey date	All valid
2	219862	0	219862	159199 (72%)	171591 (78%)	145753 (66%)	132790 (60%)
3	225729	0	225729	162822 (72%)	171869 (76%)	144638 (64%)	135742 (60%)
4	211340	0	211340	154406 (73%)	162494 (77%)	137406 (65%)	129411 (61%)
5	232190	0	232190	174948 (75%)	182694 (79%)	152007 (65%)	144026 (62%)
6	211285	0	211285	160173 (76%)	168089 (80%)	140967 (67%)	133315 (63%)
7	212891	2	212889	168181 (79%)	176050 (83%)	148656 (70%)	141011 (66%)
8	214751	1015	213736	167625 (78%)	175596 (82%)	154060 (72%)	145059 (68%)
9	210000	0	210000	165456 (79%)	172807 (82%)	150315 (72%)	143118 (68%)
10	180037	0	180037	140844 (78%)	148698 (83%)	129689 (72%)	122022 (68%)
11	170715	0	170715	127407 (75%)	134783 (79%)	118398 (69%)	111187 (65%)
12	161421	0	161421	108911 (67%)	110920 (69%)	93716 (58%)	91812 (57%)
13	146460	1	146459	98231 (67%)	100230 (68%)	85018 (58%)	83124 (57%)
14	146860	0	146860	100527 (68%)	108008 (74%)	93281 (64%)	86088 (59%)
15	142905	0	142905	100112 (70%)	107976 (76%)	95791 (67%)	88504 (62%)
16	130351	0	130351	97089 (74%)	101493 (78%)	88903 (68%)	83866 (64%)
17	139789	0	139789	102174 (73%)	108764 (78%)	98579 (71%)	92478 (66%)
18	134252	1	134251	94950 (71%)	99985 (74%)	87104 (65%)	82416 (61%)
19	148782	0	148782	109181 (73%)	115177 (77%)	102850 (69%)	97275 (65%)

Table D.6: Total valid responses (individuals who responded to the follow-up survey and have a valid survey response date), the number of individuals with missing responses to each of four specific questions, and the number of individuals with responses to all four questions. Percentages in parentheses are of the "total valid" column.

R	Total valid	Is shielding	Didn't leave home	No contacts outside home	Wears a face covering	Total valid (for all four)
2	145753	2690 (1.8%)	2348 (1.6%)	5720 (3.9%)	-	138931 (95.3%)
3	144638	1507 (1.0%)	1711 (1.2%)	4831 (3.3%)	-	139807 (96.7%)
4	137406	1357 (1.0%)	1564 (1.1%)	4800 (3.5%)	-	132603 (96.5%)
5	152007	1609 (1.1%)	1856 (1.2%)	5292 (3.5%)	-	146714 (96.5%)
6	140967	133 (0.1%)	362 (0.3%)	2484 (1.8%)	894 (0.6%)	138476 (98.2%)
7	148656	1760 (1.2%)	-	8894 (6.0%)	2105 (1.4%)	139762 (94.0%)
8	154060	1135 (0.7%)	1201 (0.8%)	6793 (4.4%)	1330 (0.9%)	147266 (95.6%)
9	150315	1109 (0.7%)	1179 (0.8%)	6478 (4.3%)	1343 (0.9%)	143836 (95.7%)
10	129689	1134 (0.9%)	1233 (1%)	6617 (5.1%)	1516 (1.2%)	123068 (94.9%)
11	118398	882 (0.7%)	950 (0.8%)	5302 (4.5%)	1182 (1.0%)	113093 (95.5%)
12	93716	611 (0.7%)	662 (0.7%)	2112 (2.3%)	889 (0.9%)	91602 (97.7%)
13	85018	573 (0.7%)	613 (0.7%)	1832 (2.2%)	828 (1.0%)	83185 (97.8%)
14	93281	711 (0.8%)	763 (0.8%)	2171 (2.3%)	1107 (1.2%)	91110 (97.7%)
15	95791	622 (0.6%)	710 (0.7%)	2793 (2.9%)	803 (0.8%)	92998 (97.1%)
16	88903	582 (0.7%)	676 (0.8%)	2475 (2.8%)	749 (0.8%)	86427 (97.2%)
17	98579	681 (0.7%)	764 (0.8%)	2664 (2.7%)	829 (0.8%)	95915 (97.3%)
18	87104	566 (0.6%)	643 (0.7%)	2549 (2.9%)	725 (0.8%)	84553 (97.1%)
19	102850	626 (0.6%)	717 (0.7%)	2972 (2.9%)	806 (0.8%)	99877 (97.1%)

D.2 Additional descriptive results

D.2.1 Additional demographic stratification

In the main paper we consider responses to the four key risk-related behavioural questions by age-group and household size. In this section we consider responses to these four questions by additional demographic disaggregations: region, deprivation quintile of neighbourhood, ethnicity, and sex (Supplementary Figures D.1-D.4).

For visual clarity, we present results by region using four regions. These are created by aggregating the standard nine English regions as follows: "North" includes North West, North East, and Yorkshire and The Humber; "Midlands and East" includes West Midlands, East Midlands, East of England, "London" (London only), and "South England" including South West and South East.

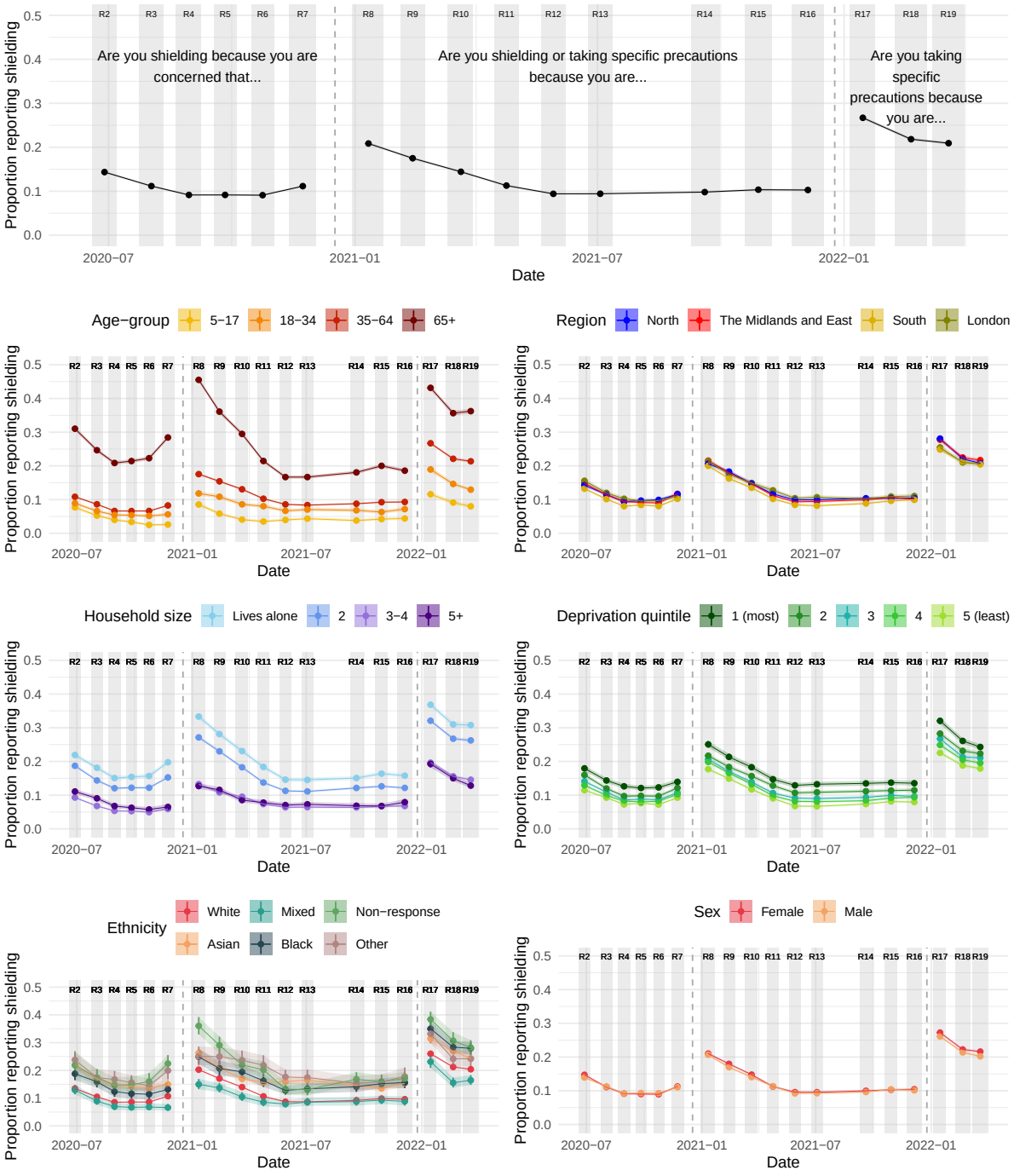

Figure D.1: The proportion of people that report shielding and/or taking specific precautions because they consider themselves "to be at risk of severe illness for COVID-19". The question wording changes between study rounds 7 and 8, and again between study rounds 16 and 17 - this is marked by a dashed grey line. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion. Individual points are joined by lines and shading for clarity, but this should not be interpreted as an interpolation.

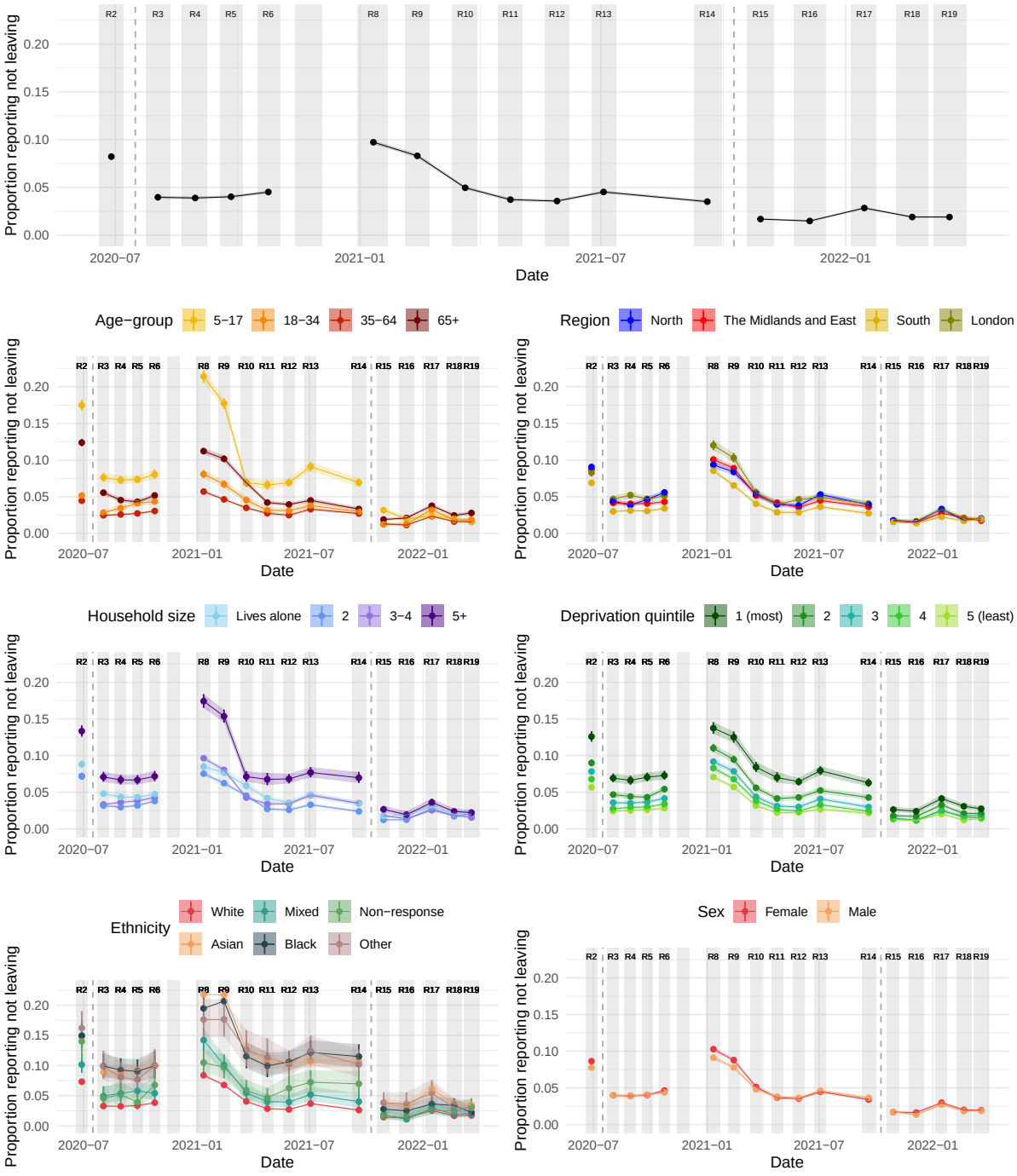

Figure D.2: The proportion of people that report not leaving the home in 7 days prior to completing the questionnaire. The question wording changes between study rounds 2 and 3, and again between study rounds 14 and 15 - this is marked by a dashed grey line. This question was not asked in round 7. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion. Individual points are joined by lines and shading for clarity, but this should not be interpreted as an interpolation.

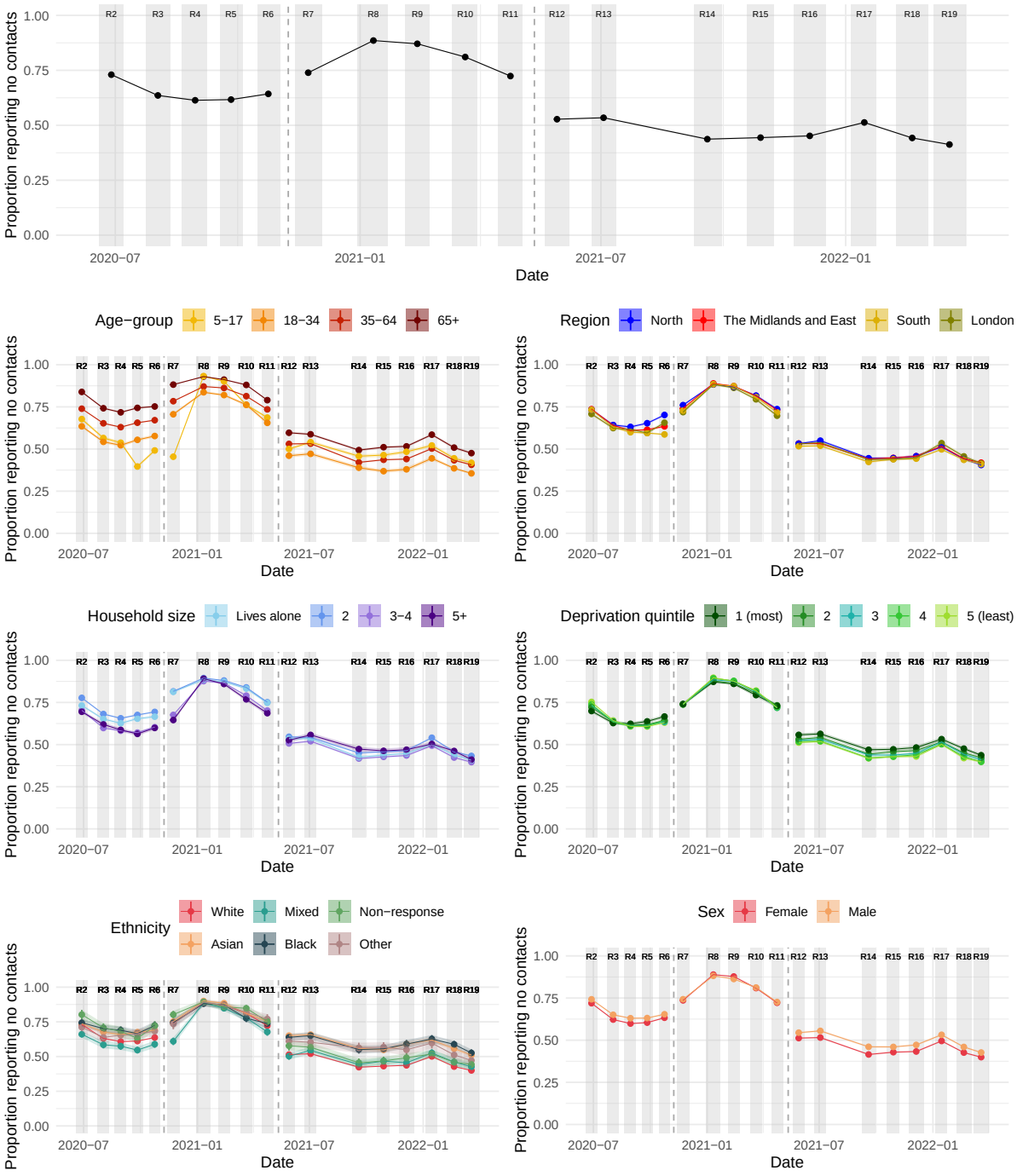

Figure D.3: The proportion of people that report having no contacts outside the home on the day preceding the questionnaire. The question wording changes between study rounds 6 and 7, and again between study rounds 11 and 12 - this is marked by a dashed grey line. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion. The mean number of contacts are also presented in Supplementary Figure D.5. Individual points are joined by lines and shading for clarity, but this should not be interpreted as an interpolation.

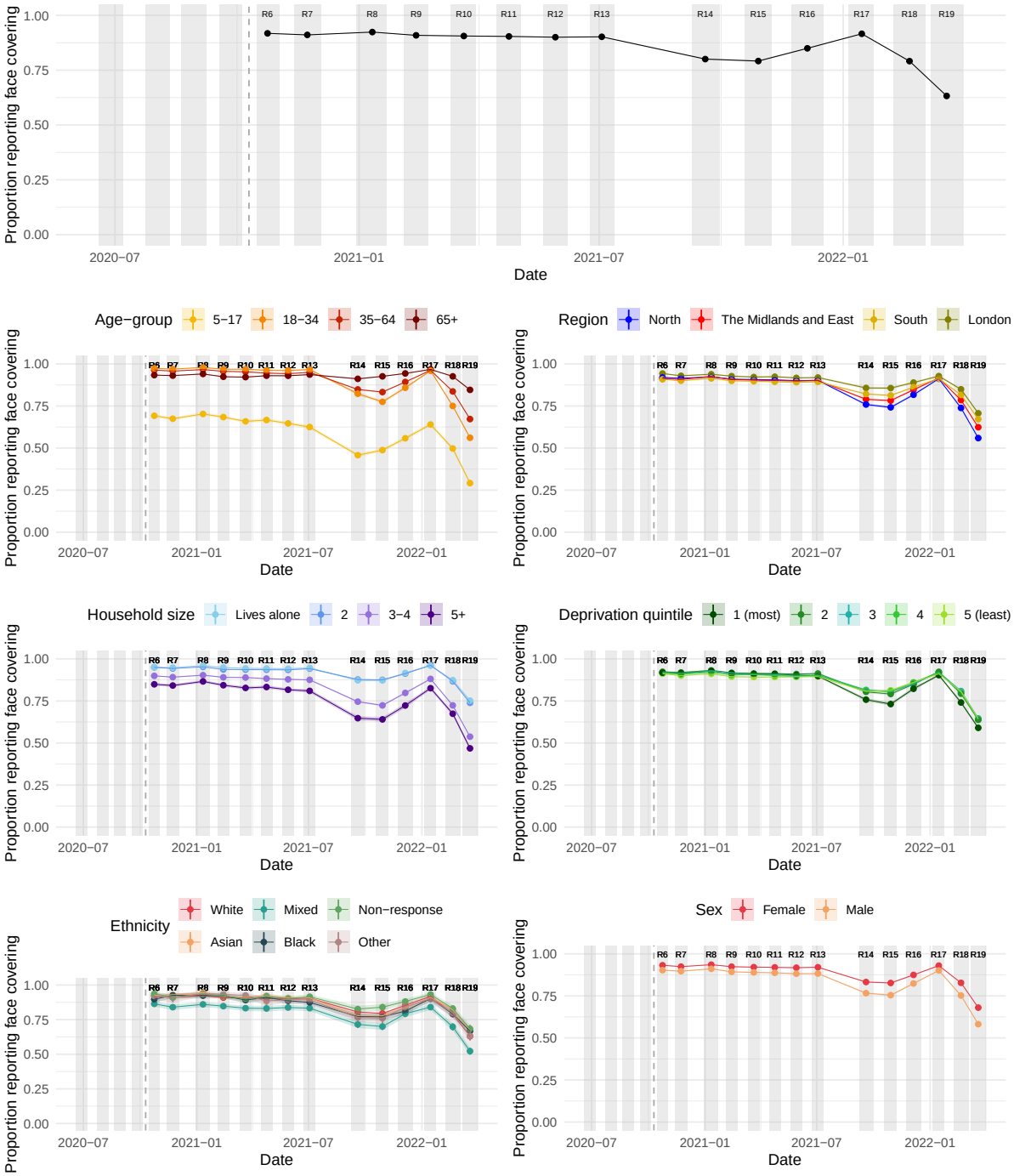

Figure D.4: The proportion of people that report wearing a face covering at least sometimes when leaving the home. This question was asked in study round 6 onwards. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion. Individual points are joined by lines and shading for clarity, but this should not be interpreted as an interpolation.

D.2.2 Mean number of contacts

Thus far, we have only considered the proportion of people that would report having no contacts on the day prior to answering the survey. However, the question outlined in Supplementary Section D.1.1 asks about the number of contacts, allowing us to additionally consider the mean number of contacts that people would report. We use the svymean() function from the survey package in R to calculate the survey mean and 95% confidence intervals. These are reported for the standard demographic stratifications in Supplementary Figure D.5. To prevent rare extremely high reported total contacts from biasing the results (over the study, a total of 350 individuals reported at least 500 contacts, 137 individuals reported at least 1,000 contacts, and 17 reported at least 10,000 contacts), we remove responses with at least 1,000 reported contacts.

Despite removing individual outliers in these data, Supplementary Figure D.5 highlights high mean total contacts in 5-17-year-olds in rounds 5, 6, and 7, coinciding with the re-opening of schools in September 2020. These high means were not repeated in the same period in 2021 (study rounds 14 and 15), suggesting that parents were more likely to report higher rates of "close contacts" for their children in 2020 than 2021.

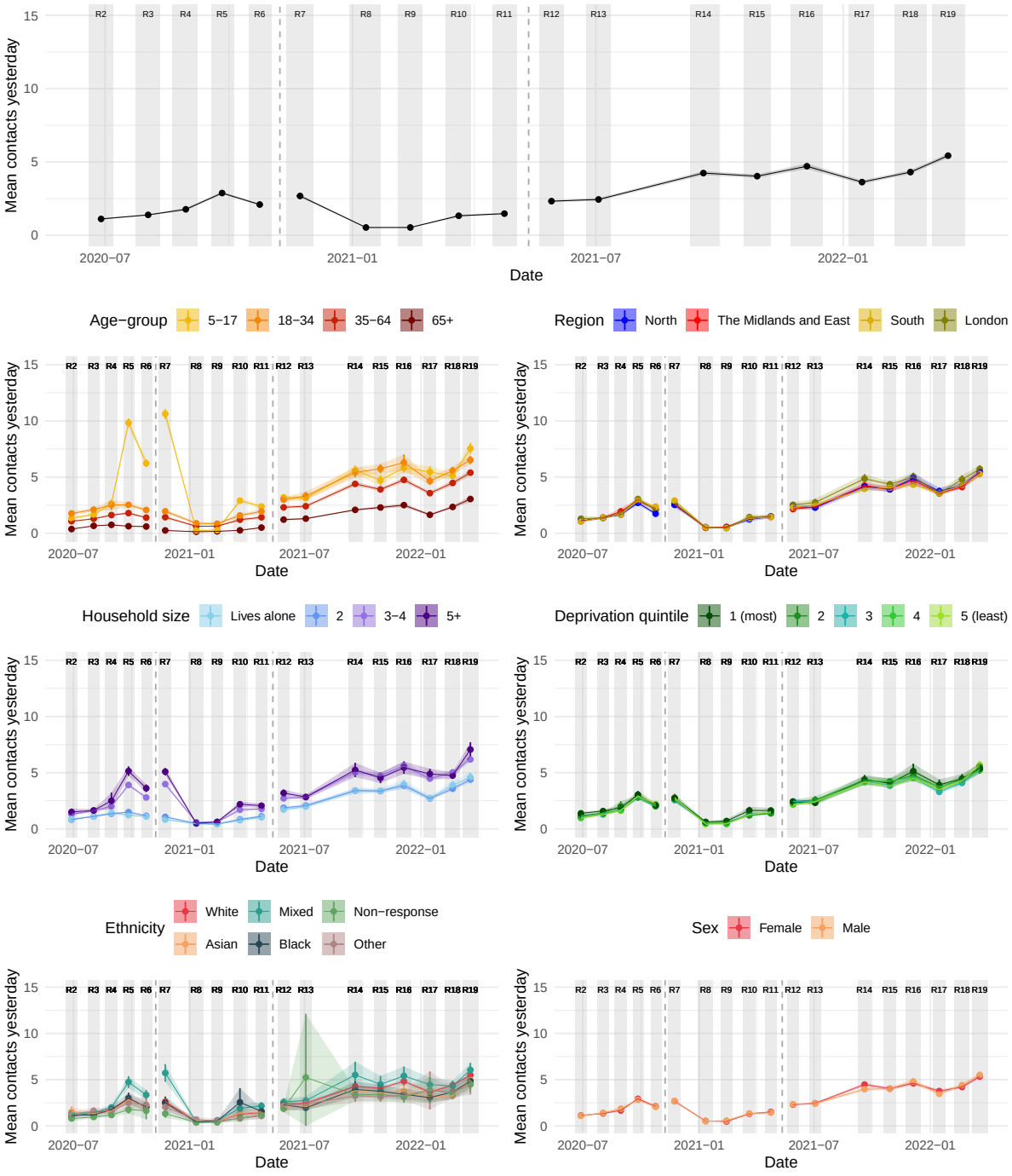

Figure D.5: The mean number of contacts that people would report having on the day preceding the questionnaire. The question wording changes between study rounds 6 and 7, and again between study rounds 11 and 12 - this is marked by a dashed grey line. Shaded areas and vertical lines show 95% confidence intervals about the estimated mean.

D.2.3 Risk beliefs, vaccination, and infection

The proportion of people that would report believing they are at risk of severe illness for COVID-19 remained relatively stable throughout the pandemic (Supplementary Figure D.6). A decrease was observed between November 2020 and January 2021 (study rounds 7 and 8), however this corresponded with the addition of "Don't know" as an answer to the question. Older people, those living in neighbourhoods with greater levels of socioeconomic deprivation, and those living alone were more likely to report believing they are at risk of severe illness.

The proportion of people that reported a suspected or confirmed past infection fell from 16.7% in May 2020 (study round 2) to 13.3% in November 2020 (study round 7) (Supplementary Figure D.8), while the proportion of people that reported a confirmed past infection (by positive test) increased from 0.3% to 1.6% over the same time period (Supplementary Figure D.9).

Supplementary Figures D.10 and D.11 show a distinct decrease in the proportion of 18-34-year-olds that report being vaccinated between January 2022 (study round 17) and February 2022 (study round 18). In fact, between these study rounds, the proportion of 18-34-year-olds that would report not being vaccinated increased from 5.1% (95% confidence interval 4.8%, 5.6%) to 9.1% (8.8%, 9.4%). The proportion of this group that would report not being double vaccinated increased from 7.9% (7.4%, 8.3%) to 14.0% (13.6%, 14.4%).

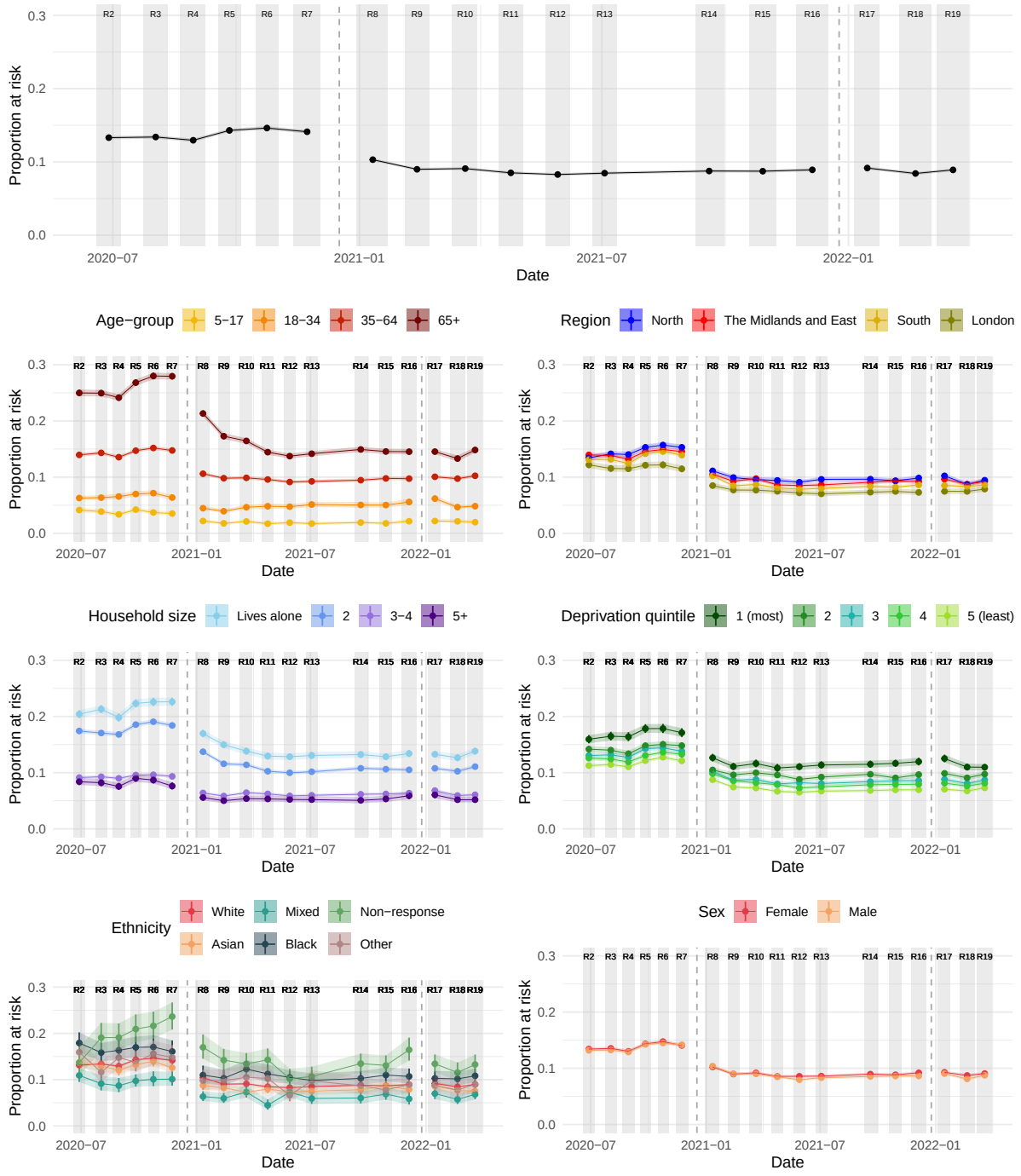

Figure D.6: The proportion of people in England that would report believing they are "at risk of severe illness for COVID-19". The question wording changes between rounds 7 and 8, and again between rounds 16 and 17 - these changes are marked by vertical dashed lines. The list of possible answers also changes between rounds 7 and 8, with the addition of a "Don't know" option. A detailed description of these changes is given in Supplementary Table D.4. We report the proportion of people that would respond "Yes" to this question. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion.

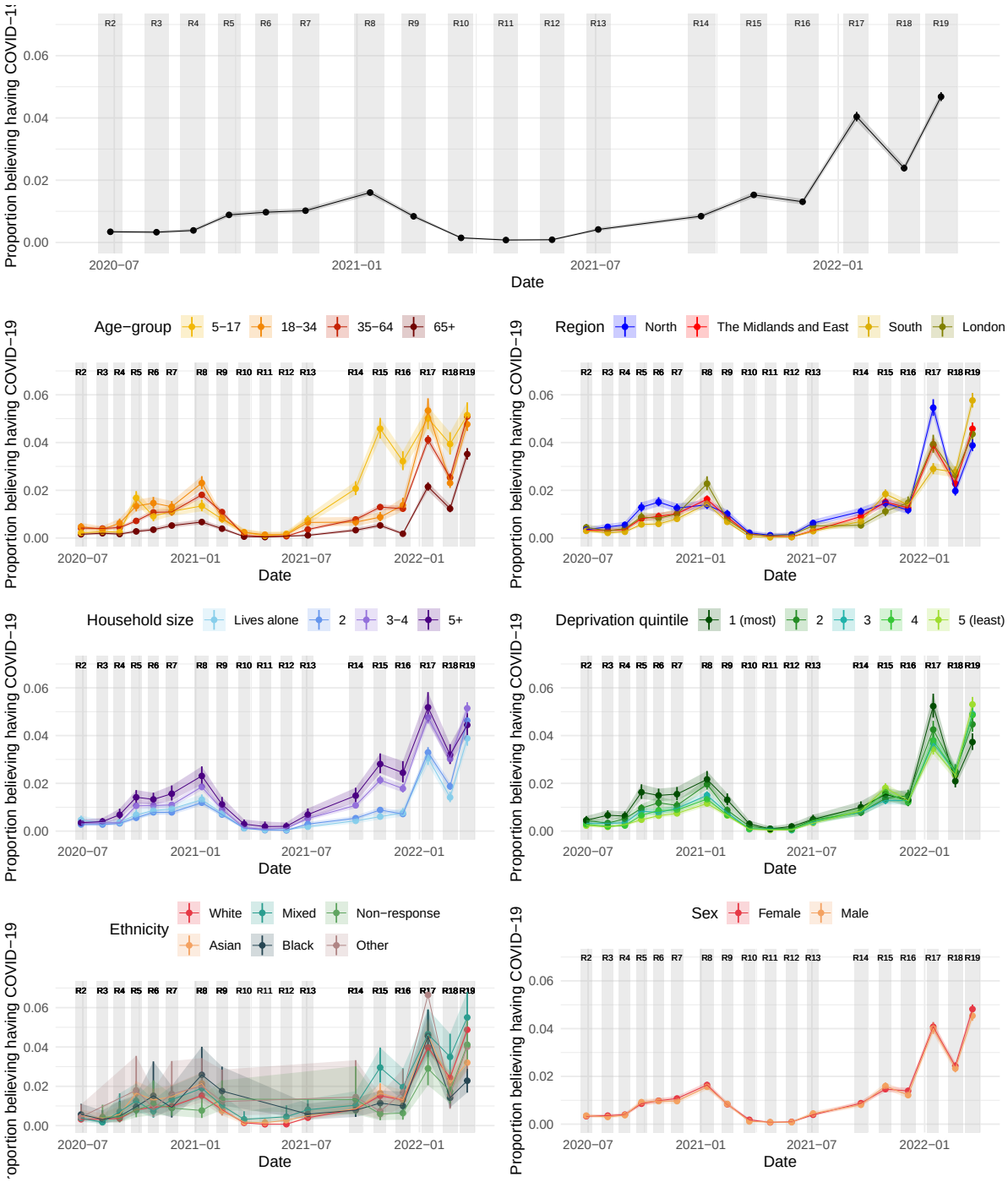

Figure D.7: The proportion of people in England that would report a recent infection (first suspected/tested positive within the preceding 13 days), either confirmed by a positive test, suspected by a doctor, or by their own suspicions. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion.

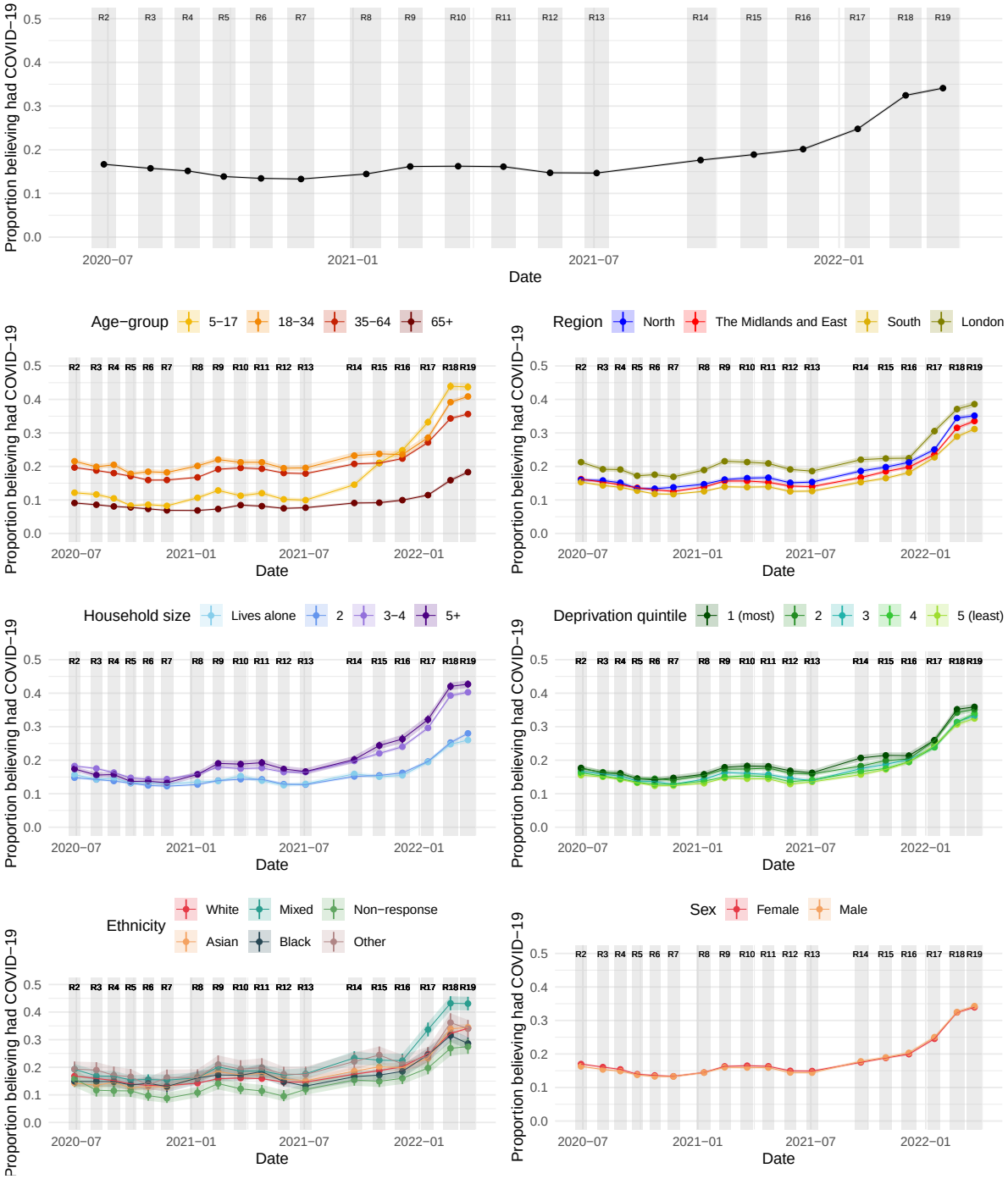

Figure D.8: The proportion of people in England that would report a past infection (first suspected/tested positive at least 14 days ago), either confirmed by a positive test, suspected by a doctor, or by their own suspicions. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion.

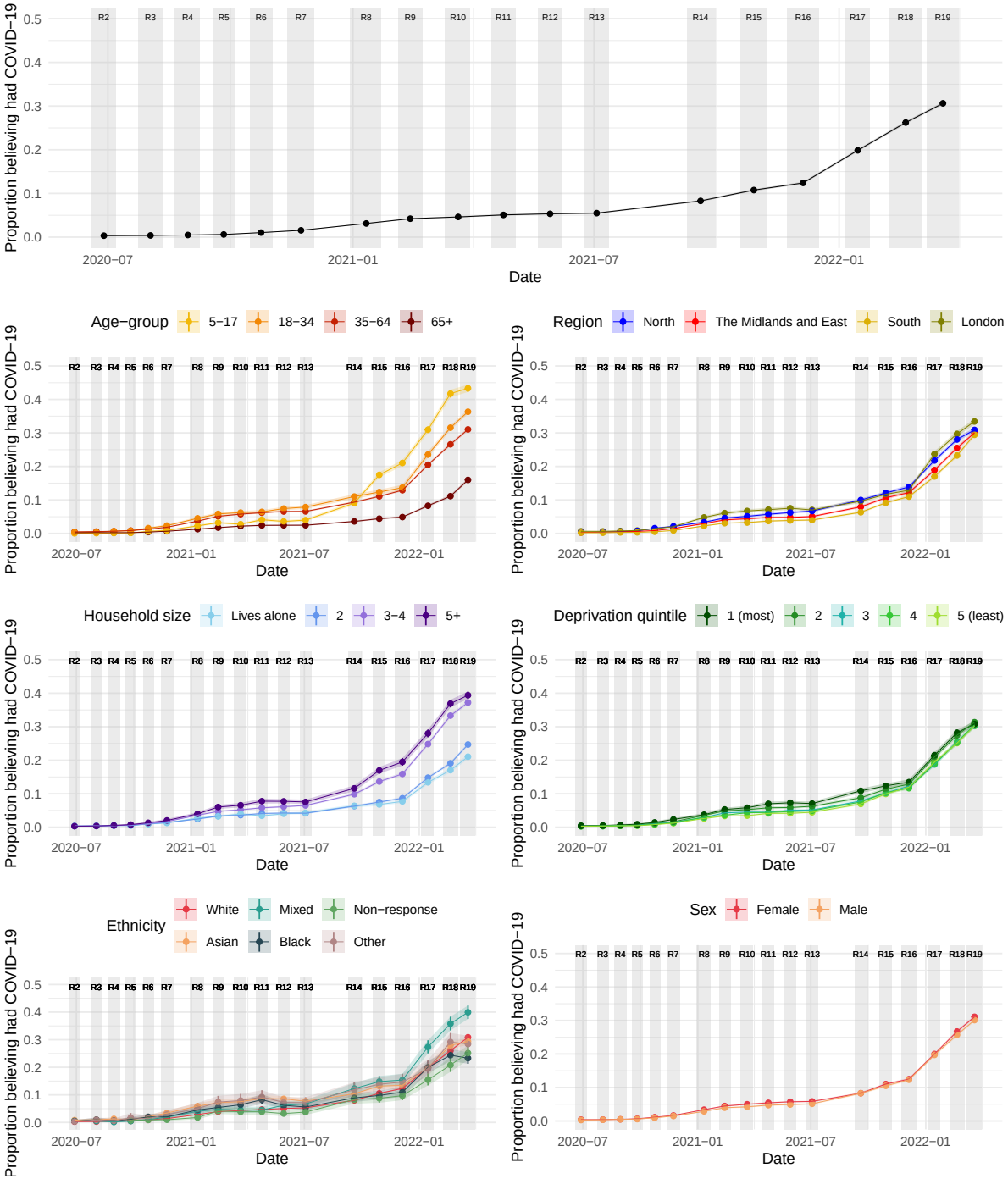

Figure D.9: The proportion of people in England that would report a past infection confirmed by a positive test, at least 14 days before completing the questionnaire. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion.

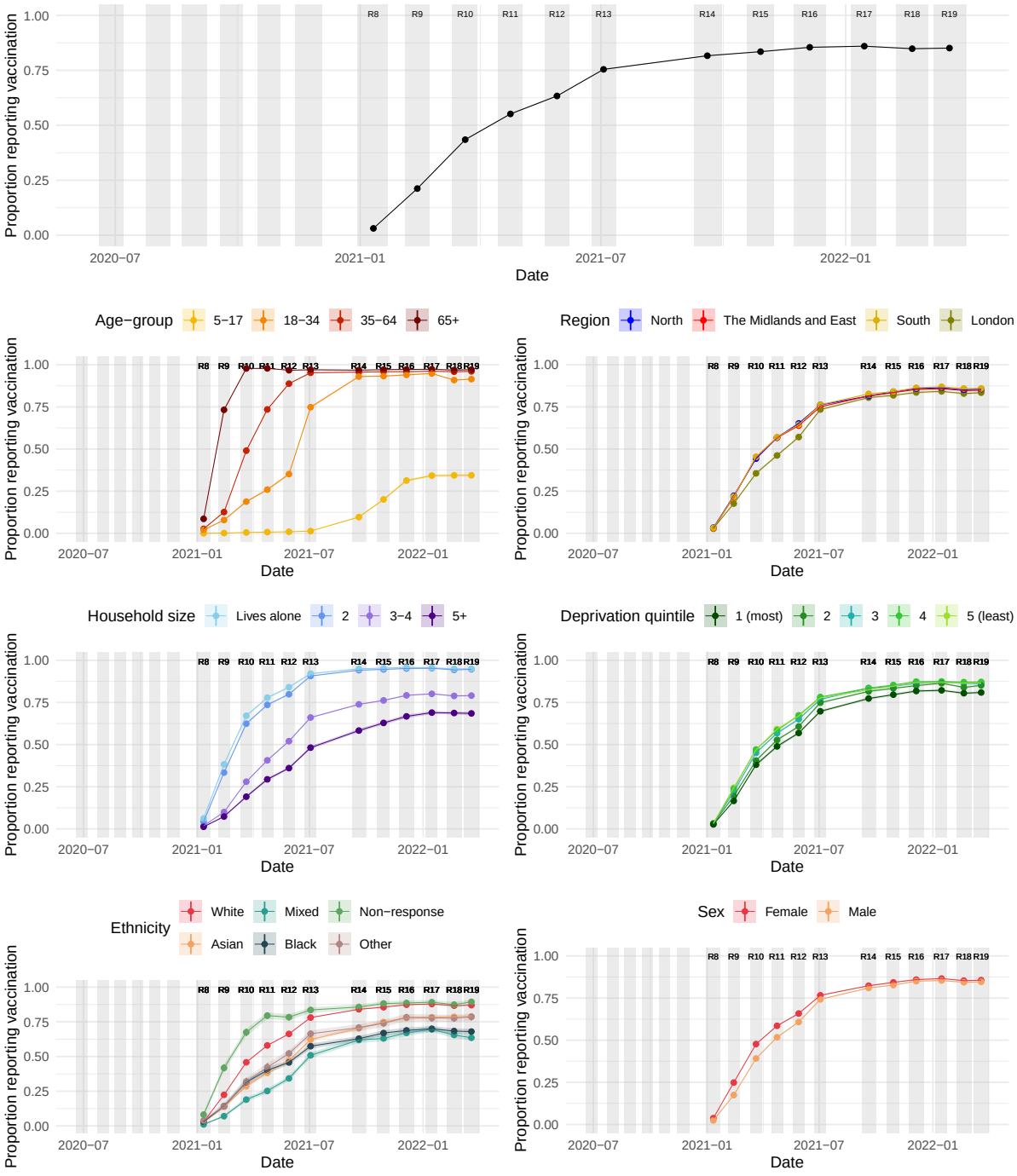

Figure D.10: The proportion of people in England that would report having received at least one dose of a COVID-19 vaccine. Shaded regions and vertical lines show 95% confidence intervals about the estimated proportions.

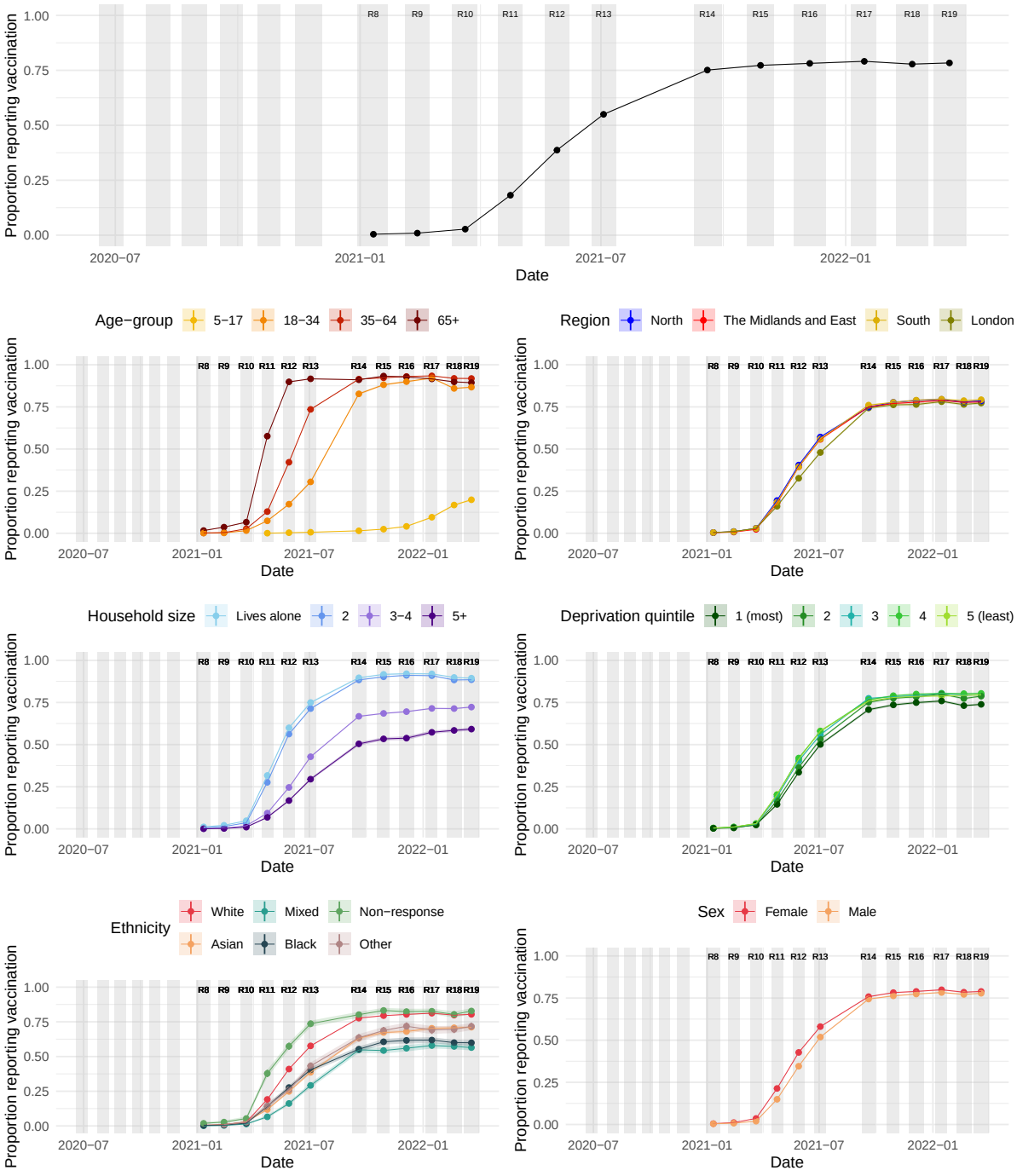

Figure D.11: The proportion of people in England that would report having received at least two doses of a COVID-19 vaccine. Shaded regions and vertical lines show 95% confidence intervals about the estimated proportions.

D.2.4 Behaviours stratified by risk beliefs, vaccination, and infection

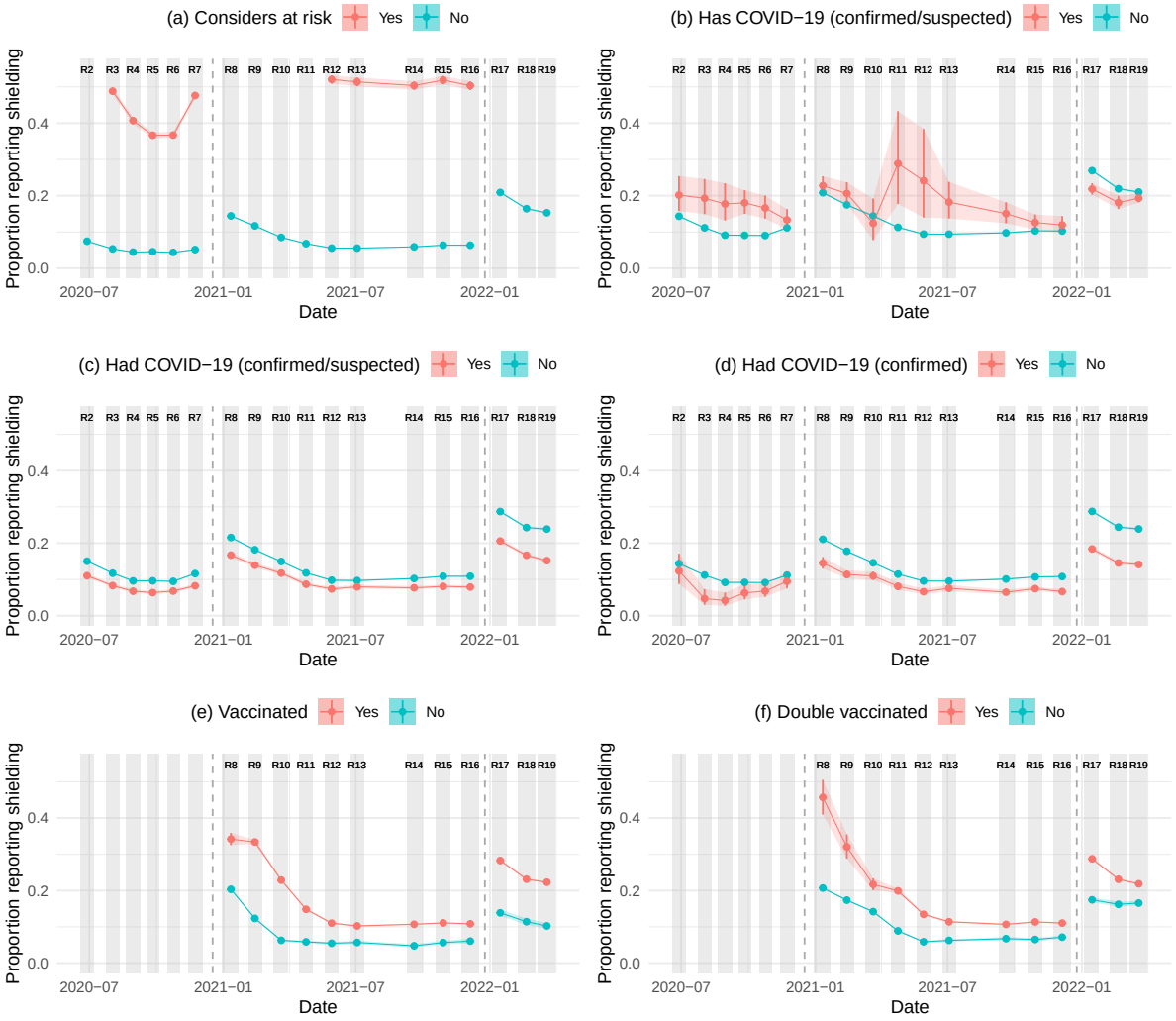

Figure D.12: The proportion of people that report shielding and/or taking specific precautions because they consider themselves "to be at risk of severe illness for COVID-19". Results are conditioned on whether the individual (a) reports considering themselves to be at risk of severe illness, (b) reports a recent confirmed and/or suspected COVID-19 infection, (c) reports a past confirmed and/or suspected COVID-19 infection, (d) reports a past COVID-19 infection confirmed by a positive test, (e) reports receiving at least one dose of a COVID-19 vaccination, and (f) reports receiving at least two doses of a COVID-19 vaccination. The question wording changes between study rounds 7 and 8, and again between study rounds 16 and 17 - this is marked by a dashed grey line. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion. The y-axis for (a) people who consider themselves at risk of severe illness is different to the remaining subplots.

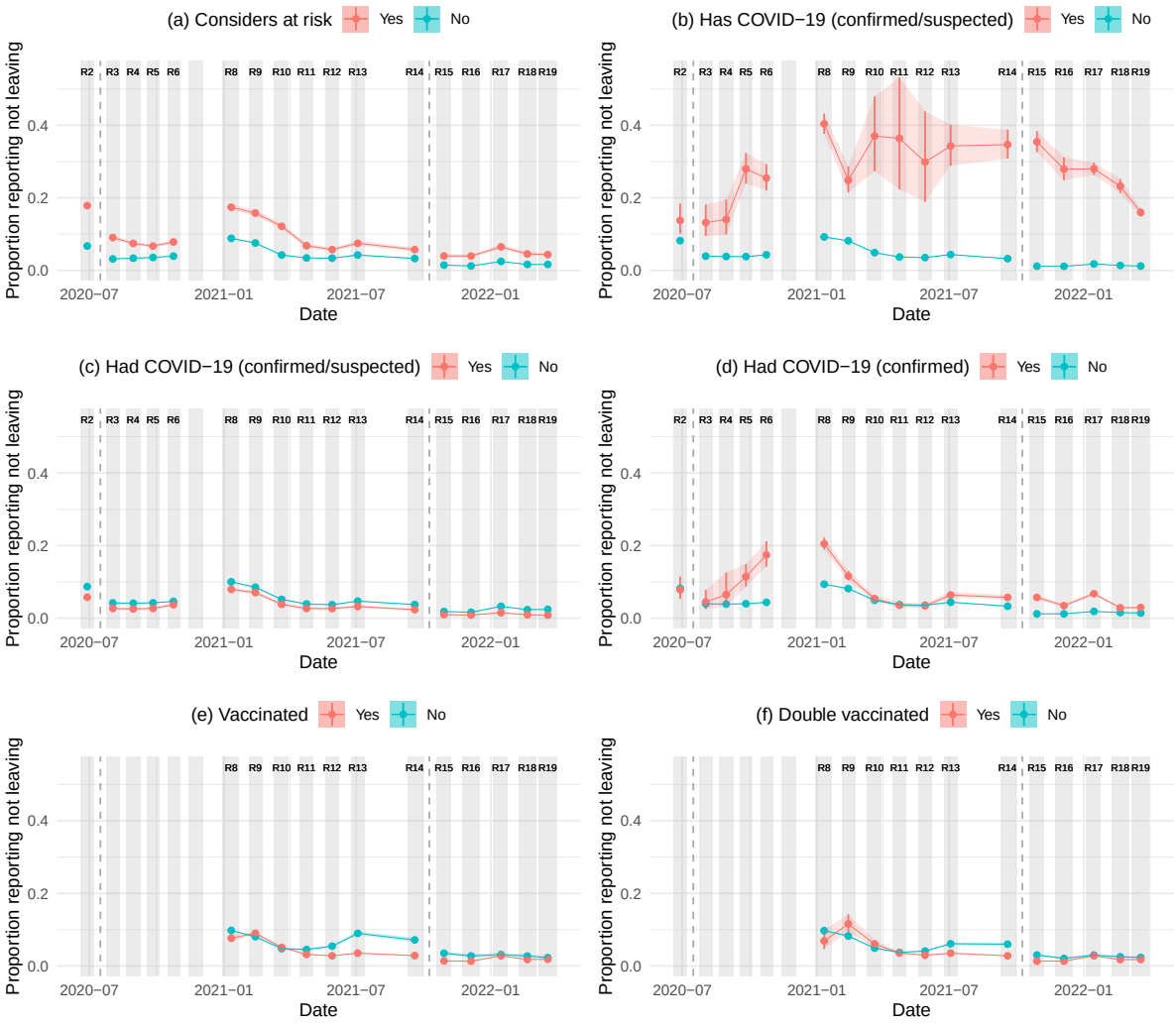

Figure D.13: The proportion of people that report not leaving the home in the 7 days prior to completing the questionnaire. Results are conditioned on whether the individual (a) reports considering themselves to be at risk of severe illness, (b) reports a recent confirmed and/or suspected COVID-19 infection, (c) reports a past confirmed and/or suspected COVID-19 infection, (d) reports a past COVID-19 infection confirmed by a positive test, (e) reports receiving at least one dose of a COVID-19 vaccination, and (f) reports receiving at least two doses of a COVID-19 vaccination. The question wording changes between study rounds 2 and 3, and again between study rounds 14 and 15 - this is marked by a dashed grey line. This question was not asked in round 7. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion.

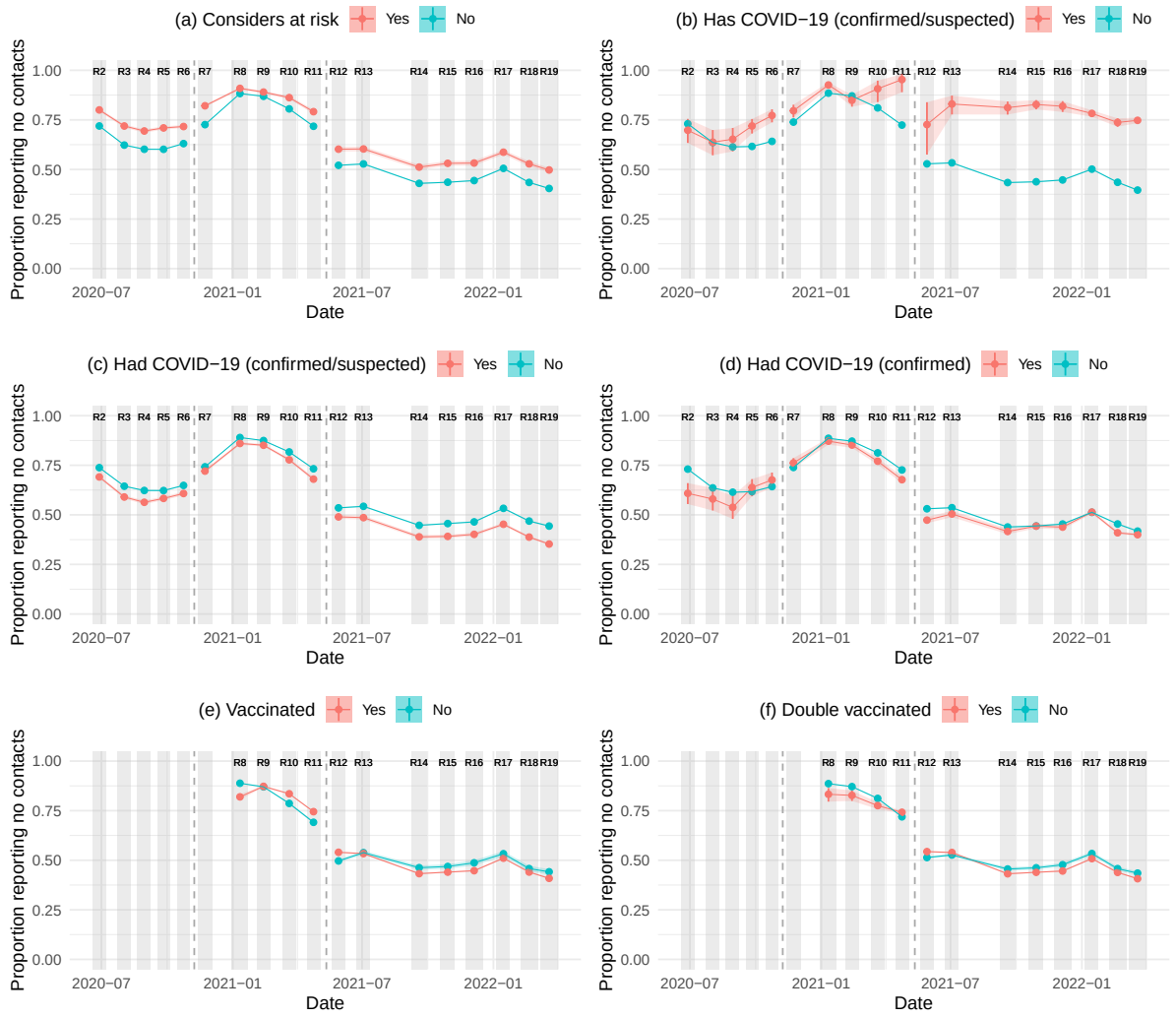

Figure D.14: The proportion of people that report having no contacts outside the home on the day preceding the questionnaire. Results are conditioned on whether the individual (a) reports considering themselves to be at risk of severe illness, (b) reports a recent confirmed and/or suspected COVID-19 infection, (c) reports a past confirmed and/or suspected COVID-19 infection, (d) reports a past COVID-19 infection confirmed by a positive test, (e) reports receiving at least one dose of a COVID-19 vaccination, and (f) reports receiving at least two doses of a COVID-19 vaccination. The question wording changes between study rounds 6 and 7, and again between study rounds 11 and 12 - this is marked by a dashed grey line. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion.

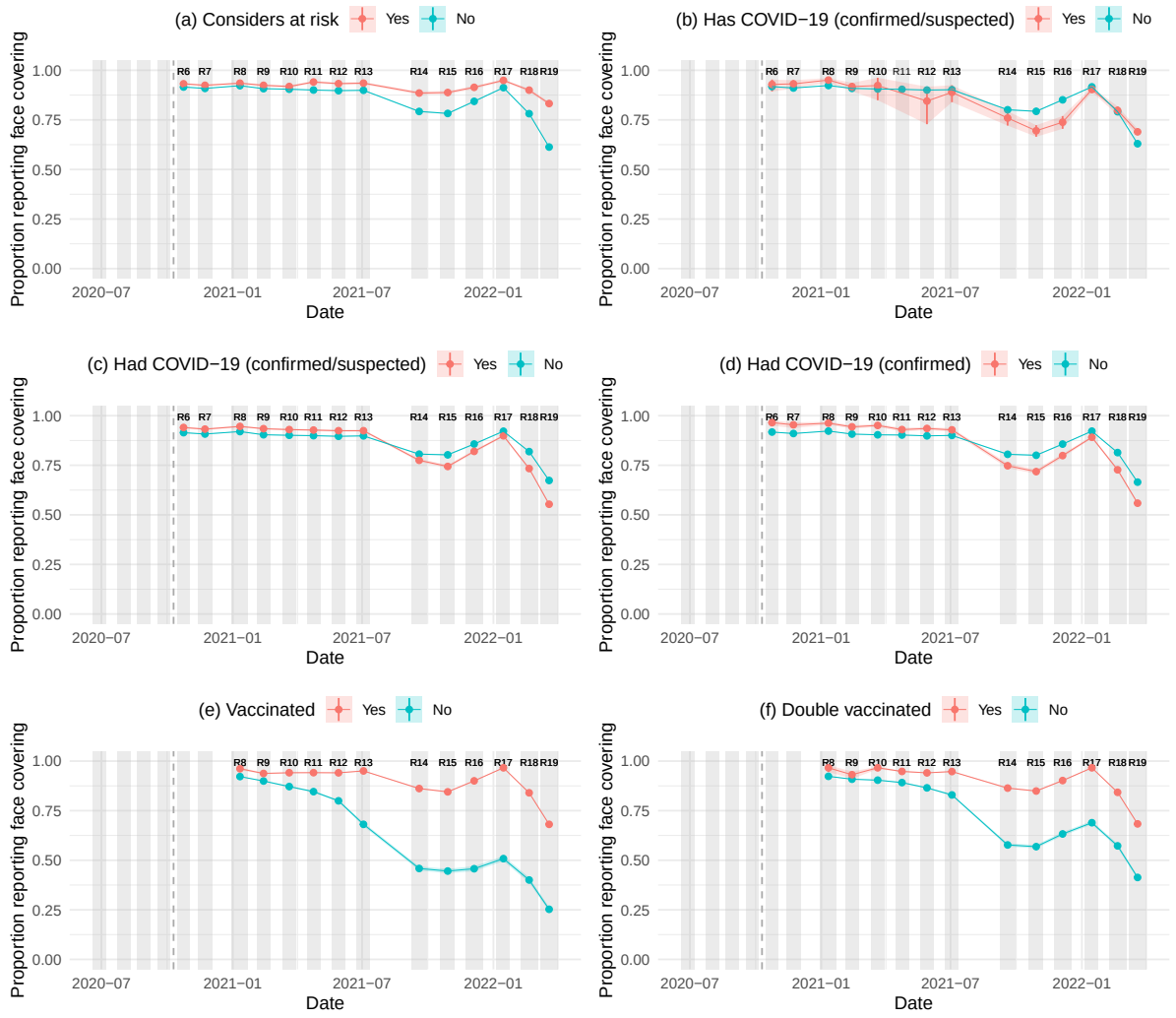

Figure D.15: The proportion of people that report wearing a face covering at least sometimes when leaving the home. Results are conditioned on whether the individual (a) reports considering themselves to be at risk of severe illness, (b) reports a recent confirmed and/or suspected COVID-19 infection, (c) reports a past confirmed and/or suspected COVID-19 infection, (d) reports a past COVID-19 infection confirmed by a positive test, (e) reports receiving at least one dose of a COVID-19 vaccination, and (f) reports receiving at least two doses of a COVID-19 vaccination. This question was asked in study round 6 onwards. Shaded areas and vertical lines show 95% confidence intervals about the estimated proportion.

D.2.5 Reasons for leaving the home

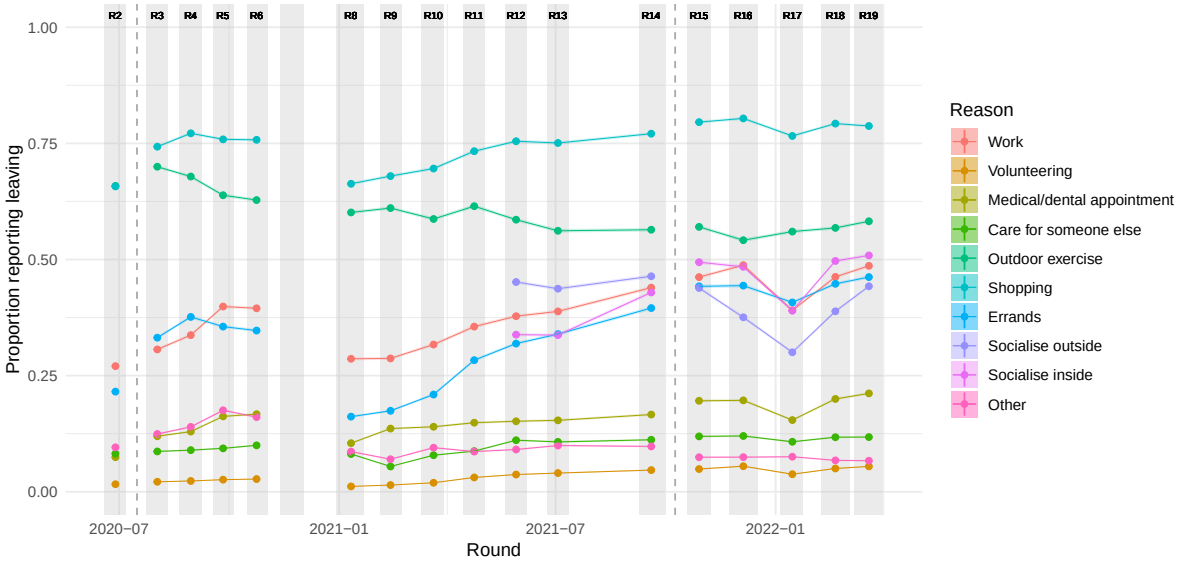

Figure D.16: The proportion of people that report leaving the home for various reasons. Coloured shaded areas and vertical lines show 95% confidence intervals about the estimated proportion. Vertical dashed lines indicate changes in question wording. This question was not asked in study round 7.

D.3 Logistic regression models

Supplementary Figures D.17 to D.27 present the results from logistic regression models that isolate the impact of each demographic variable on various reported behaviours. We include a summary of the results here.

Those who are more likely to report leaving to go to work (Figure S17) are:

- Those aged 18-34 years, and to a lesser extent aged 35-64, when compared to those aged 65+ and 5-17.
- Males
- Those not living in London in rounds 2 to 13, with the strongest effect being in rounds 8 and 9, as well as round 17. This effect reverses in round 19, with those living in London being more likely to report leaving home to go to work.
- Those in areas with greater socioeconomic deprivation (deprivation quintiles 4 & 5 [least deprived] are less likely to report leaving to work than those in deprivation quintile 1 [most deprived])
- Those in larger households

Those who are more likely to report leaving to volunteer (Figure S18) are:

- Older people (those aged 35-64 and 65+ years)
- Females in rounds 10 onwards
- People living in London (and those living in the South West in rounds 4 to 6 and 14 to 17)
- People in areas with lower socioeconomic deprivation
- People that live alone

Those who are more likely to report leaving to attend a medical or dental appointment (Figure S19) are:

- Older people (those aged 35-64 and 65+ years)

- Females

Those who are more likely to report leaving to care for someone else (Figure S20) are:

- People aged 35-64 years, and also those aged 65+, when compared to those aged 18-34 and those aged 5-17
- People living outside of London, particularly in the North East in earlier rounds
- People in areas with lower socioeconomic deprivation
- People living alone or with one other person (compared to those living in larger households)

Those who are more likely to report leaving for outdoor exercise (Figure S21) are:

- People living in London
- People in areas with lower socioeconomic deprivation
- People identifying as white
- People living in smaller households

Those who are more likely to report leaving to go shopping (Figure S22) are:

- People aged 35-64 years. Also those aged 65+ in rounds 4 to 6 and rounds 11 to 19.
- Females
- People in areas with lower socioeconomic deprivation in rounds 10 to 19
- People living alone

Those who are more likely to report leaving for errands (Figure S23) are:

- People aged 18-34 years
- People in London

Those who are more likely to report leaving for other reasons (Figure S24) are:

- People aged 65+ years and 5-17 years (with exceptions to the latter in study rounds 17, 18, and 19)

- Females
- People in wealthier areas
- People not identifying as white

These "other" reasons should not be compared across rounds because the proportion of people that ticked "other" depends on which pre-set options are available.

Figure D.17: Odds ratios for whether an individual would report shielding and/or taking specific precautions. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. The question wording changed between study rounds 7 and 8, and again between study rounds 16 and 17 as indicated by vertical dashed lines - see Supplementary Section D.1.1 for further detail.

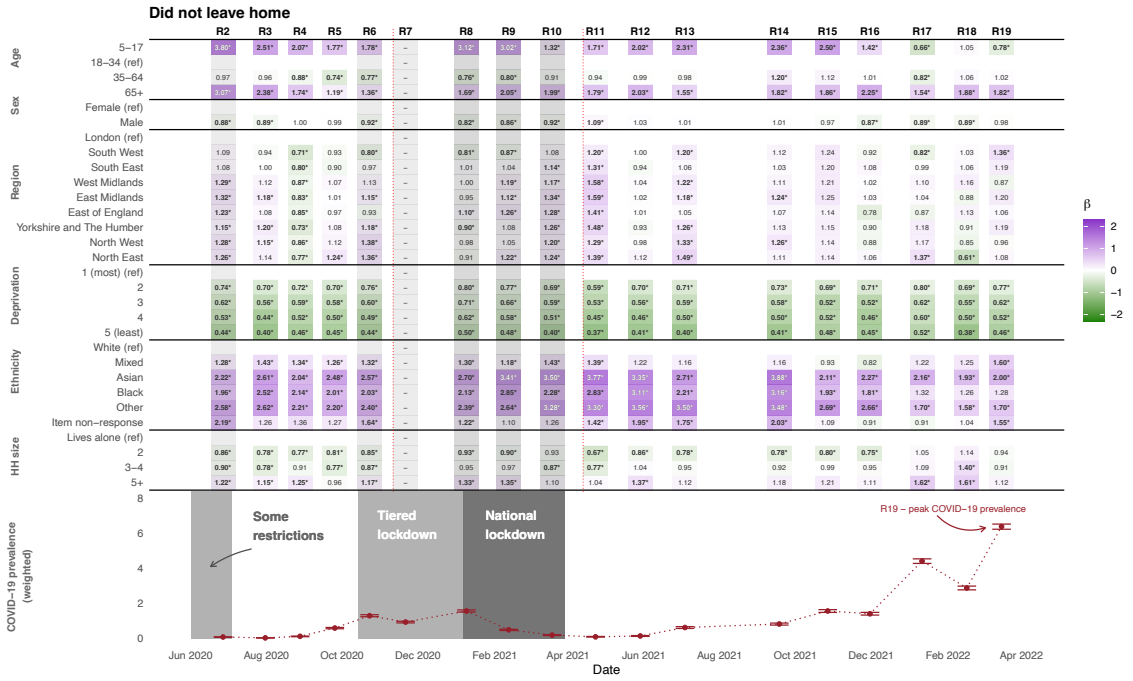

Figure D.18: Odds ratios for whether an individual would report not leaving the house in the 7 days preceding completing the questionnaire. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line.

Figure D.19: Odds ratios for whether an individual would report wearing a face covering outside the home. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked prior to study round 6.

Figure D.20: Odds ratios for whether an individual would report leaving home to go to work. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line.

Figure D.21: Odds ratios for whether an individual would report leaving home to volunteer. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line.

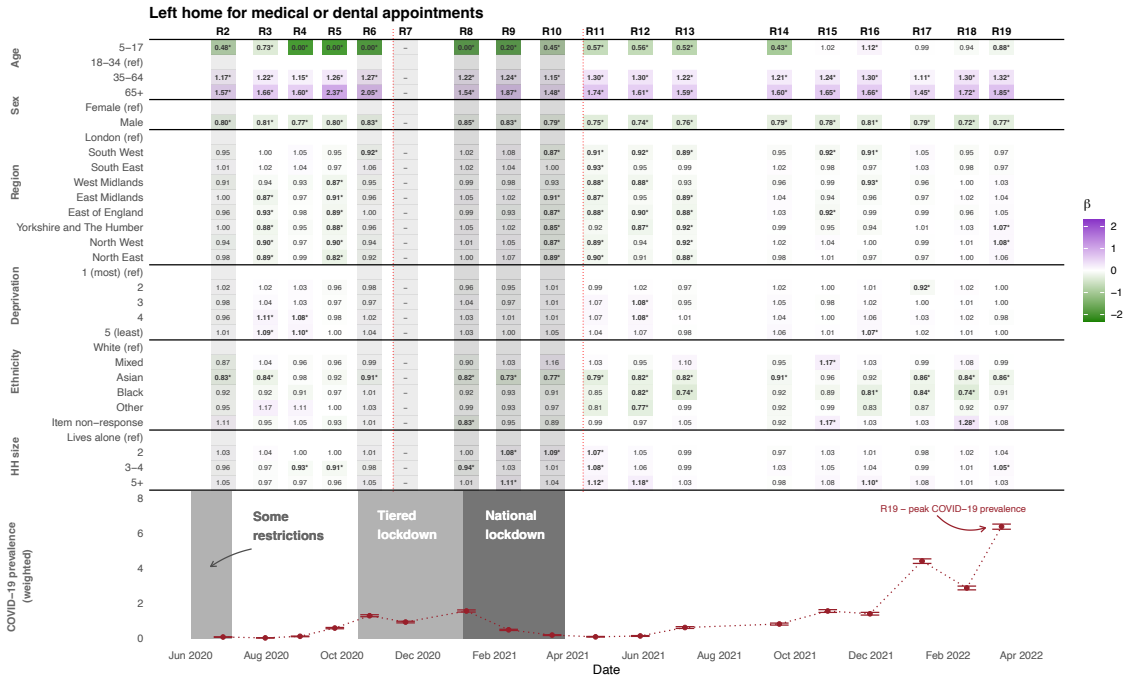

Figure D.22: Odds ratios for whether an individual would report leaving home for medical and/or dental care. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line.

Figure D.23: Odds ratios for whether an individual would report leaving home to care for somebody else. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line.

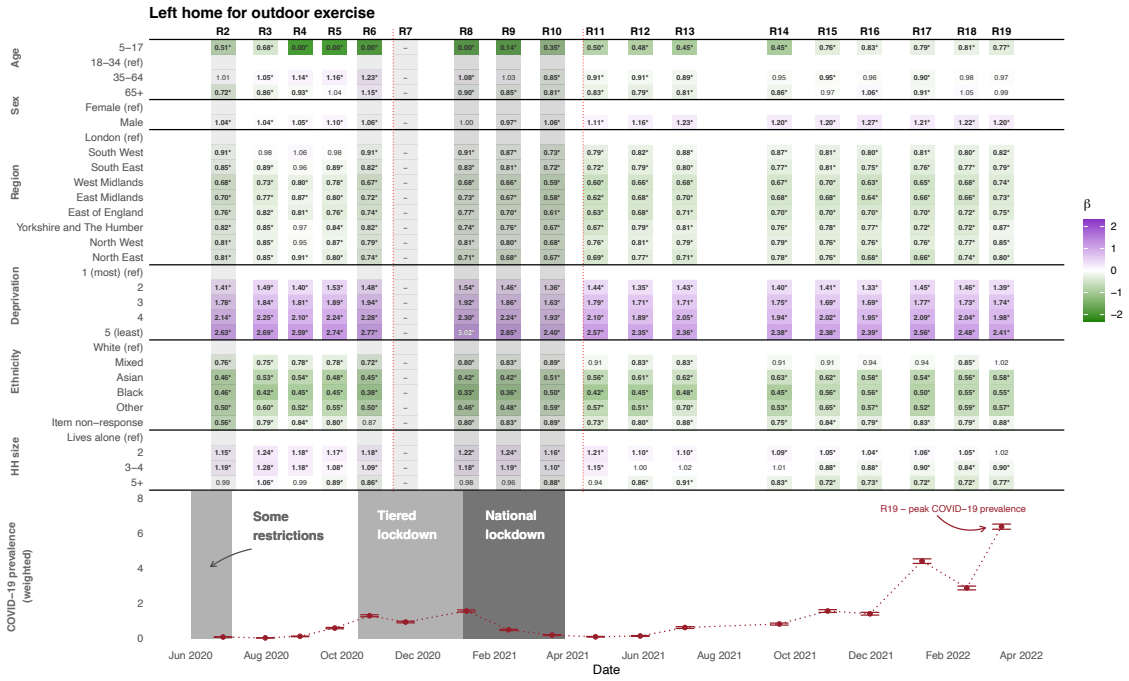

Figure D.24: Odds ratios for whether an individual would report leaving home to exercise outdoors. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line.

Figure D.25: Odds ratios for whether an individual would report leaving home for shopping. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line.

Figure D.26: Odds ratios for whether an individual would report leaving home for errands. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line.

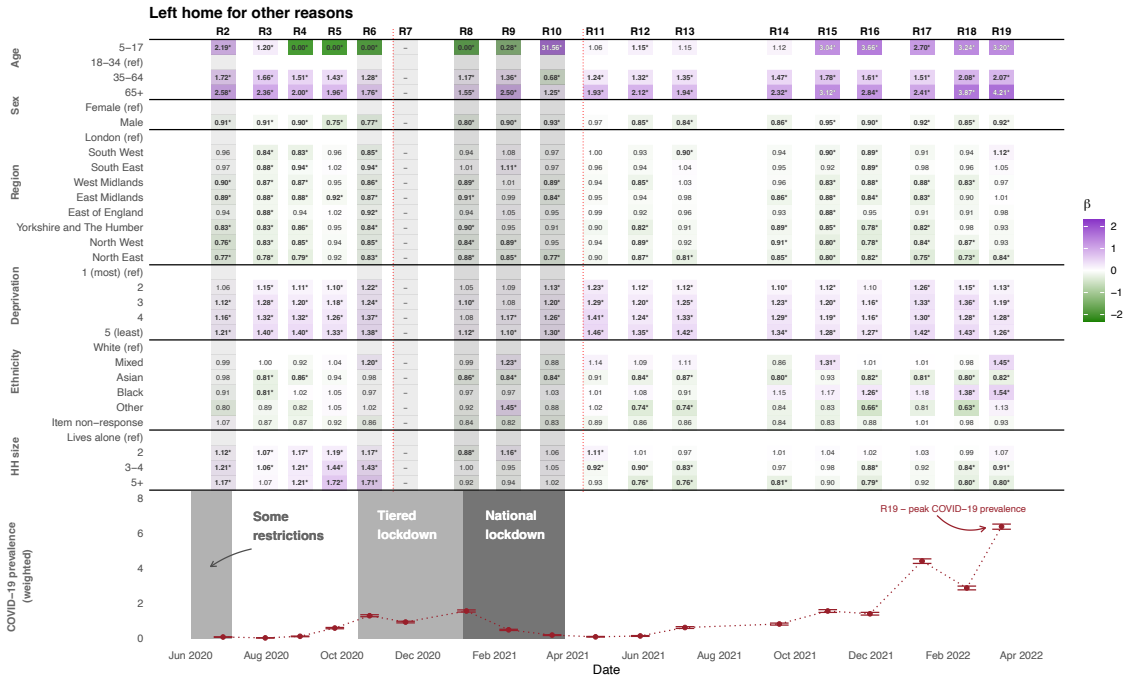

Figure D.27: Odds ratios for whether an individual would report leaving home for other reasons. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. The change in question wording is particularly important here, as "other" may include or exclude specific reasons depending on the available options (see Supplementary Table D.3).

Supplementary Figures D.28 to D.39 present the results from logistic regression models that isolate the impact of each demographic variable on various reported behaviours, with a further adjustment for a binary variable indicating whether the respondent thought they had had COVID in the preceding two weeks (question "COVIDA", described in Supplementary Section D.1.1). The results of the logistic regression models are robust to this additional adjustment.

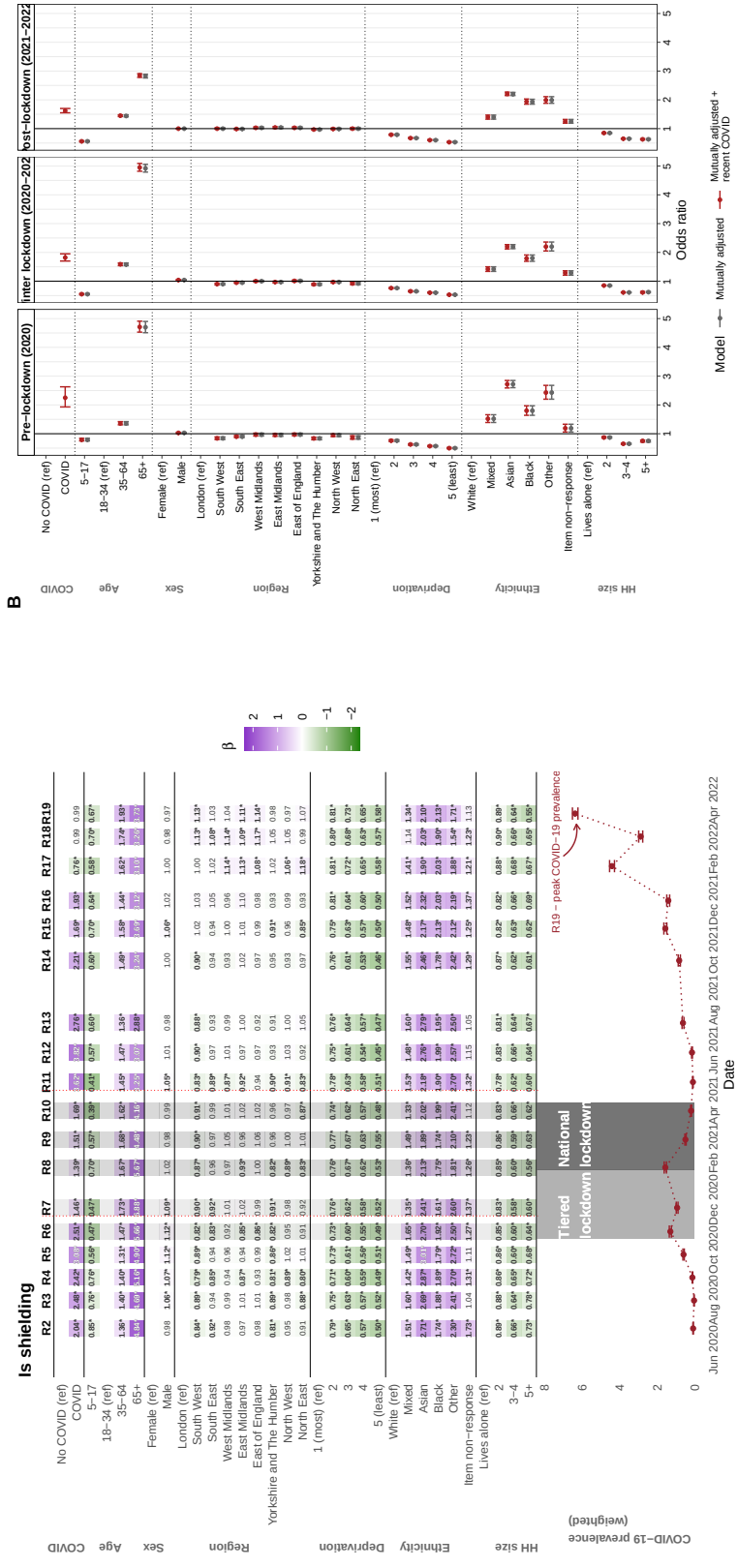

Figure D.28: A. Odds ratios for whether an individual would report shielding and/or taking specific precautions. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. The question wording changed between study rounds 7 and 8, and again between study rounds 16 and 17 as indicated by vertical dashed red lines - see Supplementary Section D.1.1 for further detail. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

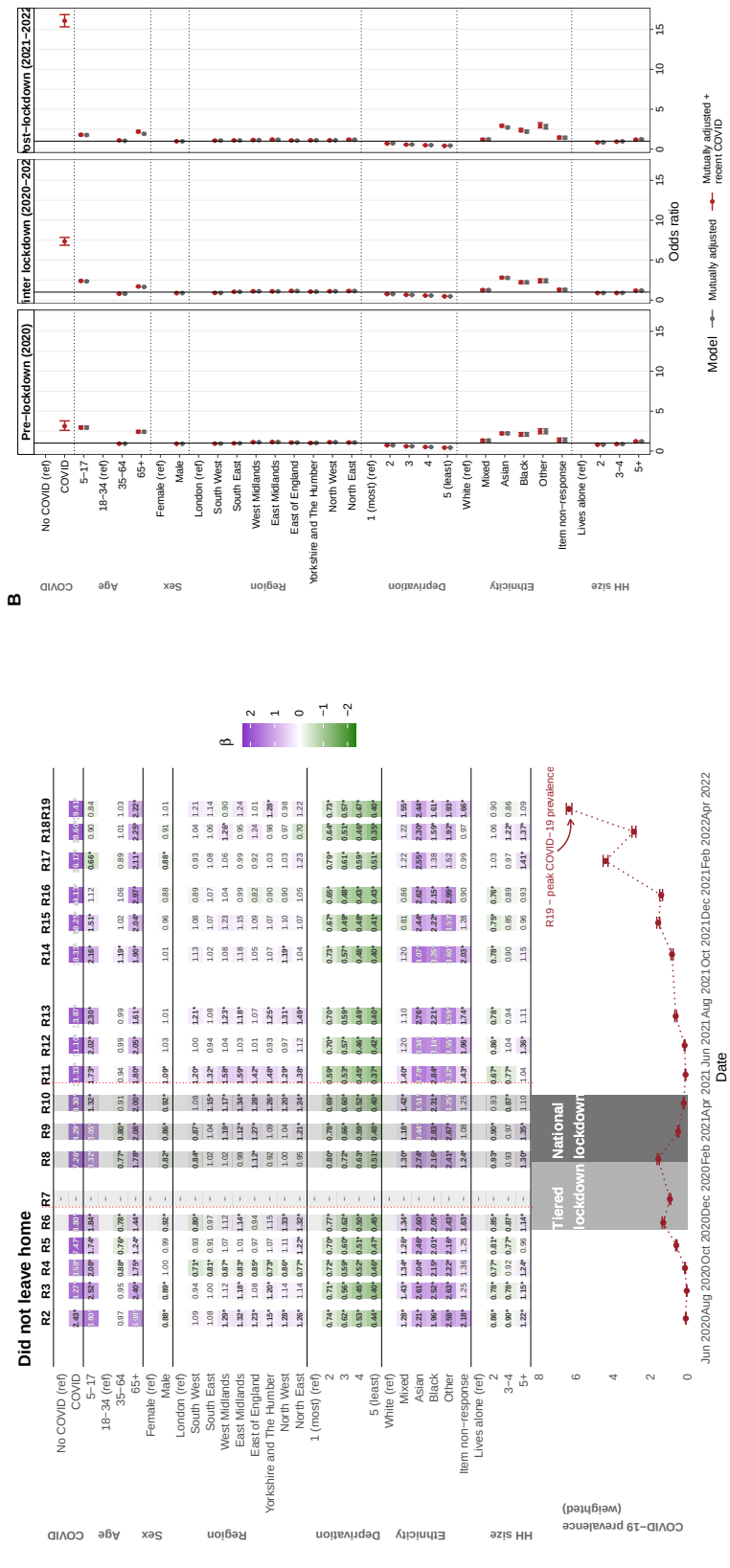

Figure D.29: A. Odds ratios for whether an individual would report not leaving the house in the seven days preceding completing the questionnaire. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

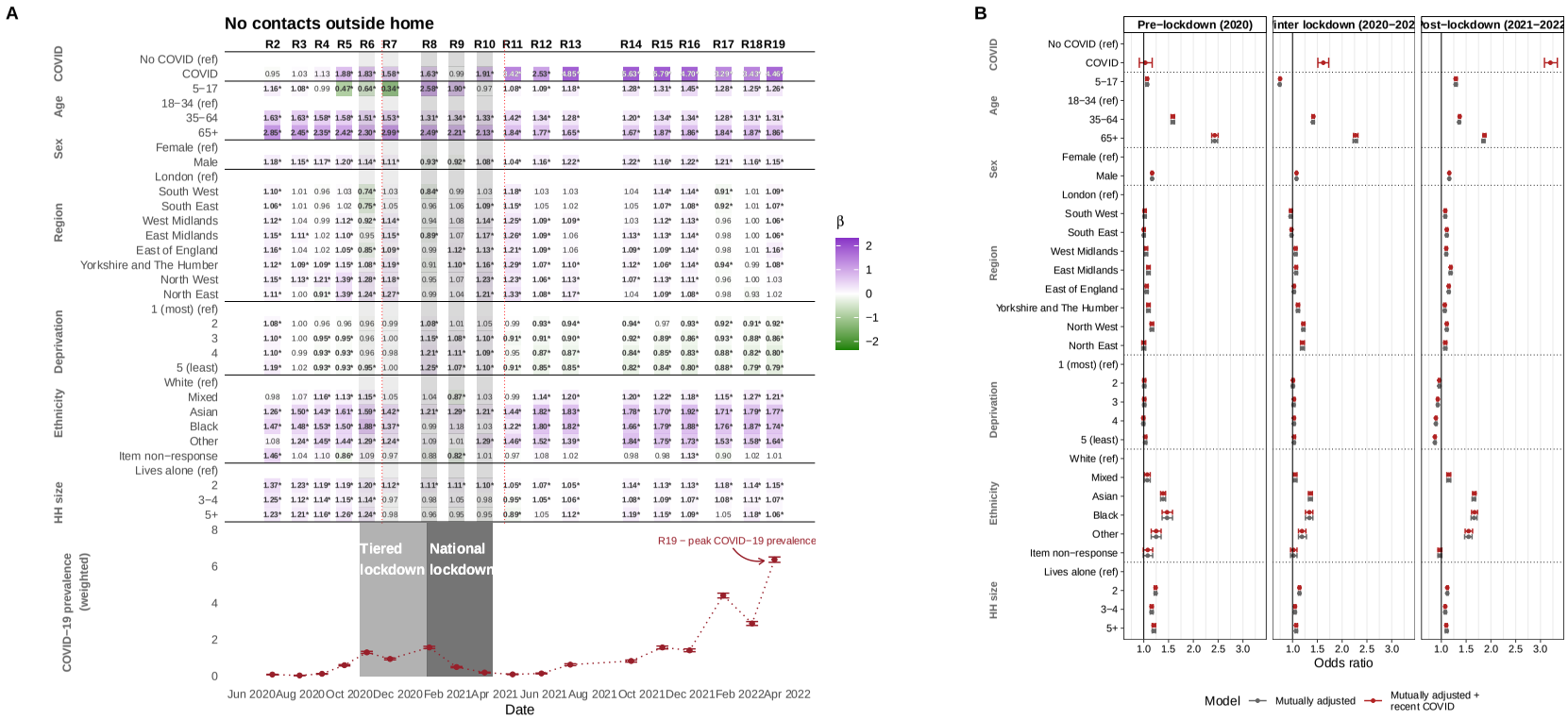

Figure D.30: A. Odds ratios for whether an individual would report having no contacts outside the home on the day preceding the questionnaire. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question wording changed between study rounds 6 and 7, and again between study rounds 11 and 12, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

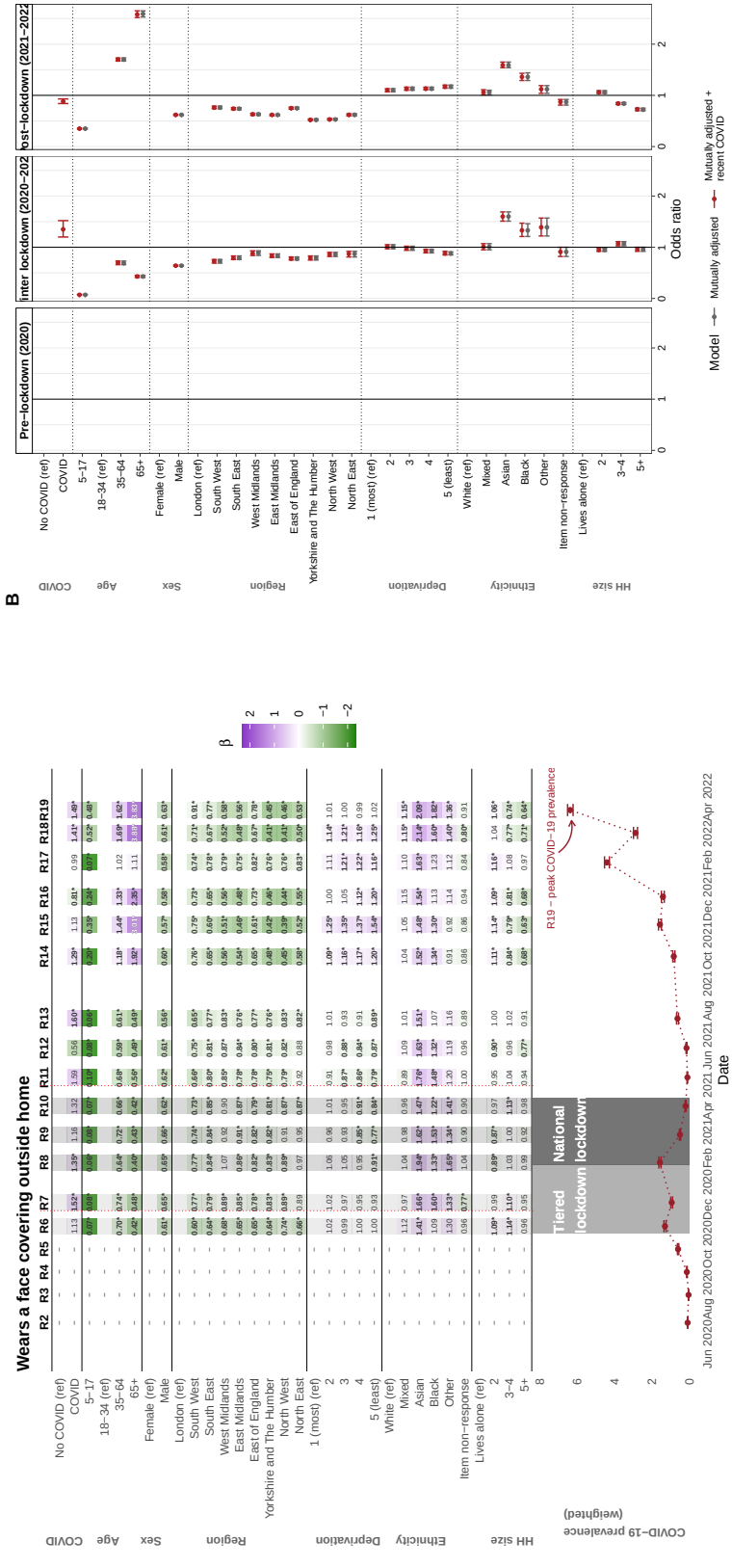

Figure D.31: A. Odds ratios for whether an individual would report wearing a face covering outside the home. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked prior to study round 6. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

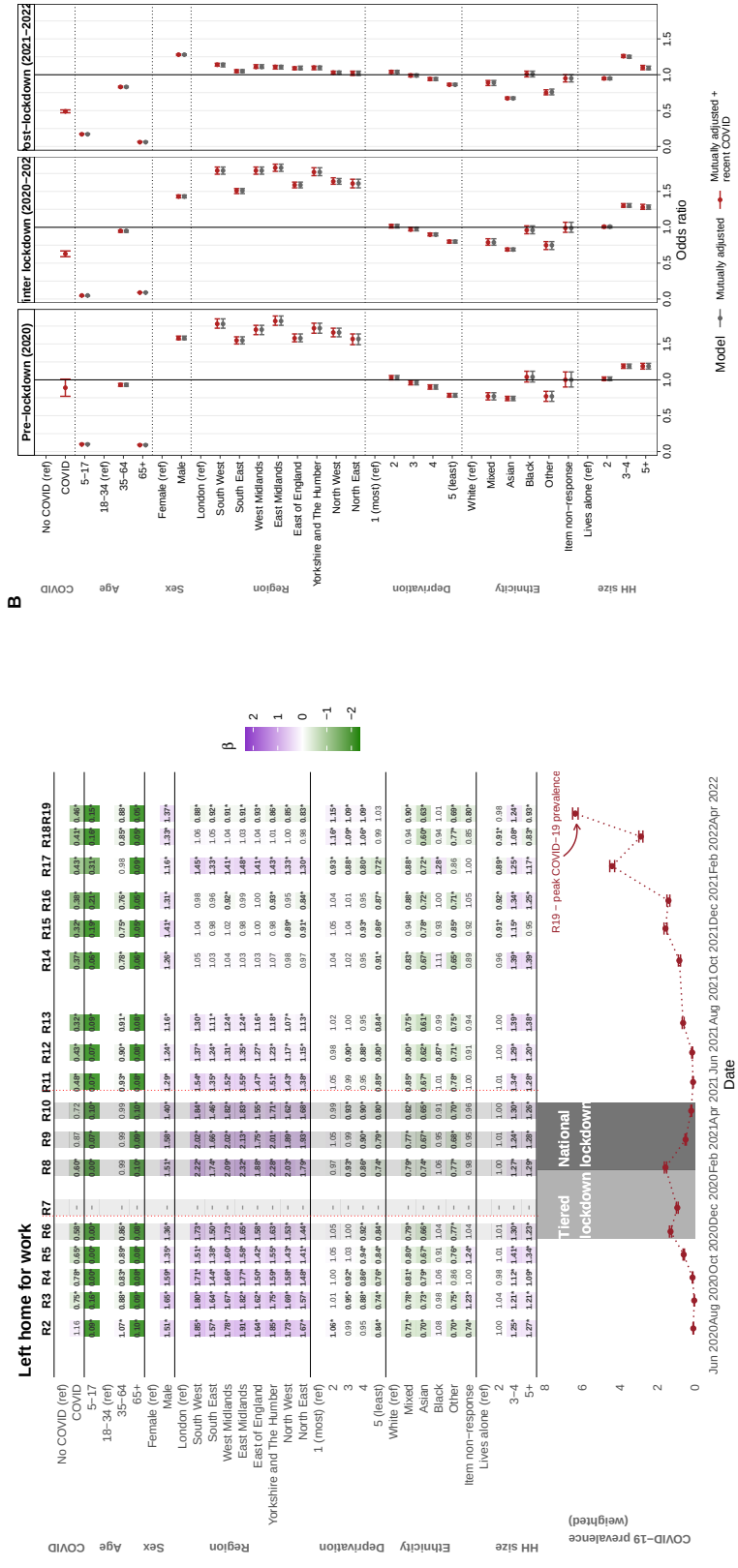

Figure D.32: A. Odds ratios for whether an individual would report leaving home to go to work. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

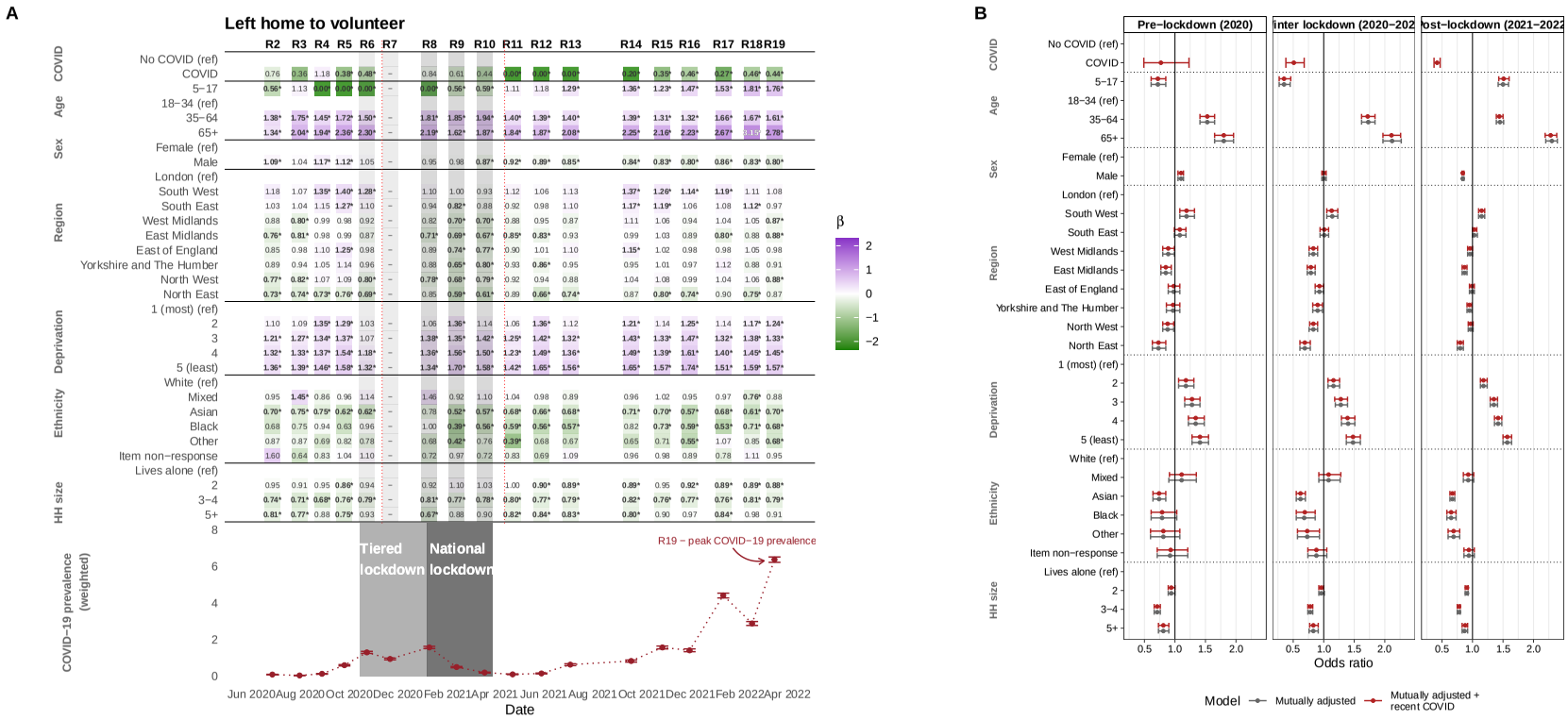

Figure D.33: A. Odds ratios for whether an individual would report leaving home to volunteer. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

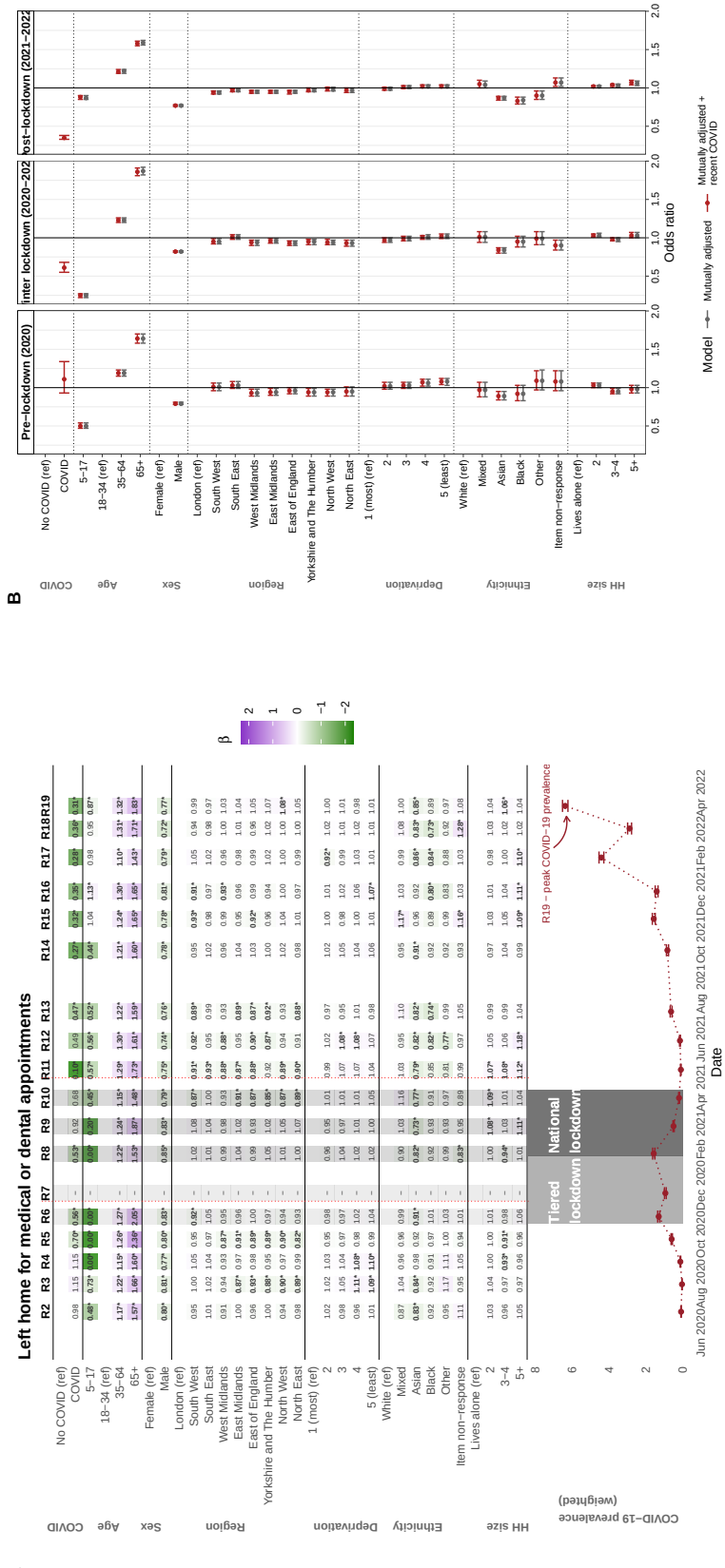

Figure D.34: A. Odds ratios for whether an individual would report leaving home for medical and/or dental care. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

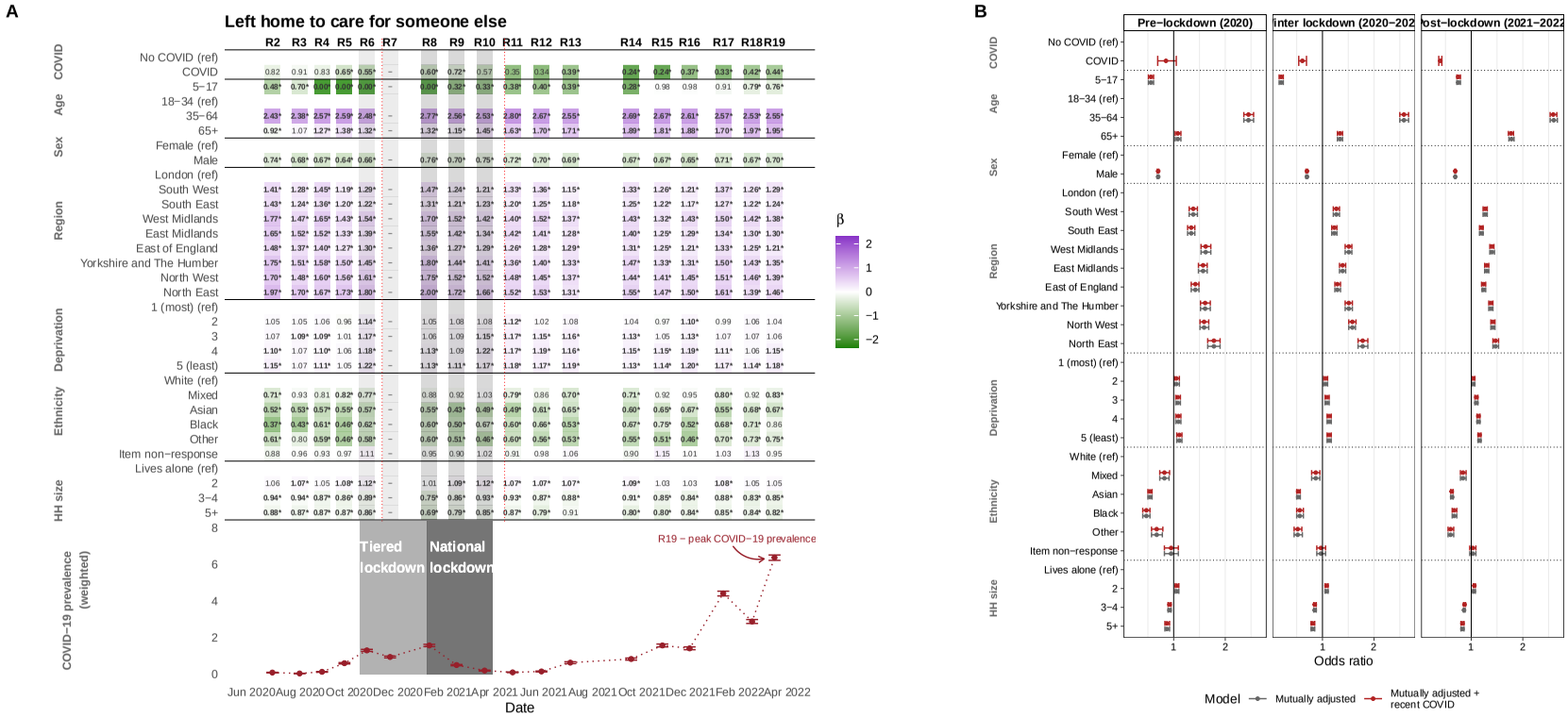

Figure D.35: A. Odds ratios for whether an individual would report leaving home to care for somebody else. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

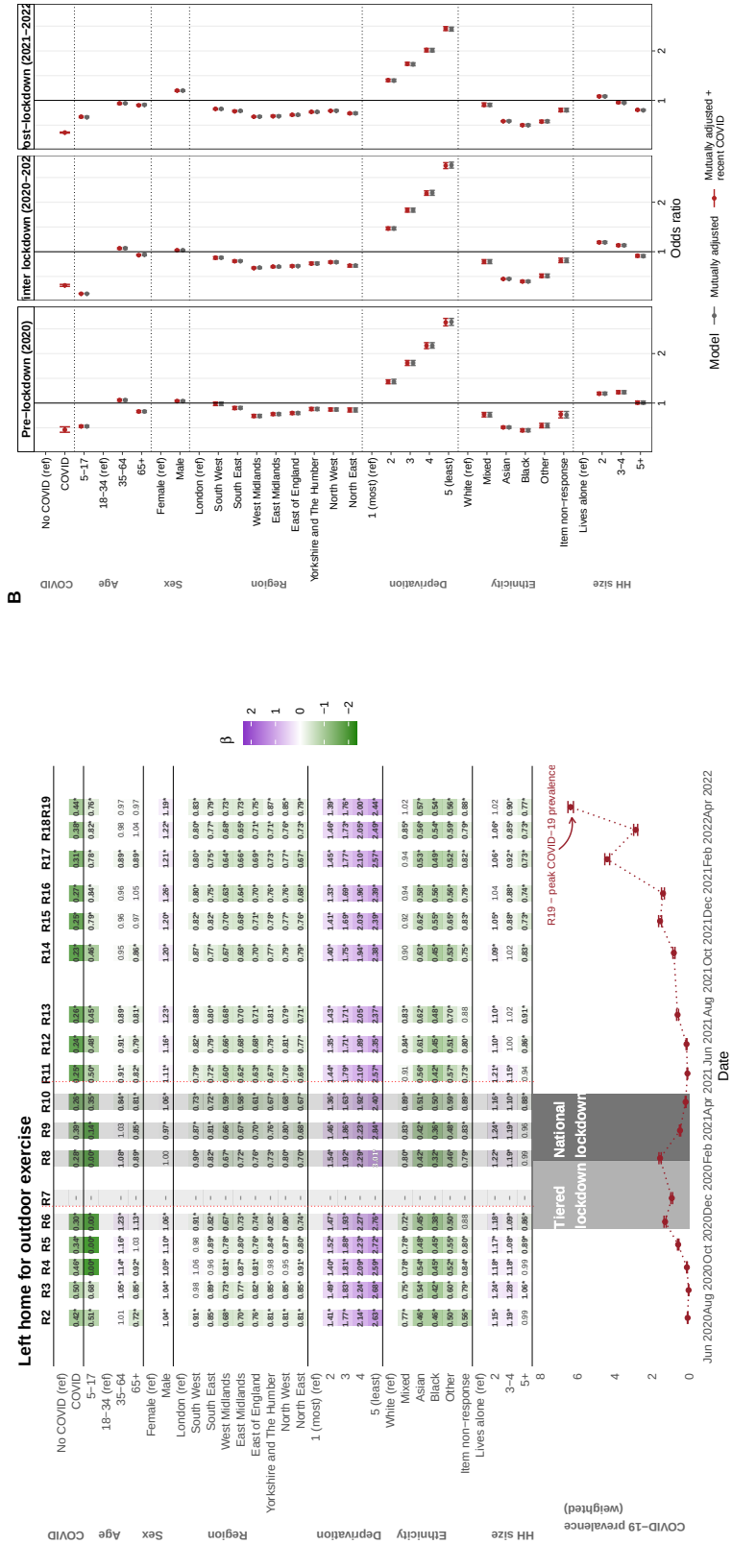

Figure D.36: A. Odds ratios for whether an individual would report leaving home to exercise outdoors. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

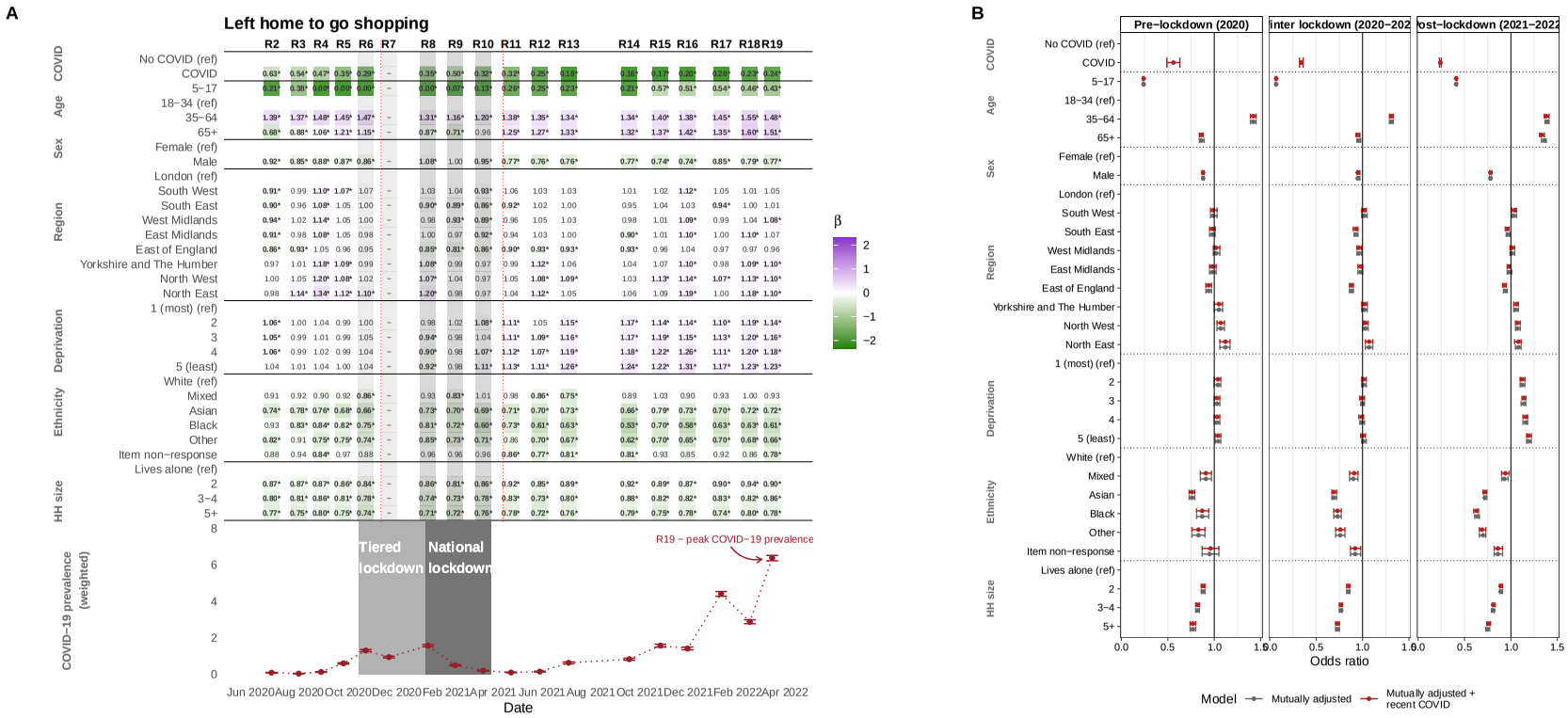

Figure D.37: A. Odds ratios for whether an individual would report leaving home for shopping. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

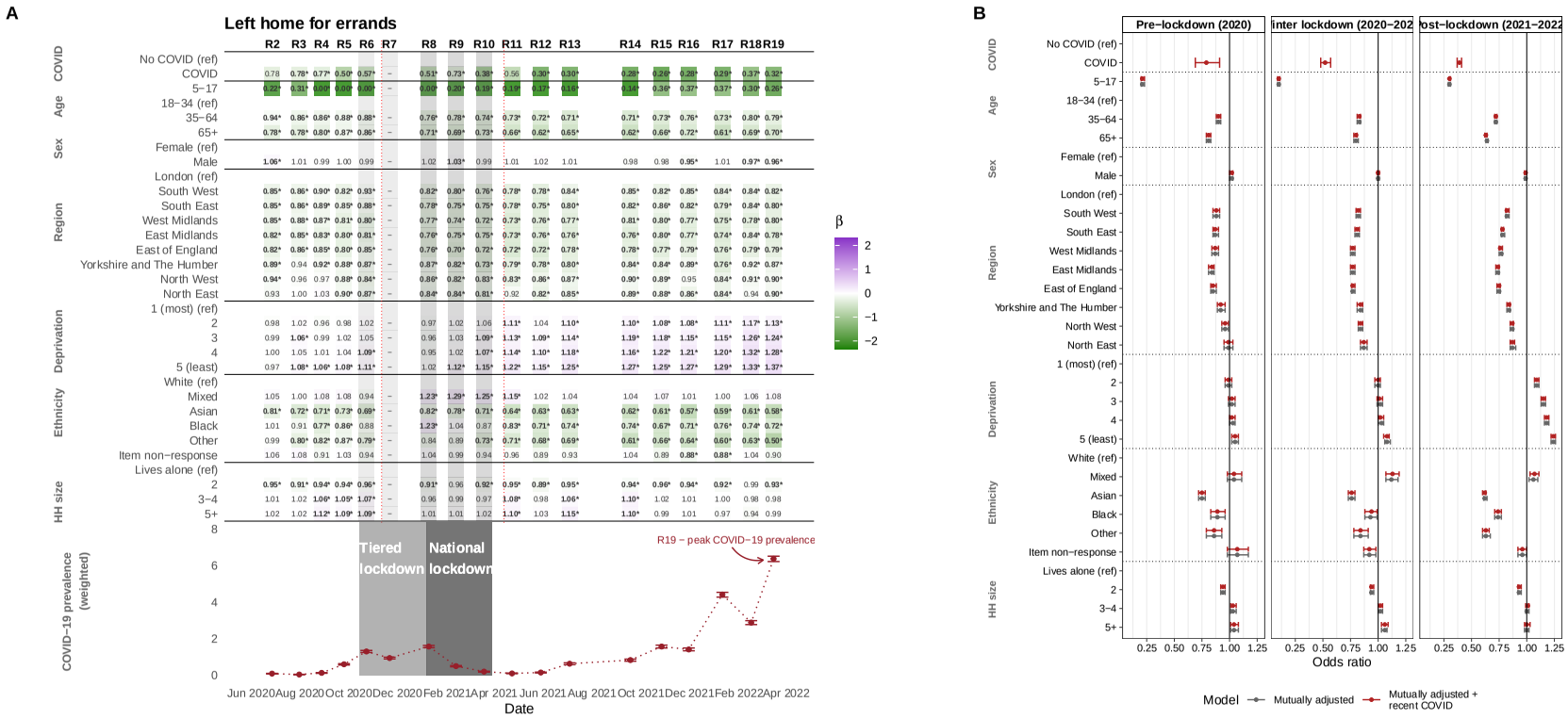

Figure D.38: A. Odds ratios for whether an individual would report leaving home for errands. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

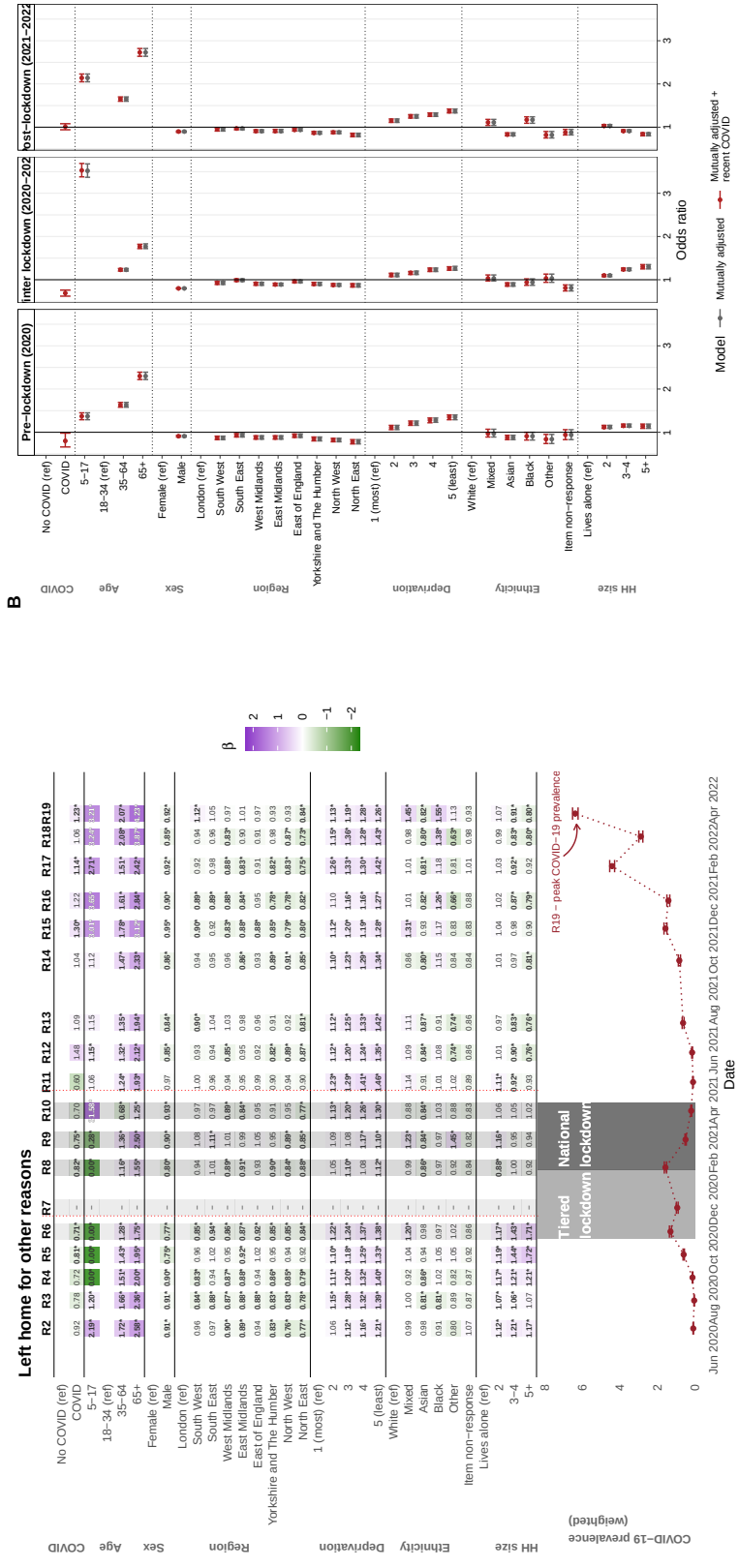

Figure D.39: A. Odds ratios for whether an individual would report leaving home for other reasons. Estimated effects are mutually adjusted for the other variables listed in the table. * indicates the odds ratio is statistically significantly different from 1 at the $\alpha = 0.05$ significance level. This question was not asked in round 7, and the question wording changed between study rounds 2 and 3, and again between study rounds 14 to 15, denoted with a vertical dashed line. The change in question wording is particularly important here, as "other" may include or exclude specific reasons depending on the available options (Supplementary Table D.3). B. Forest plot comparing odds ratios for models with and without adjustment for suspected recent COVID-19.

D.4 Correlations between community-level mobility data and reported behaviour measures

Due to complications that could arise due to multicollinearity and autocorrelation in the behavioural data, here we report pairwise sample correlation coefficients between each of six community-level mobility measures and reported behaviour measures (Supplementary Table D.7). Overall, correlations are weaker in study rounds 15-to-19, despite the random forest model often predicting mobility series just as well.

These results are very reassuring for cross-validation in terms of the magnitude and direction of the correlations. For example, for the mobility measure relating to time spent in residential locations, there are strong negative correlations with shopping, errand and medical appointments, for example, and strong positive correlations with "didn't leave" the home. Similarly, reporting "shopping" as a reason for leaving the home is strongly positively correlated with "grocery" and "retail" mobility measures.

Table D.7: Pairwise sample correlation coefficients between the six community-level mobility measures (Google data) and the proportions in REACT-1 survey rounds regarding not leaving the house or reporting particular reasons for leaving the house in the day before the survey was completed. These correlations were computed separately for study rounds 2-to-6, rounds 8-to-14 and rounds 15-to-19. Positive correlations are shaded blue and negative correlations are shaded red, with stronger shades indicating stronger correlations.

A) Rounds 2 to 6

	Grocery	Retail	Parks	Transit	Workplaces	Residential
Work	0.49	0.30	-0.38	0.23	0.47	-0.32
Volunteering	0.44	0.48	-0.14	0.49	0.68	-0.56
Medical appointment	0.63	0.68	-0.24	0.62	0.87	-0.74
Caring for someone	0.63	0.35	-0.48	0.16	0.51	-0.29
Socialise (in public)	-0.25	0.76	0.69	0.68	-0.04	-0.64
Socialise (in private)	-0.42	0.61	0.82	0.62	-0.23	-0.52
Exercise	-0.43	-0.10	0.56	0.00	-0.44	0.11
Shopping	0.42	0.89	0.02	0.75	0.62	-0.83
Errands	0.34	0.95	0.15	0.83	0.60	-0.90
Other	0.61	0.64	-0.13	0.71	0.87	-0.78
Didn't leave	-0.29	-0.87	-0.30	-0.82	-0.53	0.86

B) Rounds 8 to 14

	Grocery	Retail	Parks	Transit	Workplaces	Residential
Work	0.30	0.34	0.32	0.34	0.26	-0.32
Volunteering	0.84	0.90	0.74	0.91	0.88	-0.90
Medical appointment	0.63	0.61	0.46	0.66	0.73	-0.68
Caring for someone	0.72	0.82	0.74	0.80	0.72	-0.79
Exercise	-0.66	-0.61	-0.52	-0.63	-0.68	0.65
Shopping	0.74	0.84	0.73	0.81	0.69	-0.79
Errands	0.89	0.96	0.82	0.96	0.87	-0.94
Didn't leave	-0.88	-0.78	-0.79	-0.82	-0.81	0.84

C) Rounds 15 to 19

	Grocery	Retail	Parks	Transit	Workplaces	Residential
Work	0.38	0.38	0.16	0.48	0.48	-0.55
Volunteering	0.28	0.29	0.09	0.36	0.26	-0.31
Medical appointment	0.54	0.58	0.21	0.72	0.65	-0.73
Caring for someone	0.38	0.36	-0.02	0.22	0.02	-0.07
Socialising (outside)	0.38	0.45	0.76	0.67	0.32	-0.76
Socialising (inside)	0.38	0.52	0.38	0.68	0.31	-0.69
Exercise	-0.19	-0.15	0.25	-0.07	-0.14	0.05
Shopping	0.40	0.47	-0.10	0.33	0.03	-0.13
Errands	0.34	0.39	0.29	0.60	0.48	-0.67
Other	0.05	0.00	-0.17	-0.18	-0.22	0.31
Walking a dog/pet	0.07	0.11	0.12	0.07	-0.11	0.00
School/university	0.01	-0.03	0.21	0.26	0.46	-0.45
Holiday	0.29	0.43	0.61	0.38	-0.21	-0.30
Haven't left	-0.54	-0.64	-0.14	-0.68	-0.39	0.56

D.5 Stringency and mobility data

We also test the ability of the Google mobility data to model the stringency and containment indices using the same framework that we use to model the stringency and containment indices from REACT data. Random forest regression is used to predict both the stringency and containment indices using the six Google mobility series.

To ensure comparisons with REACT-based models are as similar as possible, we only leverage data on days from which we have REACT data. We also fit two separate models for each index, reflecting the separate models estimated in the main paper. However, this is not an entirely like-for-like comparison, as we use four covariates in the primary analysis (four behavioural questions), whereas here we leverage all six covariates in the mobility series.

We find that estimates derived from REACT data are slightly outperformed by estimates derived from mobility data in study rounds 2-to-5, when there is little overall change in the response. However, estimates derived from REACT data in study rounds 6-to-19 substantially outperform those derived from mobility data.

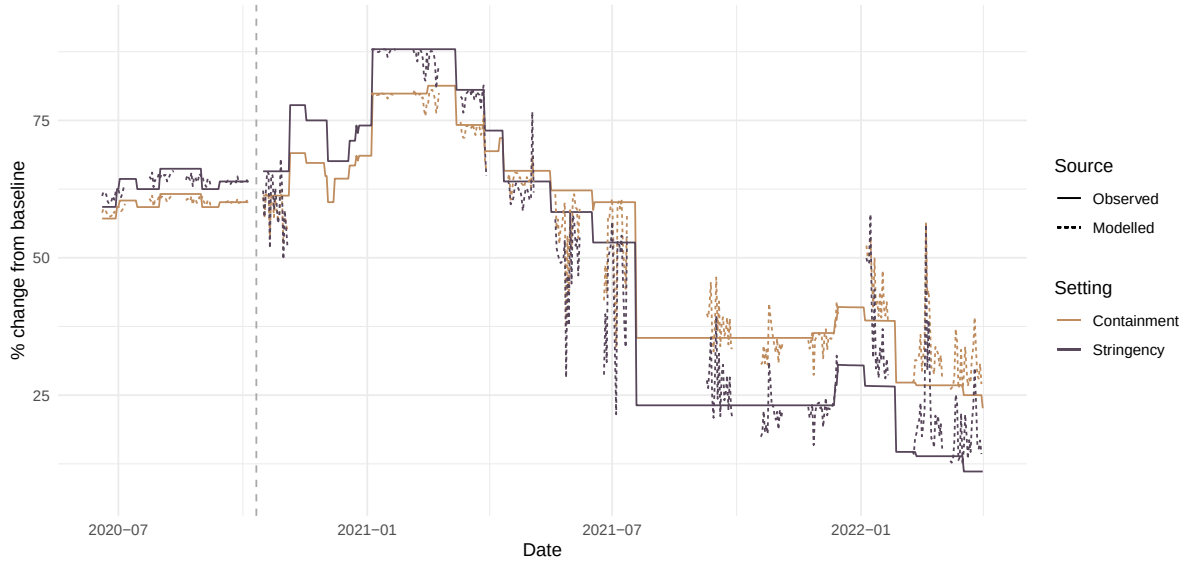

Figure D.40: The OxCGRT stringency and containment and health indices (solid lines) between June 2020 and March 2022 and the model predictions based on Google mobility data. The vertical dashed lines demarcate individual model fits, selected to align with models fit in the primary analysis.

We also present the out-of-bag and within-bag proportion of variance explained for the four models considered in the main paper (three fitting mobility data and one fitting OxCGRT

indices) in Supplementary Tables D.8 and D.9.

Table D.8: Number of days analysed, sample variance and proportion of variance explained (PVE) for each of six mobility measures by the model (analogous to R^2) for out-of-bag prediction (where the model is trained on a subset of the data and performance is evaluated by predicting out of sample), and within-bag prediction.

	Rounds 2-to-6	Rounds 8-to-14	Rounds 15-to-19
Number of days	94	136	105
Grocery			
Sample variance	3.12	130.36	12.55
Out-of-bag PVE	0.73	0.96	0.68
Within-bag PVE	0.95	0.99	0.95
Retail			
Sample variance	132.72	464.09	32.56
Out-of-bag PVE	0.96	0.97	0.81
Within-bag PVE	0.99	0.99	0.97
Parks			
Sample variance	811.22	1363.29	196.1
Out-of-bag PVE	0.85	0.86	0.72
Within-bag PVE	0.97	0.97	0.95
Transit			
Sample variance	23.91	216.54	32.52
Out-of-bag PVE	0.87	0.97	0.86
Within-bag PVE	0.98	1.00	0.98
Workplaces			
Sample variance	26.66	84.69	22.36
Out-of-bag PVE	0.89	0.93	0.76
Within-bag PVE	0.98	0.99	0.95
Residential			
Sample variance	6.46	27.35	4.28
Out-of-bag PVE	0.93	0.97	0.84
Within-bag PVE	0.99	0.99	0.97

Table D.9: Number of days analysed, sample variance and proportion of variance explained (PVE) by the model (analogous to R^2) for out-of-bag prediction (where the model is trained on a subset of the data and performance is evaluated by predicting out of sample), and within-bag prediction.

	Rounds 6-to-19
Number of days	259
Stringency	
Sample variance	760.14
Out-of-bag PVE	0.91
Within-bag PVE	0.98
Containment	
Sample variance	381.75
Out-of-bag PVE	0.90
Within-bag PVE	0.98

D.6 Data portal

To enable future research, we have made aggregated data available in a portal here: https://m-whit-ic.shinyapps.io/react-social-shiny/. This tool allows researchers to access estimates of the proportion of various groups of people in England that report performing various behaviours.

Results by round present survey-weighted estimates of the proportion of England that would report the selected outcome. Confidence intervals reflect uncertainty from extrapolating the survey to the general population. Daily results present the unweighted proportion of participants who reported the selected outcome. Confidence intervals are the 95% binomial proportion interval, calculated using the "exact" method from the Hmisc package in R.

Users can select from a range of outcomes (e.g., shielding, left home for work, reports a previous positive test), grouping (e.g., age-group, ethnicity), and conditioning (e.g., those who report belief of risk of severe illness, a previous positive test). Data are censored such that any single group with fewer than 10 responses, or fewer than 5 true and/or 5 false responses are excluded. This means we will never report proportions of 0% or 100% - any downstream analysis should ensure this is accounted for.

At the time of publication this portal does not contain all possible outcomes, although we plan on adding more over time. If you would benefit from the inclusion of additional outputs, please contact Professor Paul Elliott at Imperial College London.