Mathematical and statistical models of human behaviour in digital epidemiology

Dennis Liu

August 29, 2021

Thesis submitted for the degree of
Doctor of Philosophy
in
Applied Mathematics
at The University of Adelaide
Faculty of Engineering, Computer and Mathematical Sciences
School of Mathematical Sciences

THE UNIVERSITY
of ADELAIDE

Contents

List of Tables

List of Figures

Signed Statement

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the scholarships and support I have received for my research from the Data to Decisions Cooperative Research Centre, Data61, and also through the provision of an Australian Government Research Training Program Scholarship.

Signed: .. .. Date: 29/08/2021

Acknowledgements

This thesis would not have occurred without the generous support of the many kind people in my life.

To my patient and outstanding supervisors Professor Joshua Ross and Associate Professor Lewis Mitchell, I could not have completed this without your guidance and mentorship. Your teaching, support and supervision has been amazing, which often go above and beyond the line of duty.

To the Maths Chats crew of Andrew, Angus, Caitlin, Max and Phill, our in depth discussions of which I barely understood have been invaluable in helping me navigate the difficult and diverse world of mathematics. To George, Wai Keen, Mike and Peter, thank you for paving the way for me and for the guidance through the roller coaster of research.

To all the health workers both in Australia and the world, this research could not have occurred without your tireless contribution.

Finally, to my friends, for standing by me in my most difficult times, and my family, of which I would be nothing without them. And to Nicole, for all of the above and more.

Abstract

Understanding the dynamics of infectious disease spread is crucial to improving public health. Traditional surveillance relies on observations gathered in public health institutions, which can result in delays due to the time taken to collect and publish data and account for only a subset of all individuals infected. Digital data can used to improve models of epidemiology, but come with their own inherent biases that need to be considered and accounted for when modelling. In this thesis we examine how digital datasets can be used to understand and improve models of infectious disease surveillance, using Bayesian statistical methods to infer latent parameters, and stochastic models to forecast observations of disease transmission. Specifically, Chapter 3 examines participation bias in a voluntary digital health surveillance platform to improve on disease prevalence estimates and analyse the observation process. This is further extended in Chapter 4 where we explore the predictors in receiving an influenza test from a primary care physician using the digital health surveillance platform. Traditional data sources contain information on all individuals that receive a test, but excludes those who were ill and were not given or refused a test. We then develop modelling techniques to understand and forecast the spread of SARS-CoV-2 (COVID-19) in all jurisdictions of Australia in Chapter 5. A hierarchical model using social mobility metrics and behavioural surveys is combined with historical case data to link social distancing to estimates of the reproduction number over time in each jurisdiction. The social distancing trends are forecasted, and a branching process model used to simulate COVID-19 over time from the forecasted reproduction number. This branching process model is then adapted to examine the effectiveness of a range of contact tracing strategies in Chapter 6. A case study in using performance metrics to determine health systems under stress is examined. Throughout this thesis we emphasise the importance of using Bayesian methods to understand the uncertainty of our estimates and the importance of uncertainty in decision making, particularly in public health contexts.

Thesis publications

Published Articles

Dennis Liu, Lewis Mitchell, Robert C. Cope, Sandra J. Carlson, and Joshua V. Ross. Elucidating user behaviours in a digital health surveillance system to correct prevalence estimates. *Epidemics*, 33:100404, 2020

Nick Golding, Freya M Shearer, Robert Moss, Peter Dawson, Dennis Liu, Joshua V Ross, Rob Hyndman, Cameron Zachreson, Nic Geard, Jodie Mcvernon, and David J Price. Estimating temporal variation in transmission of SARS-CoV-2 and physical distancing behaviour in Australia. *Doherty Institute*, (July):1–50, 2020

Chapter 1

Introduction

Modelling the spread of infectious diseases is a substantial public health concern. This cannot be more apparent than in the appearance of a novel disease, such as the 2009 H1N1 swine flu pandemic or the 2020 SARS-CoV-2 pandemic. With human respiratory diseases, transmission commonly occurs through contact with infected individuals, and is therefore highly sensitive to human behaviours. Behaviours ranging from hand washing, avoidance of high transmission risk areas, social isolation and quarantining of infected individuals all have significant implications for the rate of disease transmission. By examining how individual or collective behaviour can influence the dynamics of epidemics, models of infectious disease spread can better reflect the actual dynamics, particularly in rapidly evolving scenarios.

Reliable and frequent observations of disease transmission are crucial to understanding the dynamics of infectious disease spread. However, there are many challenges in estimating disease activity and forecasting infectious cases, given that disease transmission is largely unobserved, and progression to symptoms, if at all, may take days to weeks, before finally leading to observation. Traditional surveillance relies on observations gathered in physical public health institutions, such as hospital admissions and notifications of laboratory confirmed cases. This may limit analyses of disease dynamics due to the time taken to collect, collate, confirm and publish, resulting in substantial delay between observation of illness and confirmation [10, 104]. The observations in traditional surveillance also only account for a subset of all individuals infected, which are often those with severe symptoms, health conscious, or with greater access to health care. These differences can significantly alter our understanding of disease transmission [73].

Following the challenges of observing disease transmission, are the challenges in then using those observations to generate models for prediction and forecasting. Epidemic forecasting has been likened to weather prediction in its scale and complexity [83], but unlike the weather, changes in human behaviour, through awareness of disease, knowledge of the forecasts predictions and social distancing interventions, can change the actual outcome. The dynamics of human behaviour, whether influenced through public health policy or individual risk perception, are complex, uncertain phenomena in their own right, without the additional challenge of accounting for this in epidemic models. This uncertainty makes Bayesian methods a natural choice for inferring model parameters, where uncertainty in both the process and the inference can be incorporated in forecasts.

In this thesis we examine how digital data sets can be used to understand and improve models of infectious disease surveillance, using Bayesian statistical methods to infer latent parameters, and stochastic epidemiological models to forecast observations of disease transmission. We highlight the importance of using Bayesian methods to understand the uncertainty of our estimates and the importance of uncertainty in decision making, particularly in public health contexts.

Digital data can used to address challenges in observing disease transmission and to improve epidemiological models, but come with their own inherent biases that need to be considered and accounted for when modelling [73]. After presenting relevant mathematical background in Chapter 2, Chapter 3 of this thesis utilises digital data to improve estimates of influenza prevalence, using Bayesian methods to account for the uncertainty inherent in the data and model. We continue this theme in Chapter 4, where we determine a model for estimating the probability an individual will receive an influenza test and therefore be observed in traditional laboratory surveillance of influenza. Unlike traditional surveillance, digital web surveillance platforms can provide a richer and more complete picture, enabling demographic attributes that are less likely to be associated with a test to be determined.

We tackle the problem of epidemic forecasting in Chapter 5, where we develop a modelling framework to forecast the spread of SARS-CoV-2 in each public health jurisdiction of Australia. While it is the SARS-CoV-2 virus that spreads the COVID-19 disease, for the purposes of simplicity we will refer to both as COVID-19 subsequently throughout this thesis. Using social mobility metrics generated by digital traces of users and individual behaviour surveys, we determine the changes in behaviour in response to public awareness and jurisdictional public health policies. We leverage a hierarchical model to

relate social mobility and adherence to distancing behaviours to the effective reproduction number, while allowing for differences across jurisdictions. We further use information in jurisdictions with more cases to inform parameters for other jurisdictions with less information.

With a predictive model of the effective reproduction number, we forecast the effective reproduction number using trends in social mobility, rather than relying solely on historical case data. We then build a probabilistic forecasting model to predict cases on future dates and dates without complete observations using this estimate of the effective reproduction number. This is achieved using a branching process model, where the offspring distribution has a mean value of the time varying effective reproduction number. The forecasting model is validated using a series of forecasts and compared to a random walk with drift using the continuous ranked probability score (CRPS) and a skill score. The work in this chapter has formed the basis of forecasts which contribute to weekly reports to the Commonwealth Government of Australia since May 2020.

The quarantine or isolation of infectious individuals is one of the oldest forms of disease control [67]. For COVID-19 in Australia, this is largely achieved through the tracing of close contacts (contact tracing) of known infectious individuals, and after identification, mandatory isolation. However, in a dynamic and time critical environment with resource constraints, it is not known which aspects of contact tracing, such as the breadth of time to trace contacts up to (say two or three days prior to symptom onset) or the depth of generations of close contacts (close contacts of close contacts, or simply the close contacts of a case), provide the most value in reducing transmission at different epidemic stages. Chapter 6 extends the branching process model in Chapter 5 and analyses different contact tracing strategies, while evaluating the functionality of targets and metrics of contact tracing used by the Australian Government as a case study. The work in this chapter was conducted under contract and formed the basis of a report to the Commonwealth Government of Australia.

The use of digital data in providing information, particularly on the behaviour of individuals, and its careful utilisation to adjust for its potential biases are consistent themes throughout this thesis. This is all the more significant in a world of global pandemics and burgeoning digital footprints.

Chapter 2

Background

In this thesis, we will focus on utilising digital data to provide insight to social behaviours that are associated with disease transmission and observations. In this chapter, we introduce how behaviours may influence the observations of an epidemic, and how digital epidemiology can be used to understand this. We then introduce the Bayesian inference methods we will employ throughout this thesis to learn the influence of behaviour from digital data, and describe the statistical models that link the data to the outcome of interest. Finally, we describe branching process models that can simulate and forecast outbreaks using parameters inferred from data, and methods for validating the performance of the forecast and models.

2.1 Human behaviour in epidemics

The relationship between human behaviour and epidemics has been noted throughout history. Even in the notes of John Snow, it is stated that the spatial relationship of cases in the Broad Street cholera outbreak was not perfect, due to individual behaviours where brewery workers preferred beer to water, and distant families preferring water from the affected pump to more nearby pumps [107].

There are many other examples of human behaviour affecting the course of disease transmission, such as vaccination choices [34, 102, 71], awareness and risk perception [61, 64, 82, 99], mixing patterns [67, 85], and general spatial mobility patterns [7, 80, 93]. Digital data sources are uniquely positioned to provide insight into such behaviour, through their ubiquity, convenience and

velocity, giving richer, timely and potentially more complete data relative to traditional sources. Examples of the types of human behaviour impacting epidemics examined in other work include mobility mapping [21, 48, 65] and risk perception [14, 33, 96].

A number of studies have used data sources from Google searches, Twitter posts, electronic health records, crowd sourced influenza reporting systems, and even a combination of the above to detect and forecast influenza spread. See for example Nagar *et al.* (2014) [87], Lu *et al.* (2018) [77], and Kalimeri *et al.* (2018) [66]. Despite these attempts in using digital data in epidemiological models, care must be taken to ensure these sources are used appropriately. Most digital data is not intended to produce valid and reliable samples of the population, and can be misrepresentative. The interaction of awareness of the disease, or even contraction of disease, with online behaviour is not well understood, and clearly has implications on epidemiological insights derived from these datasets. These issues are raised and discussed in further detail by Lazer *et al.* (2014) [73], Moran *et al.* (2016) [83] and Lee *et al.* (2016) [74].

In this thesis we will focus primarily on participation in health surveillance and changes in mobility patterns from policy and individual risk perception, and emphasise that it is digital data sources that are uniquely positioned to provide insight into behaviour trends in a convenient and near real-time setting.

2.1.1 Mobility and transmission

The influence of population-level mobility patterns and disease transmission is a key component of understanding the relationship between behaviour and epidemics. Studies examining measles [51], Ebola [80, 32], influenza [43, 46, 55, 68], and COVID-19 [20, 37, 95] have all found mobility patterns to be an important consideration to understanding transmission and outbreaks.

Gog *et al.* (2014) [46] analyse the spatial dynamics of 2009 H1N1 pandemic in the US using digital medical claims data, compiled as weekly numbers of influenza-like illness (ILI) patients stratified by zip code. They find that the best model contained both spatial and temporal (school openings) correlations, but spatial alone resulted in a better fit than school terms alone. Charu *et al.* (2017) [21] extend the work by Gog *et al.* (2014) [46] and explore the drivers of influenza spread in the US during 2002 to 2010, and their analysis also relies on insight from digital medical claims data. They again confirm

the importance of mobility in understanding disease spread.

Due to the significant proportion of COVID-19 cases that are mildly symptomatic or asymptomatic, studies such as those by Zachreson *et al.* (2020) [118], Chang *et al.* (2020) [20] and Price *et al.* (2020) [95] use mobility patterns to provide critical insights into the spread of COVID-19.

2.1.2 Participatory surveillance

Accurate and expeditious surveillance of the transmission of infectious diseases such as influenza and COVID-19 are crucial for public health response and policy decisions. The ubiquity and velocity of news and social media can lead to public awareness well before individual observation of disease. Policy decisions may also be implemented within hours due to the highly connected digital world of communications. Changes in both the quality and quantity of the social contacts required for disease transmission can therefore change much faster than any change in disease prevalence, and well before being observed in traditional health care settings such as hospitals and primary care practices.

Changes in behaviour due to public awareness or in response to policy changes are important considerations that digital systems are uniquely placed to address. This is particularly important where there are a significant number of cases that are mild but still infectious, and therefore not observed in traditional health institutions. Online participatory health surveillance platforms seeks to address these challenges by reaching a wider section of the public, providing near real-time estimates of disease prevalence and behaviours at a finer scale.

FluTracking [28] is an online participatory health surveillance system for monitoring influenza-like-illness (ILI), with one of its principal aims to detect the onset of influenza in Australia. Since 2010, there are over 10,000 participants each year, with participation continuing to increase, particularly during the COVID-19 pandemic in 2020. Similar platforms in the US (Flu Near You [22]) and Europe (Influenzanet [91]) have been successfully used in surveillance and modelling of influenza spread in those regions, since 2011 and 2010 respectively. However, in-depth analysis of reporting behaviour is yet to be conducted, and may have a significant effect in the accurate interpretation of online participatory surveillance data.

A number of analyses of both Influenzanet and Flu Near You [9, 69, 91, 94, 98, 113] examine only 'active users' and exclude other users. Active users are

defined in various ways, but all of these studies simply exclude individuals who do not report regularly enough. In an examination of the determinants of repeated participation [6], low participation was defined as enrolled participants that had not filled in at least two surveys between 30 and 60 days after their first survey (or at least 3 surveys within 2 months of starting). In this definition, it was found that approximately 44% (8,185/18,477) of users in the 2011-12 season were classified as having lower participation, leaving only 56% of participants as 'active users'. In comparison, a study of Flu Near You, defining 'active' users as those who report at least 3 times in a season, found that approximately 35% (10,871/30,737) of users were active [9]. By excluding so many reports and users, a substantial amount of information is lost, and bias in derived estimates is a possibility if there is a relationship between 'active users' and contracting disease.

A number of studies [9, 16, 53] have examined the demographics and representativeness of the participating populations and compared them to national populations, but few have analysed how these demographics or external factors such as school holidays, ILI incidence or occupation affect participation in any given week. In Chapter 3, we will examine these relationships and construct a framework where behaviour can inform our estimates of disease prevalence.

2.1.3 Healthcare-seeking behaviour

The influence of healthcare-seeking behaviours provide insight into the interpretation of population-level incidence of disease, where issues of representativeness and sampling bias must be considered carefully [59]. This is particularly the case where surveillance is time critical and the characteristics of transmission require evaluation for intervention purposes. This is where web-based health surveillance systems may be of greatest use, but are yet to be thoroughly examined for these insights.

The extent to which these behaviours differ among individuals has been examined in earlier studies using surveys [61], as well as interviews and focus groups [29]. Digital data and web-based health surveillance systems provide a rich and convenient source of information to examine similar behaviours [94, 103].

Previous work has shown socioeconomic status [90], media reports [24], and demographic factors [59] can all impact the observations seen in digital data streams. Web health surveillance systems have also been observed to have

participation biases amongst their users [6, 9, 16], as noted in the previous section.

While these insights are critical to understanding digital data streams, there has been little research into using these data streams to provide insight into traditional surveillance systems, such as confirmed influenza cases from testing, and how they may be affected by human behaviour. Some evidence of this has been observed, where healthcare-seeking behaviour has impacted the testing rate of influenza over time [84]. In Chapter 4, we will use a digital data stream to examine the healthcare-seeking behaviour of individuals and whether there is bias among the testing of influenza associated with primary care physicians.

2.2 Bayesian inference and modelling

To relate changes in behaviour to outcomes in epidemiology, we seek to infer the influence of latent parameters of our mathematical models. We can achieve this through Bayesian inference.

We seek to infer the posterior distribution $P(\theta|X, y)$ of a set of parameters θ given some condition X and observations y, which is often the data. There is some *a priori* knowledge of the parameters, which is quantified as the prior distribution $P(\theta)$.

The posterior distribution can then be found using Bayes' rule:

$$P(\theta|X, y) = \frac{P(y|X, \theta)P(\theta)}{P(X, y)}, \tag{2.1}$$

where $P(y|X, \theta)$ is defined as the likelihood of observing y given the parameter, and $P(X, y)$ the evidence. It can also be re-written as:

$$P(X, y) = \int_{\theta} P(X, y|\theta)P(\theta)d\theta, \tag{2.2}$$

which, particularly when θ is high dimensional, may be intractable or difficult to calculate. As such, the exact posterior distribution (2.1) can only be evaluated analytically in a closed form when using specific prior distributions that are conjugate to specific choices of likelihood functions. While this is mathematically elegant, this restricts the choice of prior distributions and

likelihoods to only those which are conjugate, which may not be appropriate to the problem.

As the evidence is not a function of the parameters θ, we can instead have

$$P(\theta|X, y) \propto P(y|X, \theta)P(\theta). \tag{2.3}$$

We can therefore use sampling methods to estimate the posterior distribution without having to evaluate the evidence. We will introduce the sampling method primarily used in this thesis for Bayesian inference, chosen for its appropriateness and flexibility in accommodating these applications.

2.2.1 Hamiltonian Monte Carlo

While there are many other methods for sampling and inference, we have chosen one which avoids forced modelling choices and is computationally efficient. Hamiltonian Monte Carlo (HMC) sampling, sometimes known as Hybrid Monte Carlo sampling, is a specific Markov chain Monte Carlo (MCMC) algorithm that is capable of handling highly complex models [11, 78].

MCMC methods use a Markov chain to produce samples from a stationary distribution that matches the distribution of interest, which is the posterior distribution here. MCMC techniques seek to create a Markov chain that iteratively explores the space of parameter values through an appropriate choice of transition probabilities. After a number of transitions, the distribution of proposed parameter values is unchanged after further transitions, and the Markov chain is said to have converged to the stationary distribution. Once stationary, the Markov chain can be used to produce a large number of samples from the posterior distribution.

However, the time to convergence and autocorrelation of samples are important considerations when choosing the appropriate MCMC method. Simpler methods, such as Metropolis-Hastings [44], may have highly correlated samples in regions of high probability in the parameter space, particularly in higher dimensions. High autocorrelation may also potentially lead to an extended period of time before convergence to the posterior distribution. A choice of Markov chain which allows for large transitions in space can prevent this and may converge to the posterior distribution in fewer samples, but will not be able to adequately sample from regions of high probability density, instead moving away from them too quickly.

In contrast, small transitions are inefficient and will result in a long time to convergence, with samples that are highly autocorrelated, but will be more capable of exploring regions of high density but low volume. Adaptive methods are available, but are often problem specific. HMC attempts to address these challenges by using the geometry of the space of interest to efficiently explore the space of interest using Hamiltonian dynamics.

Conceptually, HMC can be visualised as a frictionless ball sliding over a surface of varying slopes. The ball has both position $\boldsymbol{q}$ and momentum $\boldsymbol{\rho}$ at any point in time, with a corresponding potential energy $U(\boldsymbol{q})$ and kinetic energy $K(\boldsymbol{\rho})$, and as the ball slides across the surface, energy is converted between the two. In HMC, the position of the ball on the surface $\boldsymbol{q}$ is identified with the current state of our parameter of interest θ, and has potential energy analogous to the posterior density, such that $U(\boldsymbol{q}) = -\log(\pi(\boldsymbol{q}))$, where $\pi(\boldsymbol{q})$ is the target posterior distribution $P(\theta|X, y)$. The ball, and thus the sampler, is then drawn towards regions of high density in the posterior distribution like a ball moving downhill.

However, this can then result in the ball trapped in valleys and result in poor exploration of the space. The momentum and kinetic energy of the ball addresses this problem. Given an initial momentum, the ball may then escape regions of low potential energy to continue exploring the surface. The balance of potential and kinetic energy, or the total energy of the system, is described by the Hamiltonian

$$H(\boldsymbol{q}, \boldsymbol{\rho}) = U(\boldsymbol{q}) + K(\boldsymbol{\rho}). \tag{2.4}$$

The joint density $P(\boldsymbol{q}, \boldsymbol{\rho})$ is defined by the Hamiltonian $H(\boldsymbol{q}, \boldsymbol{\rho})$, and can be related to the posterior distribution $\pi(\boldsymbol{q})$ via

$$\begin{aligned} P(\boldsymbol{q}, \boldsymbol{\rho}) &\propto e^{-H(\boldsymbol{q},\boldsymbol{\rho})} \\ &= e^{-U(\boldsymbol{q})-K(\boldsymbol{\rho})} \\ &= e^{-U(\boldsymbol{q})}e^{-K(\boldsymbol{\rho})} \\ &= \pi(\boldsymbol{q})e^{-K(\boldsymbol{\rho})}, \quad \text{as} \quad U(\boldsymbol{q}) = -\log(\pi(\boldsymbol{q})). \end{aligned}$$

If we then define $K(\boldsymbol{\rho}) = -\log P(\boldsymbol{\rho})$, where $P(\boldsymbol{\rho})$ is the distribution of the momentum $\boldsymbol{\rho}$, then we have

$$P(\boldsymbol{q}, \boldsymbol{\rho}) \propto \pi(\boldsymbol{q})P(\boldsymbol{\rho}). \tag{2.5}$$

$P(\boldsymbol{\rho})$ will depend on the implementation but is typically a multivariate normal distribution $\mathcal{N}(0, \Sigma)$. Defined in this way, sampling from the joint parameter space $P(\boldsymbol{q}, \boldsymbol{\rho})$ also samples from the posterior distribution $\pi(\boldsymbol{q})$ by projecting back onto the original target space.

The Hamiltonian dynamics describe the change in both $\boldsymbol{q}$ and $\boldsymbol{\rho}$ through

$$\frac{d\boldsymbol{q}}{dt} = \frac{\partial H}{\partial \boldsymbol{\rho}} \tag{2.6}$$

$$\frac{d\boldsymbol{\rho}}{dt} = -\frac{\partial H}{\partial \boldsymbol{q}}. \tag{2.7}$$

The dynamics can be simplified using Equation (2.4),

$$\frac{d\boldsymbol{q}}{dt} = \frac{\partial K}{\partial \boldsymbol{\rho}} \tag{2.8}$$

$$\frac{d\boldsymbol{\rho}}{dt} = -\frac{\partial U}{\partial \boldsymbol{q}}. \tag{2.9}$$

Given a position and a momentum, and therefore a total energy, we can then use the Hamiltonian dynamics to map a trajectory that explores the parameter space while being constrained by the probability density of the posterior distribution. However, by using Hamiltonian dynamics, HMC can only be implemented on continuous random variables due to the algorithm's reliance on derivatives. Betancourt *et al.* (2017) [11] provides further detailed discussion on HMC and its ability to incorporate the geometry of the space.

By utilising the geometry of the parameter space, HMC reduces autocorrelation between samples by rapidly exploring the space, and is able to adequately sample from regions of high probability density without being trapped. While calculations of the Hamiltonian dynamics are expensive, it efficiently samples complex parameter spaces [78], particularly with multilevel models with high numbers of parameters, as used in this thesis. The Stan package [109] is robust and flexible implementation of HMC, and in Chapters 3, 4 and 5 we will use the PyStan implementation in Python.

2.2.2 Logistic regression

A logistic function (2.10) is a suitable model for a binary outcome, such as whether an event occurred, and provides a probability of the modelled event

occurring. As we will frequently encounter such outcomes in this thesis, in this section we will describe how logistic functions are used within our statistical framework. Using a set of predictors $\boldsymbol{x}$, it can link the predictors $\boldsymbol{x}$ to the response y, with the influence of each predictor on the response quantified by the corresponding element of $\boldsymbol{\theta}$.

$$y = \frac{C}{1 + e^{-\boldsymbol{\theta}^T \boldsymbol{x}}}. \tag{2.10}$$

This method is usually termed logistic regression when used to model a probability p of an outcome y occurring given information or predictors $\boldsymbol{x}$, and $C = 1$ to appropriately bound $p \in (0, 1)$. Equation 2.10 can be manipulated to provide an alternative form:

$$\log\left(\frac{p}{1-p}\right) = \boldsymbol{\theta}^T \boldsymbol{x}. \tag{2.11}$$

From Equation 2.11, we can interpret the value of θ_i as the relative increase in the log odds of y occurring for a unit increase in x_i, given the other $x_j \quad \forall j \neq i$. For example, if $\theta_i = 1.099$, then for a unit increase in x_i and no other change in the other x_j for $j \neq i$, the relative odds of y occurring increase by $e^{1.099} \approx 3$, or y is 3 times more likely to occur. Please see Jewell (2003) [63] for further discussion on the interpretation of the log odds. To infer the value of the parameters $\boldsymbol{\theta}$ we can apply the techniques discussed in Section 2.2.

As we will see in Chapter 5, the logistic function is also a useful model for a response variable that is bounded and is expected to asymptotically approach those bounds, and where diminishing returns is expected. This property will prove useful in modelling the reduction in transmission potential as behaviour changes.

2.2.3 Hierarchical models

A common modelling consideration is how to account for similarities within observations that may be naturally grouped. This may occur when there are repeated, but a differing number of, samples across groups. We will encounter such a model in Chapter 5. If each group is believed to be identical, we can ignore the grouping and assume that each observation is independent and is described as complete pooling of observations. If instead we believe

each group has no information about the other groups, and each group is independent of each other, than this is termed no pooling.

However, what if we would like to use observations from one group to inform our knowledge of another group, and we expect there to be some similarities across groups? This is where hierarchical models (sometimes called multilevel models) can be used to partially pool the observations. Groups where there are few observations can then be informed by groups with a greater number of observations, without the larger groups overwhelming the inference of parameters of the smaller groups.

Bayesian models provide a natural framework for hierarchical models. In the no pooling paradigm, when given a set of parameters $\boldsymbol{\theta}_i$ for each group i, the assumed prior distribution $g(\boldsymbol{a})$ for some given constants $\boldsymbol{a}$ can be applied independently to each $\boldsymbol{\theta}_i$, and observations $\boldsymbol{y}_i$ in each i only affect the corresponding $\boldsymbol{\theta}_i$.

If instead we would like to partially pool our observations and construct a hierarchical model, we can instead use hyperparameters $\boldsymbol{\theta}$, such that

$$\boldsymbol{\theta_i} \sim g(\boldsymbol{\theta}, \boldsymbol{\sigma}), \tag{2.12}$$

where the hyperparameter $\boldsymbol{\theta}$ represents the location of the distribution of all $\boldsymbol{\theta}_i$, and $\boldsymbol{\sigma}$ the scale. We can then learn the values of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$ jointly with our inference of $\boldsymbol{\theta}_i$. The hyperparameters are then given hyperpriors,

$$\begin{aligned} \boldsymbol{\theta} &\sim h_1(\boldsymbol{a}) \\ \boldsymbol{\sigma} &\sim h_2(\boldsymbol{b}), \end{aligned}$$

for some appropriate choice of distributions $h_1(.)$ and $h_2(.)$, and constants $\boldsymbol{a}$ and $\boldsymbol{b}$. If the groups are similar, the inference will learn that the scale hyperparameter $\boldsymbol{\sigma}$ will be small, and the posterior distribution of $\boldsymbol{\theta}$ will be similar to the posterior distribution of $\boldsymbol{\theta}_i$.

In Chapter 5 we will see the advantages of using a hierarchical model to infer parameters of groups with fewer observations, and how this can provide a suitable model for inference across different geographical regions and public health jurisdictions.

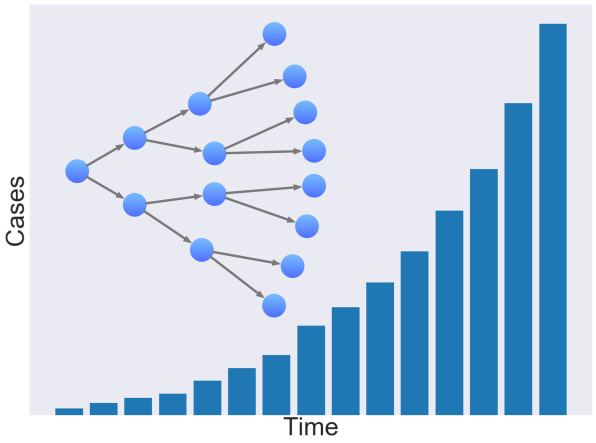

Figure 2.1: A simple branching process, where each parent case generates two offspring, and the number of new cases plotted over time.

2.3 Branching process models

A branching process is a stochastic process that is often used to model the growth of discrete random variables, often individual agents. Branching process models are an intuitive framework for modelling the growth of an epidemic from a small number of initial cases, and they will be used extensively in Chapters 5 and 6 for modelling COVID-19 cases in Australia.

Let χ_n denote the number of new offspring at generation $n \in \mathbb{N}$. As each new offspring must be generated from a previous case, we have

$$\chi_n = \sum_{i=1}^{\chi_{n-1}} S_{n,i} \tag{2.13}$$

where $S_{n,i}$ is the number of offspring generated by individual i in generation n. Figure 2.1 graphically depicts the growth of the generations and the number of new offspring χ_n over time.

We can then find the total number of individuals I_n up to generation n by calculating $I_n = \sum_{i=1}^{n} \chi_i$. However, to model an epidemic, we may be interested in how the number of offspring, or new cases, $\chi(t)$ evolves in discrete time t, rather than in generation n. Evolution over discrete time can be

more intuitive for informing policy in the time dependent nature of outbreak response.

Equation 2.13 can be rewritten to vary in time t by

$$\chi(t) = \sum_{i=1}^{I(t)} N_i(t) \tag{2.14}$$

where $I(t)$ is now the total number of cases at time t, and $N_i(t)$ the number of offspring generated by individual i at time t. This form is a more intuitive notation for modelling and forecasting an epidemic through time.

2.3.1 Reproduction number

In order to examine the distribution of offspring in a branching process that models an epidemic, we must first introduce the basic reproduction number R_0. The basic reproduction number R_0 is defined as the average number of offspring, or secondary cases, arising from an average primary case in a completely susceptible and well mixed population [31, 67]. When $R_0 < 1$, then then each generation will by expectation have fewer individuals than the previous, and the epidemic will die out almost surely.

The spread of an infection will vary over time, whether due to the depletion of susceptible individuals in the population or through changes in behaviour. The parameter $R_{\text{eff}}(t)$ is defined as the time-varying effective reproduction number, which generalises the basic reproduction number R_0 to also represent the average number of secondary infections of an index case infected at time t.

If the number of offspring $S_i(t)$ of an individual is assumed to be Poisson distributed given a mean $R_{\text{eff}}(t)$, which is itself a random variable that is Gamma distributed with shape parameter k and scale parameter θ, then the marginal distribution of $S_i(t)$ is the Negative Binomial distribution,

$$S_i(t; k, \theta) \sim \text{NegBin}\left(k, \frac{1}{1+\theta}\right), \tag{2.15}$$

which has mean $\frac{k}{\theta}$. We would like to set the mean of the distribution of the number of offspring to $R_{\text{eff}}(t)$. We then require

$$\frac{k}{\theta} = R_{\text{eff}}(t)$$
$$\implies \theta = \frac{k}{R_{\text{eff}}(t)}.$$

Finally, this gives

$$S_i(t; k, \theta) \sim \text{NegBin}\left(k, \frac{1}{1 + \frac{k}{R_{\text{eff}}(t)}}\right) \tag{2.16}$$
$$= \text{NegBin}\left(k, \frac{R_{\text{eff}}(t)}{R_{\text{eff}}(t) + k}\right). \tag{2.17}$$

The parameter k is a measure of the variance in the offspring distribution, and is referred to as the dispersion parameter. For further discussion and analysis of models of offspring distributions and dispersion, please see Lloyd-Smith *et al.* (2005) [76].

2.3.2 Application in continuous time

To simulate an epidemic in real time, we must define the distribution of time between the parent and generation of the offspring, or the infector to the infectee in our context. This is referred to as the generation interval. This distribution has important ramifications for estimation of the reproduction number in near real-time. When infector-infectee pairs are unknown, as is often the case in active outbreaks, inference performed jointly on both parameters will yield correlations between samples [42, 52, 115].

Another important distribution is the time from infection to onset of symptoms. This is referred to as the incubation period. The onset of symptoms is often where observation of a case becomes possible, as large scale testing of asymptomatic and pre-symptomatic individuals can be impractical. Combining all these components together, we may simulate an epidemic using a branching process, as detailed in Algorithm 1 for a simple epidemic example.

input: χ_0 initial cases and their infection times g_i
$n = 0$;
while $\chi_n > 0$ **do**
  $\chi_{n+1} = 0$;
  for $i = 1$ **to** χ_n **do**
    using $t = g_i$, calculate reproduction number $R_{\text{eff}}(t)$;
    sample number of offspring N_i from $S_i(t; k, \theta)$;
    $\chi_{n+1} = \chi_{n+1} + N_i$;
    for $j = 1$ **to** N_i **do**
      sample g_j from generation interval;
      sample h_j from incubation period;
    end
  end
  $n \leftarrow n + 1$;
end

Algorithm 1: A simple epidemic modelled using a branching process in continuous time.

If the generation interval is often shorter than the incubation period, as with COVID-19 [38, 72, 111], then control of the outbreak becomes difficult without social distancing measures and changes in individual behaviour. The effect and compliance of social distancing measures is difficult to observe using traditional public health datasets. However, digital sources can provide this insight in near real-time, as we have discussed earlier in this chapter.

Here we have introduced to the reader some background into the importance of behaviour in understanding epidemics, and how digital data sources can be used to examine this, as well an understanding of the methods used to infer and model the relationship between behaviour and disease transmission. We have also introduced a simple epidemiological model, the branching process, and some of the key parameters involved. More details will be provided in each chapter, as well as references for further reading.

Chapter 3

Elucidating user behaviours in a digital health surveillance system to correct prevalence estimates

This chapter contains contents from the publication:

Dennis Liu, Lewis Mitchell, Robert C. Cope, Sandra J. Carlson, and Joshua V. Ross. Elucidating user behaviours in a digital health surveillance system to correct prevalence estimates. *Epidemics*, 33:100404, 2020

Please see the Appendix for the full authorship statements.

Influenza is a substantial public health concern, with approximately 3-5 million severe cases worldwide each year [117]. There are many challenges in estimating influenza activity and forecasting the spread of influenza, given that disease transmission in the general population is largely unobserved. Traditional influenza surveillance relies on public health system monitoring, such as through hospital admissions, notifications of laboratory confirmed cases, or voluntary reporting from local physicians [4, 23]. However, in these systems, estimates of disease prevalence can be limited by the time taken to collect, collate and publish through public health systems, resulting in delays between infection and observation.

Online participatory health surveillance systems attempt to address these challenges by providing a convenient, simple and near real-time platform for

self-reporting of symptoms. FluTracking [28] is one such system for monitoring influenza-like-illness (ILI), with a principal aim to contribute to community level ILI surveillance in Australia and New Zealand. Recall that similar platforms exist in the US (Flu Near You [22]) and Europe (Influenzanet [91]).

Online influenza surveillance platforms often publish estimates of the incidence or prevalence of ILI in the population, usually derived as a proportion of the total number of reports received that week. These estimates have been found to correlate well with clinical surveillance by public health bodies [8, 98, 113], including across different definitions of ILI cases [91]. However, user behaviour is not adjusted for other than by excluding 'inconsistent' users, and potentially information is lost on the interaction of behaviour with the prevalence of disease.

The work in this chapter addresses this interaction of behaviour with the prevalence of disease by using information about how individuals behave with regard to survey completion to inform our estimate of the weekly prevalence of ILI. While the scope of this work is limited to Australia, the analysis is generalisable to other types of voluntary, web-based surveillance in other locations and settings, of which there are many.

3.1 Background

While systems such as FluTracking show promise in estimating the incidence of ILI in the population, there is evidence that these systems can be affected by variations in user participation [8, 26, 54, 97]. Very little is known about how these biases could affect disease prevalence estimates and there is a clear lack of studies that attempt to quantitatively adjust for them. Attempts to correct for these biases have all been based on the removal of data, such as only considering users who participate frequently [16, 69, 91, 113], removing the first report of a user [54], or by creating noise filtering algorithms that minimise sudden departures from sentinel data due to changes in participation [106]. While not unreasonable, these methods do not give insight into user behaviour, and do not examine the effect of their corrections on the estimates they examine.

Some examinations of the heterogeneity of users and their behaviours have been conducted, such as inferring significant predictors for and classifying users on their participation levels [6, 9], and examining the demographic representativeness of the participating population [9, 16, 53]. However, user

behaviour has not been analysed within a systematic, statistical modelling framework, in particular one which can simultaneously correct disease prevalence estimates.

3.1.1 FluTracking data

Participants can register in FluTracking at any stage of the year, and upon registration, they provide various demographic details about themselves, such as age, sex, and postcode. Participants can optionally register and report on behalf of others in their household. Those participants who are submitting reports, on behalf of themselves and/or their household, will henceforth be referred to as 'masters', while any individual observed, master or not, will continue to be referred to generally as 'participants'. After registering, masters are sent an email on the Monday of each week during the influenza season to respond to an online questionnaire about the presence of fever or cough that they or their household members may have experienced in the week prior, and whether they have received the influenza vaccine this year.

Surveys can be submitted for up to 5 weeks from their first reminder. Note that FluTracking publishes estimates of ILI incidence to examine the spread of new cases of ILI [17], where incidence is the number of *new* cases per unit of time. However, as we wish to examine the influence of symptoms on behaviour, in this study we will focus on estimates of prevalence in order to determine if user behaviour is influenced by ILI status, where prevalence is the number of cases per unit of time.

If a participant reports both a fever and a cough, follow up questions are revealed, such as enquiring about any health-seeking behaviour taken or if a sore throat was experienced. In this study, we define an ILI case as a survey response with a participant reporting both a fever and a cough, which closely resembles the World Health Organization surveillance case definition [116].

Note that a report, and therefore an observation of ILI status in a household, is generated by the action of the master, and so it is the report of the master and subsequent reports on behalf of the rest of the household that is the outcome of interest. Also note that in order to build a framework for near real-time prediction, we have chosen to define a report submitted less than 7 days after the initial request as an 'on-time' report and otherwise label it as a 'late' report. This is the interval of time before the subsequent request for the next survey is sent. However, if symptoms are submitted in a late report, this information is not excluded, but used in the derivation of predictor variables

for subsequent survey weeks, irrespective of submission date.

FluTracking operates between May and October every year in the winter season in the Southern Hemisphere[1]. We examine all Australian reports submitted between 2011 and 2017, totalling 3,459,339 unique reports from 30,564 households and 52,773 participants. Of these reports, 352,287 (approx. 10%) are late reports.

Using the registration date of households, we can determine the weeks in which a household was registered but never submitted a report. Survey weeks before the registration date are therefore not included as missed reports. This includes another 496,175 missed surveys.

3.1.2 The model

The conventional estimate of prevalence in the literature is given by the total number of individuals that have reported on-time with ILI, divided by the number of on-time reports in the given week. We improve this by developing a framework to adjust ILI prevalence estimates by correcting for user behaviour, and construct a model to predict the probability of a user reporting in a given week, based on their prior behaviour and demography (Figure 3.1). An individual household i in a given week w will receive a survey request to participate, and will then proceed to either report and provide information on their symptoms, or not report. Given that they report, and their symptoms are therefore known, they will either report on-time, or report late. We will compare the following two estimates of the prevalence of ILI:

- the naïve estimate (an extension of the conventional estimate); and,
- a behaviour corrected estimate (our new framework).

Our naïve estimate extends the conventional estimate by considering a Bayesian perspective, which respects the same mode value as the conventional estimate. The behaviour corrected estimate is inferred from a framework that incorporates an observation process in the model, whereas the naïve model does not. Both estimates assume the number of individuals with ILI in the population is binomially distributed.

[1]This has been expanded in 2020 to assist in surveillance of COVID-19, with the addition of more survey questions.

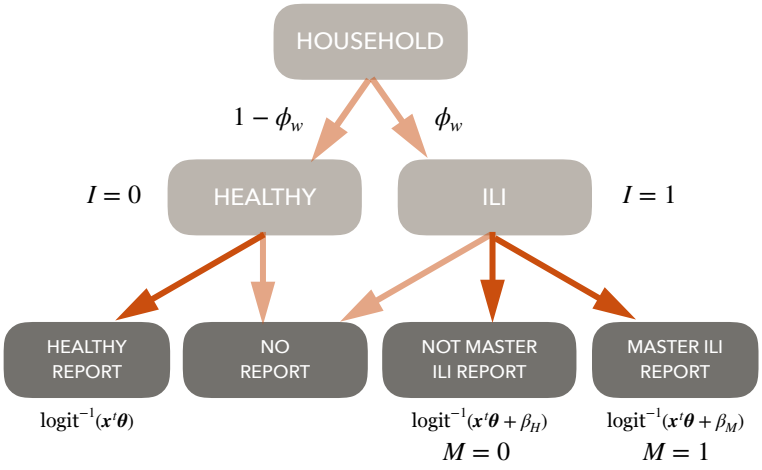

Figure 3.1: Visualisation of model of user outcomes for a household i in week w, where the probability of reporting given their health status is modelled with the link function listed below each respective compartment. The probability of having at least one household member with ILI is modelled with ϕ_w (Equation 3.2).

3.2 Materials and Methods

3.2.1 Naïve model

The conventional estimate of prevalence takes the total number of individuals that have reported on-time with ILI this week X_w and divides this by the number of on-time reports this week $\mathcal{N}_w$. If we consider that each individual in the population has an equal probability $\hat{\pi}_w$ of having ILI in week w, then X_w can be modelled with a binomial likelihood with probability $\hat{\pi}_w$ and trials $\mathcal{N}_w$. Note that this interpretation considers $\mathcal{N}_w$ as fixed and not a random variable. Given a uniform prior (Beta$(1,1)$) and a binomial likelihood, the posterior distribution of $\hat{\pi}_w$ is then Beta$(X_w+1, \mathcal{N}_w - X_w + 1)$. This results in the mode of the posterior distribution of $\hat{\pi}_w$ as $X_w/\mathcal{N}_w$, which is the conventional point estimate often used in the literature (See [106, 91, 113, 66]). We define this posterior distribution of $\hat{\pi}_w$ as the naïve estimate, which is an estimate for the prevalence of ILI in the population, without consideration of user behaviours.

In the work presented here, we did not split the cohort by vaccination status

and model estimates for vaccinated and unvaccinated prevalence. Models using a split cohort with separate parameters for vaccinated and unvaccinated prevalence were examined, with only marginal differences in the corresponding parameters, typically near the peak ILI week. As such, we have simplified the model and reduced the number of parameters inferred. While it could be tempting to examine the difference between such vaccinated and unvaccinated estimates as a measure of influenza vaccine effectiveness, this study examines cases of ILI, not influenza. Any examination of vaccine effectiveness would require further modelling or data specific to influenza.

3.2.2 Behaviour model

We construct a model (Figure 3.1) that accounts for user behaviour and informs an estimate for the prevalence of ILI in the population. For a household i in week w, let Y be a binary indicator of an on-time report submitted by the master of the household, I be a binary indicator of at least one participant of the household having ILI, and M be a binary indicator of the master of the household having ILI.

For every household in every week, there can be one of four outcomes:

- A household submits a report on-time, and no participants report having ILI ($Y = 1, I = 0, M = 0$),
- A household submits a report on-time, and at least one member that is not the master reports having ILI, ($Y = 1, I = 1, M = 0$),
- A household submits a report on-time, and the master reports having ILI, ($Y = 1, I = 1, M = 1$), or,
- A household does not submit a report on-time ($Y = 0$).

Logistic regression can be used to estimate a household's probability of reporting p, for some set of predictors $\boldsymbol{x}_{i,w}$ and parameters $\boldsymbol{\theta}$, β_H and β_M. The link function relates the log odds ratio of p to a linear combination of the predictors and their coefficients, and is defined as

$$\log\left(\frac{p(I,M)}{1-p(I,M)}\right) = \quad \boldsymbol{x}_{i,w}^t\boldsymbol{\theta} + I\left((1-M)\beta_H + M\beta_M\right), \tag{3.1}$$

where:

- $\boldsymbol{x}_{i,w}$ is the vector of predictors for reporting,
- $\boldsymbol{\theta}$ is the parameter vector of regression coefficients,
- β_H is the parameter for the change in reporting behaviour due to the household having at least one member with ILI, but not the master, and,
- β_M is the parameter for the change in reporting behaviour due to the master having ILI.

The odds ratio of reporting are therefore multiplied by e^{θ_d} for every unit increase in x_d for each predictor d.

Assume each household member has an independent probability π_w of having ILI in week w, then we can define $\phi_{i,w}$ to be the probability that a household of size n_i has at least one individual with ILI in a given week w. The probability $\phi_{i,w}$ can then be modelled as

$$\phi_{i,w} = 1 - (1 - \pi_w)^{n_i}. \tag{3.2}$$

The probability of the master having ILI, $M = 1$, and reporting, $Y = 1$, is below. The other three possible outcomes for a given household in a given week (Figure 1) can be similarly derived, and are not presented.

$$\begin{aligned} P(M, Y, I) =& P(Y|I, M)P(M|I)P(I) \\ =& P(Y|M)P(M|I)P(I), \text{ as } M \subset I \\ =& p(I, M)P(M|I)\phi_{i,w}. \end{aligned}$$

Defining $\zeta = P(M|I)$, we have

$$\begin{aligned} \zeta &= P(M|I) \\ &= \frac{P(I|M)P(M)}{P(I)}, \text{ by Bayes' Rule,} \\ &= \frac{P(M)}{P(I)} \text{ as } P(I|M) = 1, \\ &= \frac{\pi_w}{\phi_{i,w}}. \end{aligned}$$

Table 3.1: Variables from the dataset used in constructing predictors.

Chronological	Epidemiological	Demographic
Survey week	Had fever	Participant ID
Year	Had cough	Household ID
Join date	Days of absence	Postcode
Submit date	Sought medical advice?	Age group
	Diagnosis	Health worker
	Test results	Vaccinated

The likelihood $\mathcal{L}$ of the observations occurring given a set of parameters is then the product of the likelihood of each household i in each week w in the training set

$$\mathcal{L} = \prod_{i,w} \mathcal{L}_{i,w} \tag{3.3}$$

where

$$\mathcal{L}_{i,w}(M, Y, I) = P(M, Y, I) \tag{3.4}$$

and

$$\begin{gathered} P(M, Y, I) = p(I, M) \left(\phi_{i,w} \zeta^{M} (1-\zeta)^{1-M}\right)^{I} (1-\phi_{i,w})^{1-I} \\ \forall \text{ M, Y, I } \in \{0, 1\}. \end{gathered} \tag{3.5}$$

The likelihood $\mathcal{L}$ is the probability of the data given the set of parameters.

3.2.3 Model fitting

We use a Bayesian framework to estimate the posterior distribution of the parameters $\boldsymbol{\theta}$, $\boldsymbol{\beta} = (\beta_H, \beta_M)$, and $\boldsymbol{\pi}$, conditional on the data $\boldsymbol{X}$ and $\boldsymbol{y}$, via Bayes' rule:

$$\mathrm{P}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\pi} | \boldsymbol{X}, \boldsymbol{y}) \propto \mathrm{P}\left(\boldsymbol{y} | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{X}\right) \mathrm{P}\left(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\pi}\right),$$

where $\mathrm{P}\left(\boldsymbol{y} | \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{X}\right) = \mathcal{L}$ (Equation 3.3). The inference of $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and π_w jointly, means our estimate of the prevalence of ILI in week w, π_w, accounts for the impact of user behaviours via $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$.

Prior distributions for the parameters were:

$$\boldsymbol{\theta}, \boldsymbol{\beta} \sim \text{Norm}(\mathbf{0}, 0.7\mathcal{I}),$$
$$\boldsymbol{\pi} \sim \text{Unif}(0, 1),$$

where $\mathbf{0}$ is the zero vector and $\mathcal{I}$ is the identity matrix.

The variance of the prior distributions for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ were chosen to provide a near uniform prior density for the probability of reporting and reporting on-time, given the distribution of the training set predictors. Figure S1 in the Appendix shows the prior distribution transformed from the log odds scale to the probability scale, after multiplication with samples from the training set predictors. Results were not sensitive to this choice of variance.

All predictors $\boldsymbol{x}_{i,w}$ were scaled to be between 0 and 1, to allow for simplicity in comparing the predictive strength of each parameter. The set of predictors used for regression were derived from variables listed in the Appendix (Table 3.1), and the complete set of predictors used can be seen in Table A.1.

To estimate the posterior distribution of the parameters of the model, we used Hamiltonian Monte Carlo (HMC) implemented in the software package Stan (version 2.18) [18]. HMC is particularly effective over other Markov chain Monte Carlo methods, such as Random Walk Metropolis-Hastings, in high dimensional parameter spaces such as in this analysis, and the Stan implementation uses adaptive tuning of algorithm parameters, reducing the need to tune the inference algorithm [11].

For model training and inference of results, 80% of the households in each year were used in the training set for the model, and the remainder left in a test set for cross-validation of model predictions. For each season examined, reports in the first week were used in determining the prior behaviours of users and used in calculating predictor values, but were not used to infer parameters of the model, as predictors involving prior behaviours could not be well defined in the first week.

3.2.4 Data access

Reproduction of this study would require individual level data, which is precluded by ethics approval (The University of Adelaide Human Research Ethics Committee (HREC) H-2017-131). Anonymised data may still allow the possibility of identification of individuals, given the rich set of features and the small numbers in some postcodes.

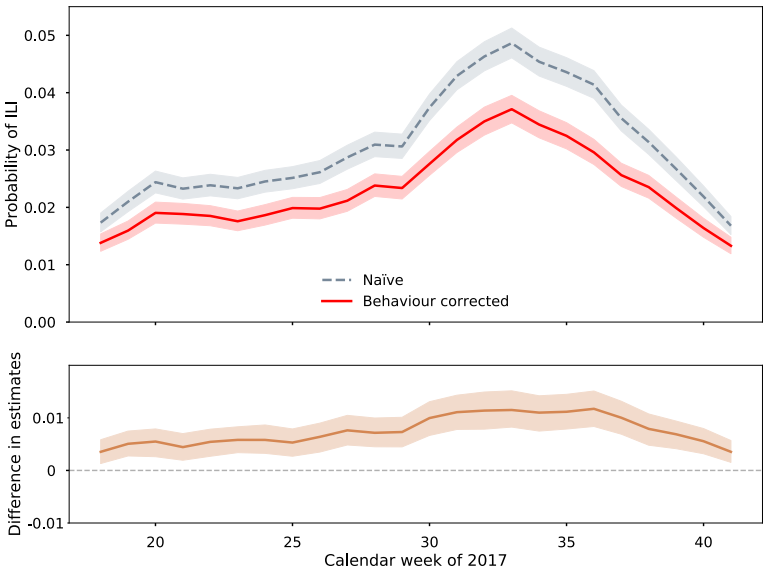

Figure 3.2: Comparison of behaviour corrected posterior estimate of the prevalence of ILI in the population (solid) to the naïve estimate (dashed) for the winter of 2017, and the difference between the two distributions over time. Lines represent the median of the distribution and shaded regions are 95% credible intervals of the posterior distributions.

3.3 Results

3.3.1 Comparison of behaviour corrected estimate to naïve estimate

Without correcting for user behaviour, our results show that naïve estimates of ILI prevalence are likely to be biased and overestimate the actual prevalence. The distribution of the naïve estimates have nearly all their probability density higher than the distribution of the behaviour corrected estimates for the ILI prevalence (Figure 3.2). This bias was found to be greatest when prevalence of ILI was greatest, and implies that users may be more likely to report when they have ILI. This trend was observed for every year of analysis (see Appendix A Figures A.2 to A.7). Figure 3.2 shows the results from training the model on the year 2017.

3.3.2 Reporting behaviour

Predictors of reporting behaviour were determined from chronological, epidemiological and demographic data (Table 3.1). Given a consistent identification number for each individual and each household, predictors such as the week of the survey, whether the individual reported having a single symptom previously this year, and whether an individual is reporting on behalf of others can be determined. A full list of predictors used and their explanations can be found in the Appendix (Table A.1).

Unsurprisingly, the strongest predictor for reporting behaviour was found to be the proportion of reports submitted on-time in the year so far for an individual, with all posterior samples of log odds ratio greater than 4 (see Figure A.10). Here we will present predictors of interest, and the marginal densities of all predictors are presented in Appendix A.

A comparison of marginal posterior densities of regression coefficients for predicting the probability of reporting (p) can be seen in Figure 3.3. Those reporting on behalf of others were not found to be much more likely to report, inconsistent with analysis in other studies [6, 9]. These studies did not consider the past reporting behaviour of users in their analysis, and this may have confounded this predictor in previous work. Individuals who have reported being vaccinated for this season were also found to be more likely to report in any given week.

Individuals who had reported having a symptom in previous weeks were much more likely to submit a report. However, the increasing number of weeks since the symptom occurred was a strong predictor for not submitting a report. Given no other differences, the model then predicts an increase in probability for an individual to report when they have reported having a symptom, with the increase decaying over time as the number of weeks since reporting the symptom increases. Figure 3.3 also shows that there is a large increase in the log odds of reporting when the household, but not the master, has ILI. However, when the master has ILI, there is a substantially larger increase in the log odds of reporting. This would indicate that the status of having ILI in proximity, and in particular personally experienced, is a strong predictor of reporting and participating in FluTracking. The posterior densities are remarkably consistent across the years, with the direction of the effect of the predictor consistent in all years, with the exception of those with small effects.

The current week was also used as a predictor for the probability a user will

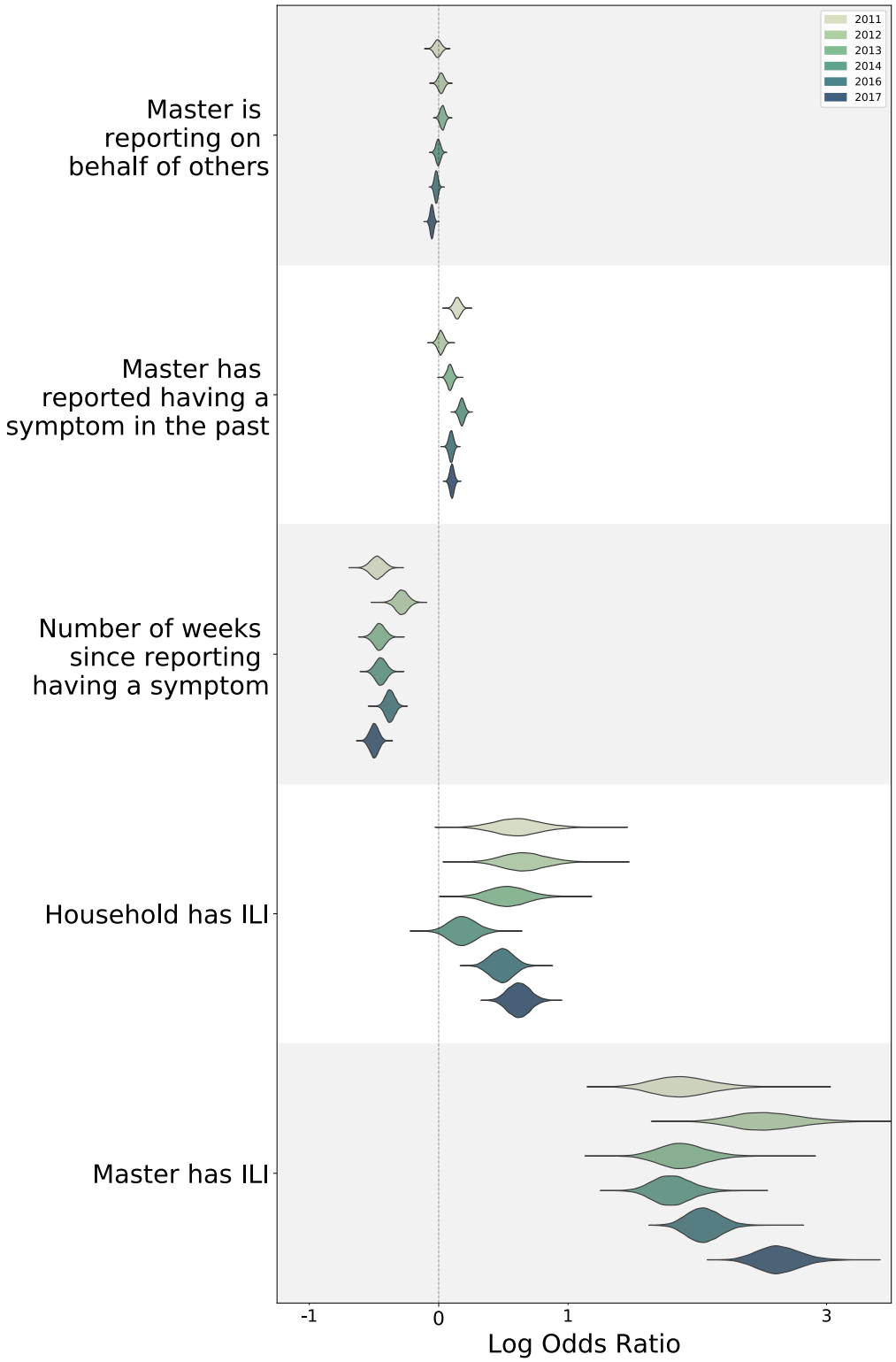

Figure 3.3: The marginal posterior densities of the log odds ratio of certain regression coefficients for predicting the probability of an individual reporting on-time across all years examined. For the complete set of regression coefficients, see Appendix A.

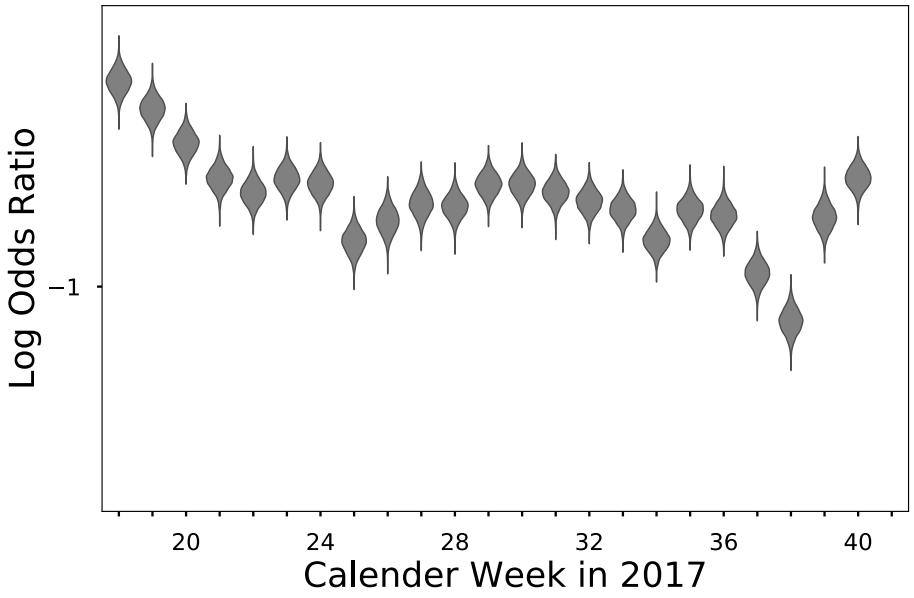

Figure 3.4: The marginal posterior densities of the log odds ratio of chronological regression coefficients for predicting the probability of an individual reporting in a given week of 2017. For all other years, please see Appendix A.

report, and their marginal probabilities of reporting declined sharply in the first third of the season, before a more shallow decline over the rest of the season (Figure 3.4). This could be explained by user fatigue, where dedication to participating in a voluntary system wanes through the year. This trend was also seen in every year studied (see Appendix A Figure A.12).

3.3.3 Model validation

To validate the model, 20% of the households in each year were excluded from training the model and kept as a test set. Model predictions were compared to observations in the test set by taking 1000 samples of the parameters from the posterior distribution, and simulating outcomes from the model for each sample and each household in the test set. The simulated data classified each household in each week into one of the four outcomes (Figure 3.1).

The proportion of reporting households with at least one participant with ILI, and the participation rate of households of the simulated data, were compared to observations from the test set. The proportion of reporting

households with at least one participant with ILI was calculated by dividing the number of households reporting on-time with ILI by the total number of households reporting on-time. The participation rate of households is determined by dividing the number of households reporting by the total number of households registered who have submitted at least one report in that year. Both summary statistics from the data were similar to simulations from the model on the test set using samples from the posterior, and these comparisons can be seen in Appendix A Figures A.13 to A.19. The simulated data from the model matches well to the actual observations from the test set and appears to retain the autocorrelation or time dependence of the actual test outcomes, without time dependence being explicitly included in the model. As the model predictions correspond well to outcomes observed in the test set in both summary statistics, the model shows no indication of bias of observable outcomes at a population level.

As the model is able to predict the weekly prevalence of households with ILI and the participation rate of users in FluTracking from a test set (Figures A.13 to A.19), we can see that the model correctly predicts user behaviour while providing insight into the drivers of participation and an improved estimate of ILI prevalence in the population.

3.4 Discussion

Online participatory health surveillance systems strive to provide near real-time estimates of disease prevalence, and yield complementary insights to traditional public health systems. However, the voluntary nature of these systems, and the reliance on user participation, need to be considered.

In this chapter, we find that the presence of ILI in the household, as well as other demographic factors, impacts the probability a user will submit a report. At a population level in FluTracking, this results in overestimation of the prevalence of ILI in the population when using the naïve estimate. This difference is greatest near the peak prevalence of ILI. This may be due to users being triggered to report by their symptoms, and being more likely to report in voluntary health surveillance systems when directly affected by the illness, through their household, or more strongly by personal experience.

Participation in FluTracking, derived as the number of on-time reports divided by the number of registered users reporting at least once in a given year, generally ranges between 70% and 90% for any given year and any

given week. This is much higher than rates observed in Influenzanet and Flu Near You, where previous studies have shown that 70% and 35% of users respectively submit more than 3 reports in a season [6, 9]. This is despite FluTracking having proportionally much higher numbers of participants than the European and US systems, relative to the respective populations of these regions. While the variance of participation over time has not been examined in the other systems, correcting for user behaviour could potentially be even more critical than observed here.

The estimates of the prevalence of ILI were not modelled with influenza vaccination status incorporated, but were used in the model of behaviour. As mentioned earlier, models with vaccination status included were considered, but are not presented here. Models which produce simple ILI prevalence estimates, as examined here, cannot be used as a measure of influenza vaccine effectiveness. The vaccine is not targeted to ILI, and the changes in prevalence over time does not consider the population size, or the number of cases of influenza reduced in a season. For these reasons, vaccination status was not incorporated into the model for prevalence estimates.

However, there exists the potential to construct a measure of vaccine effectiveness by comparing the two groups. The conventional metric of test-negative cases [62] does not extend simply to the behaviour corrected estimates presented here. However, with the ability to correct for potential bias in the data and reporting behaviour across different years, further studies may help inform vaccine effectiveness estimates in near real time.

Whilst past behaviours and illness status were used to predict user participation within the season, this study did not attempt to train the model across seasons. Training the model across every season would allow for users' behaviour in past years to inform behaviour in later seasons. However, the computational difficulty in substantially increasing the size of the training data, and the number of parameters involved, remains a challenge left for further work.

Predictors for user behaviour have only been taken from within the FluTracking data set. Social media and news coverage [82] and public awareness [41] are some examples where external factors can influence behaviour during an epidemic. Inclusion of these factors in the model may further improve estimates of disease prevalence where user behaviour may be significant.

This analysis assumes each individual has an equal and independent probability of having ILI in a given week. Incorporating a spatial or mechanistic model, particularly given the postcode and demographic information in the

data set may enable further insights into the mechanisms that drive the transmission of influenza and ILI in Australia, and provide spatially resolved prevalence estimates. Recent analysis has shown the onset of influenza epidemics is largely synchronised across regional areas [43].

In this chapter we have shown that user behaviour can have a significant effect on disease prevalence estimates drawn from the data. The framework developed herein, which elucidates drivers of user reporting in voluntary health surveillance systems and improves estimates of disease prevalence, emphasises the importance of the relationship between human behaviour and observations of disease. In the following chapter, we will consider this relationship from another dimension of disease observation.

Chapter 4

Behaviour in influenza testing associated with family doctors

In Chapter 3, we introduced a Bayesian framework that examined the role of individual behaviours on the observation of disease. We showed that the presence of symptoms and different demography influenced the probability of reporting in a participatory web surveillance platform. In this chapter, we will examine how demography can influence the probability of being associated with an influenza test from a family doctor or primary care physician. Traditional surveillance will contain information on those who have received a test, but not those where the physician chose not to test. FluTracking, and other web-based surveillance platforms, can provide this insight, and examine whether different age groups are under or over represented in traditional data sets of confirmed influenza activity, whether due to unconscious bias or risk based assessments.

4.1 Background

4.1.1 Healthcare-seeking behaviour

Seasonal influenza places a significant burden on public health systems. Traditional influenza surveillance relies on notifications from family doctors, hospitals and laboratory confirmed cases, with family doctor (general practitioner) testing providing a key source of information [110]. This unfortunately takes time and costs resources, relying on samples physically sent to

laboratories and then notifications passed on, resulting in up to several weeks of delay between observation of illness and confirmation. Web-based participatory health surveillance systems address these issues by providing timely and low cost surveillance that is complementary to traditional systems. As examined in the previous chapter, FluTracking [28] is one such online health surveillance system, primarily focussed on influenza-like-illness (ILI) surveillance. We will again see that this digital dataset has an important role in providing insight into human behaviour in observing disease prevalence.

The effect of human behaviour and healthcare-seeking behaviours are known to have a substantial influence on the surveillance of disease outbreaks [40, 59, 73], and we have already seen an example of this for FluTracking in the previous chapter, where user reporting behaviour changed when symptoms occurred. The impact of behaviour is not restricted to the patients, but includes the clinical decision making of family doctors [12, 79]. As a key source for traditional surveillance, the behaviour of family doctors directly influences observations in public health surveillance and must be understood. Web-based health surveillance platforms can provide novel insight into these behaviours that were previously difficult or unobtainable from traditional surveillance [8, 26, 84].

It has been observed that healthcare-seeking behaviour and perceived risk can be predicted from individual and household demographics [29, 61], and this could potentially influence the family doctor in their likelihood of requesting a test. If this influence does occur, then there may be subsequent unforeseen consequences in patient outcomes, as influenza testing has been observed to influence clinical decision-making and management of patients [12].

In this chapter, we demonstrate that testing behaviour can be more effectively predicted and understood using data that utilise digital surveillance systems by incorporating the complementary information they collect compared with standard notification data. For example, FluTracking collects data on individuals who may have not sought healthcare services. We present results that provide insight into the predictors of influenza testing associated with family doctors, and explore the difficulty in the prediction of rare, uncertain events and the appropriateness of certain metrics in determining the performance of models.

4.1.2 Validation metrics

The application of machine learning (ML) models to healthcare problems has been examined previously due to their predictive performance [2, 19, 25]. However, examinations of the risks and issues with intelligibility require comparisons with statistical models such as logistic regression, with consideration given to balancing interpretability with predictive ability. The most appropriate choice of performance metric is often unclear, as the risks and ethics involved in healthcare require careful consideration on what a predictive model should prioritise.

The choice of a validation and performance metric in health prediction problems is often challenging due to the consequences of model predictions. The simplest metrics commonly used are precision, recall and accuracy. Precision is the proportion of correct positive classifications among all predicted positive classifications, but neglects to consider the performance in predicting negative classifications. Recall also shares the same issue, being defined as the proportion of correct positive classifications among all true positive classifications. Accuracy is defined here to be the proportion of correct predictions of all classifications. While accuracy may appear to be an intuitive choice for a performance metric, it does not allow for consideration to the risks of classification errors and can be misleading when classes are unbalanced in the data [108], such as in disease incidence or testing rates.

The balanced accuracy score can address the unbalanced classes, as it is the average of the recall of each class. However, it does not consider sensitivity and specificity, which may have ethical or risk based consequences that must be considered in a public health setting. Though the F1-score (the harmonic mean of precision and recall) does account for sensitivity and specificity, the score is evaluated at a chosen threshold for models with probabilistic or score based predictions and does not consider the true negative rate. The Receiver Operator Characteristic (ROC) curve allows comparison at all thresholds, providing a visual description of a model's performance and allows for a model to be assessed according to varying levels of risk. For this reason, particularly in healthcare, the area under the ROC curve (AUC) [36], is often used as an overall measure of a model's performance [2, 19, 25].

In this work, we seek to understand what factors influence the probability of a patient receiving an influenza test, and not necessarily the best performing classifier. Strictly proper scoring rules are metrics that are uniquely optimised by the true probability distribution of the event. The logarithmic loss is one such rule [45], and therefore minimising the logarithmic loss will result

in correctly producing the true probabilities of the outcome. This property of the logarithmic loss and its similarity to the binomial likelihood in Bayesian logistic regression suggests it is a natural performance metric to consider in a binary classification problem. Here we will use, where applicable, the score produced by the ML algorithm in place of probabilities to evaluate the logarithmic loss.

Each performance metric has its own advantages and disadvantages, so for completeness, this study will include all metrics discussed above.

4.1.3 Dataset

We have already encountered the FluTracking dataset in the previous chapter, but we will reintroduce particular aspects for clarity and their applicability to the work in this chapter. Participants of FluTracking provide various demographic details about themselves at registration. Users may also report on behalf of other members in their household. After registering, participants are sent an email each week requesting a report on whether they experienced ILI symptoms in the week before. If they report having ILI symptoms, follow up questions about choices of healthcare-seeking behaviour and whether an influenza test was received are asked. The entire set of variables taken from FluTracking and considered in this study can be seen in Table 3.1.

In this work, we considered all participants, both masters and their household members, who reported having ILI and visited a family doctor at least once in the year in Australia between 2011 and 2017. This is composed of 831,593 unique reports from 22,824 individuals. These reports and the variables listed in Table 3.1 were used to generate the demographics and past health history of each individual and their reported households. The predictors generated by these reports and used in the model are summarised in the Appendix in Table B.1.

4.2 Methods

We seek to use the derived variables in Table B.1 to predict the probability of an individual receiving an influenza test, given they present to the family doctor with ILI symptoms. We are therefore interested only in reports where the individual reported having ILI and visited the family doctor. If an individual visited the hospital or emergency department in the same week

Table 4.1: The number of reported family doctor visits and influenza tests received by users of FluTracking in each year of study.

Year	Doctor visits reported	Tests conducted (% of visits)
2011	2,292	96 (4.2%)
2012	2,995	149 (5.0%)
2013	2,772	111 (4.0%)
2014	4,151	274 (6.6%)
2015	5,469	382 (7.0%)
2016	5,437	421 (7.7%)
2017	6,832	957 (14.0%)
Total	29,948	2390 (8.0%)

as the family doctor, where the test was requested cannot be determined in FluTracking. Therefore, all of these reports are excluded. This results in a total of 29,948 surveys of interest. The breakdown of the dataset by year can be seen in Table 4.1.

For each individual i visiting the family doctor with ILI, we seek to use a set of predictors $\boldsymbol{x}_i$ (Table B.1) to determine the probability p_i that the patient will receive a test requested by the family doctor. We define $y_i = 1$ to indicate a test was conducted, and $y_i = 0$ if not. Then

$$y_i \sim \text{Bernoulli}\left(p_i\right). \tag{4.1}$$

We use Bayesian logistic regression to estimate p_i,

$$\log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{x}_i^T\boldsymbol{\theta}, \tag{4.2}$$

where $\boldsymbol{\theta}$ is a parameter vector containing the regression coefficients, or weights, of the predictors. The posterior distribution of the parameters $\boldsymbol{\theta}$ conditional on the data $\boldsymbol{X}$ and $\boldsymbol{y}$, via Bayes' rule:

$$P(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{y}) \propto P\left(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{X}\right)P\left(\boldsymbol{\theta}\right). \tag{4.3}$$

The prior distribution chosen for the parameters is

$$\boldsymbol{\theta} \sim \text{Normal}(\mathbf{0}, 0.7I),$$

where $\mathbf{0}$ is the zero vector and I is the identity matrix. The choice of the covariance matrix in this prior for the parameters in the log odds space provides

a near uniform distribution in the probability space, while also providing a level of regularisation on the posterior distribution of the parameters.

We also compare our model to a naïve proportion:

$$\bar{p_i} = \frac{\sum_i y_i}{N}, \tag{4.4}$$

which is the assumption that the testing rate is constant through the population, and testing is uniformly random at the mean testing rate.

For validating and benchmarking our Bayesian logistic regression, we seek comparisons with other supervised learning algorithms which require little to no tuning. The ML classifiers used were the support vector machine with a radial kernel (SVM-RBF), the random forest (RF), and gradient tree boosting (GB) [57]. All ML algorithms were implemented using the scikit-learn package (version 0.20.2) [92]. Unless otherwise stated, the default parameter values for the implementation of these classifiers were used. As the data set contains unbalanced classes, the SVM-RBF and RF algorithms were trained with the samples weighted inversely proportional to their class frequencies. The GB algorithm inherently addresses unbalanced classes, as assigning weights to individual samples is a key step in the algorithm.

4.2.1 Model fitting

All but one predictor value ranged between 0 and 1. Only the parameter for the number of days absent from normal duties had a different range (from 0 to 7), and this was normalised to the unit interval to improve model training, parameter comparisons and interpretation.

To estimate the posterior distribution of the parameters of the model, Hamiltonian Monte Carlo (HMC) was used through the Python implementation of the software package Stan (version 2.18) [18], as we have done throughout this thesis.

4.2.2 Validation

To validate and measure the performance of the models, a random 80% subsample of the data was used for training, and the remaining data excluded as a test set. Comparison classifiers are similarly trained and validated on the same subsamples. All performance metrics are evaluated using the test set.

Table 4.2: A comparison of the Bayesian model against other ML classifiers. All values listed are median values (90% credible intervals) of the distribution of the metrics (excepting SVM-RBF and RF). The best value of each metric (by median where applicable) is in bold.

Algorithm	Accuracy	Balanced Accuracy	F1 score	Log loss	AUC
Naïve	0.854 (0.850-0.858)	0.500 (0.492-0.509)	0.079 (0.064-0.095)	1642	0
Logistic Regression	0.859 (0.854-0.864)	0.518 (0.508-0.527)	0.111 (0.094-0.129)	1562 (1560-1565)	0.671 (0.668-0.673)
SVM-RBF	0.675	**0.628**	**0.216**	-	-
RF	0.719 (0.713-0.725)	0.585 (0.543-0.567)	0.167 (0.157-0.177)	infinite*	0.612
GB	**0.860** (0.856-0.864)	0.518 (0.509-0.528)	0.113 (0.095-0.130)	**1556**	**0.678**

Where probability estimates p_i or scores are given by the algorithm (logistic regression, RF and GB), the performance metrics (excepting logarithmic loss and AUC) were calculated by evaluating p_i for every i using the model, simulating Bernoulli trials and then calculating the performance metric over all i. This was repeated 1,000 times to construct a distribution of each performance metric. In the Bayesian logistic regression model, p_i was calculated using a random sample of the posterior $P(\boldsymbol{\theta}|\boldsymbol{X}, \boldsymbol{y})$ for each repetition and for each individual i, carrying the uncertainty through the evaluation of the metrics.

4.3 Results

We constructed a Bayesian model to estimate the probability of receiving an influenza test requested by the family doctor given the patient presented with ILI. We benchmark our model against ML classifiers to validate the performance of our model, before examining the insight into doctor behaviour.

4.3.1 Model performance

All models were an improvement on the naïve model in nearly all metrics, indicating that there is predictive power in the data provided by the detailed

FluTracking reports. Bayesian logistic regression performed favourably in accuracy, logarithmic loss and AUC, where it is within 1% of the best performing algorithm. This provides confidence in the validity of the insights drawn from our model.

The prediction performance of the naïve model, Bayesian logistic regression and the ML classifiers are in Table 4.2 and Figure 4.1. Where necessary, distributions of metrics are summarised with the median of the distribution and the 90% credible interval. Note that RF made multiple incorrect predictions with probability 1, resulting in infinite log loss.

While the SVM-RBF performs the best in balanced accuracy and F1 score, it performs very poorly in accuracy. As SVM-RBF is unable to provide scores or probabilities without further tailoring, it was not assessed using the logarithmic loss or AUC, and therefore it is difficult to assess its appropriateness for predictions in a public health setting, particularly with its low interpretability [57]. While GB performs slightly better than the Bayesian logistic regression across all metrics, it is only a 1% improvement in AUC compared to the AUC of 0.673 of the maximum *a posteriori* (MAP) estimate, and has similarly small improvement in the other metrics.

4.3.2 Interpretation of predictors

To examine the impact of each predictor on the outcome of whether an influenza test is received, the marginal posterior densities of the log odds ratio of the regression coefficients can be compared in Figure 4.2. This level of direct interpretation in only possible for the Bayesian logistic regression model.

The greatest predictor of whether an influenza test will be received is the number of days absent from normal duties and work. This can be interpreted as a measure of the severity of the illness presented to the family doctor, and implies that the more severe the ILI is, the more likely it is an influenza test will occur. Patients where the current visit to the family doctor for ILI is not the first visit for ILI in the same year are predicted to be tested more often. This could be interpreted as patients who seek healthcare more than once, whether due to severity or risk perception, are more likely to be tested. The implied severity of ILI is unsurprisingly a strong predictor for whether an influenza test will be received.

Interestingly, if the patient is in a patient facing occupation, they are more likely to have an influenza test. This may be due to the increased risk of

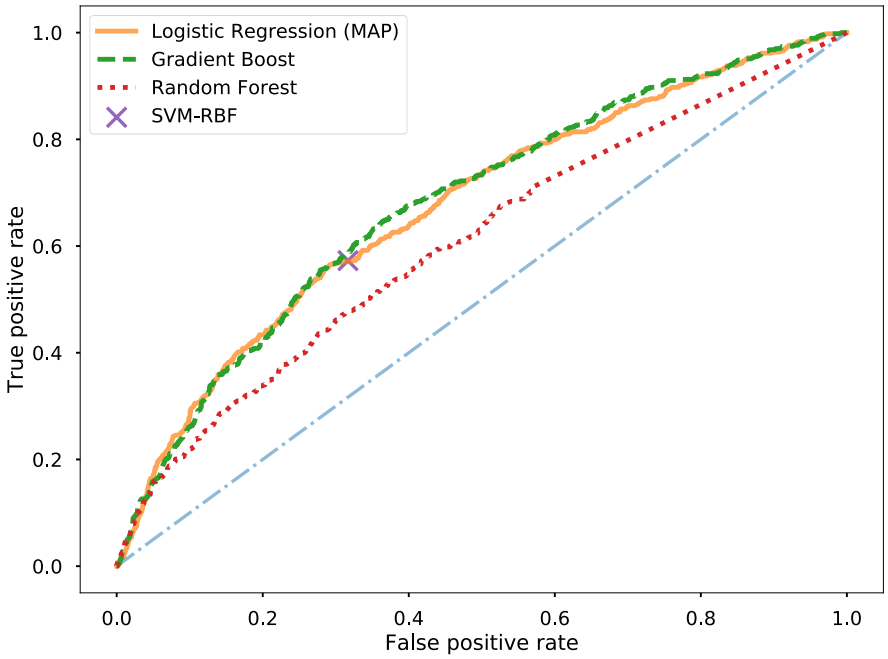

Figure 4.1: The ROC curves of the MAP estimate of the logistic regression (yellow solid), the gradient boosted trees (green dashed), the Random Forest (red dots), and the Support Vector Machine with radial basis kernel (purple cross). The naïve estimate (blue dashdot) is plotted for comparison.

transmitting influenza by the patient due to their occupation, or access to testing services. Regardless of the cause, this indicates that healthcare workers are over represented in influenza testing statistics, and may bias surveillance systems. This is similarly the case for elderly patients, and patients with elderly members in their household. Previous work has shown that these demographics are generally over represented in other web surveillance systems [9].

The model was unable to identify the predictive power of the patient being under 5 years of age, as the marginal posterior distribution is very similar to the prior distribution. This is in direct contrast to the predictive power of the patient being between 5 to 15 years of age (a youth), which the model finds is strongly predictive of not receiving an influenza test. This is despite the number of individuals in the training set under 5 years of age being quite similar to the number of individuals between 5 to 15 years of age. Alongside the model's inference that healthcare workers, elderly patients, and patients with elderly members in the household, it can be seen that influenza testing rates are also biased towards these demographics. This has important implications for interpretation of traditional summary statistics of influenza surveillance, particularly where there are differing risks for certain demographics or age profiles [110].

Figure 4.3 displays the base rate of testing, and the variation in testing across all individual years examined in this study, outside of the other included predictors. While the base rate of testing was low (negative log odds ratio), it can be seen that as the years progressed, there was an overall increase in the probability of receiving influenza tests. This has implications in comparing the population statistics from traditional surveillance systems across different years, due to the substantial change in testing rates, even accounting for the demographic differences considered in this model.

4.3.3 Individual level predictions

A further advantage of the Bayesian method over traditional ML classifiers is the ability to examine individual level predictions with an inherent characterisation of uncertainty. This can be used by health practitioners to comprehensively examine risks. Figure 4.4 describes a specific patient's posterior predictive density of receiving an influenza test (blue), and compares this to the population posterior and the naïve model prediction (mean testing rate). The individual's density is much more concentrated than the population density, demonstrating that the variation among patients is what drives the

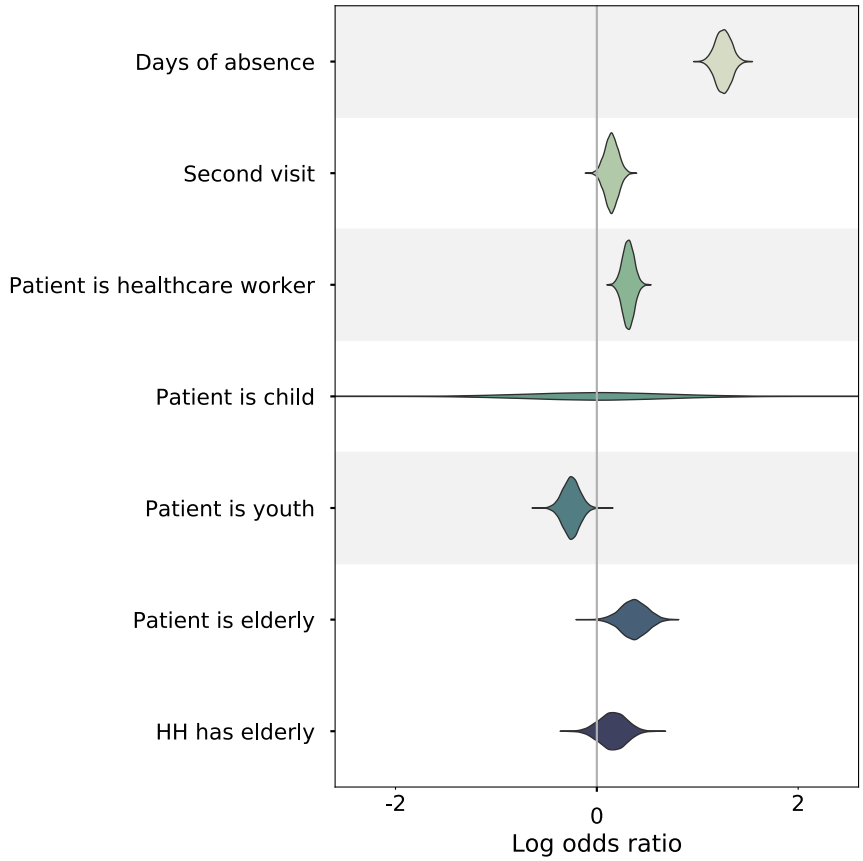

Figure 4.2: The marginal posterior densities of the log odds ratio of certain regression coefficients for predicting the probability of receiving an influenza test when presenting to the family doctor.

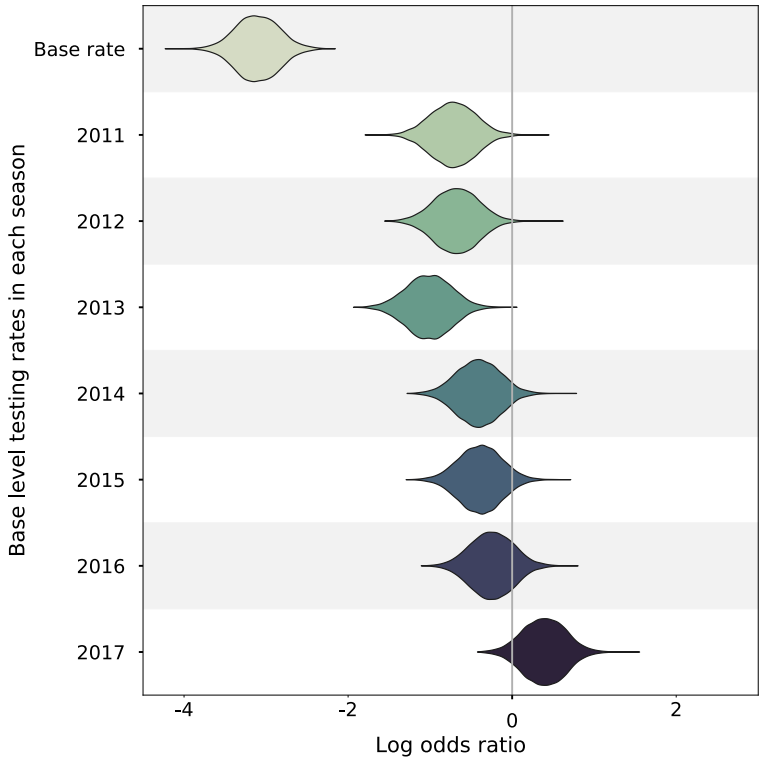

Figure 4.3: The marginal posterior densities of the log odds ratio of the regression coefficients for base level predictors of receiving an influenza test when presenting to the family doctor. The sum of the parameter values of base rate and the corresponding year gives the overall base level of testing in a given year, given the same demographic predictor values. A gradual increase in base level of testing can be seen over time, particularly in recent years.

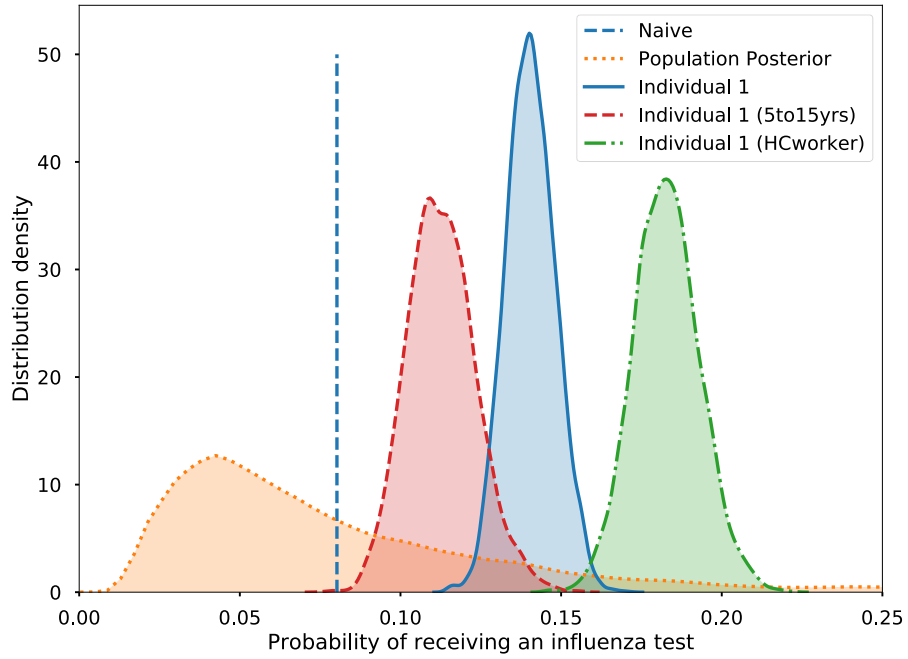

Figure 4.4: The posterior density of the probability of receiving an influenza test for a specific individual in 2017 (blue solid) when presenting to the family doctor, and the impact of changing various predictors on this individual's probability. If this individual was 5 to 15 years of age (red dashed), or was in a patient facing occupation (green dashdot), the posterior density shifts substantially. The population posterior predictive density (yellow dotted) and the naïve proportion (blue dashed) are plotted for comparison.

variance in the population posterior, not the uncertainty in the model. The effect of changing the patient's predictors and their effect on the posterior probability of receiving an influenza test can be clearly observed. For example, if the patient were actually in a patient facing occupation, the posterior predictive density is shifted substantially higher, with very low probability mass overlapping the original distribution. A similarly substantial shift, but in the opposite direction, occurs if the patient was a youth aged between 5 and 15 years of age.

4.4 Discussion

Our study shows that various demographic factors influence the decision by the family doctor to request, or the patient to receive, an influenza test. We see that age, household demographics and whether the patient is a health-care worker are important predictors in whether an influenza test will be conducted. This insight has important public health surveillance implications, and is observable through the analysis of a web-based health surveillance system. These insights would not have been easily discerned through traditional surveillance systems alone.

Computational time was in the order of seconds to minutes across algorithms. Further work will examine the impact of the insights discovered here, and their implications for the surveillance of disease outbreaks. The bias introduced into population level statistics and epidemic models may have substantial implications. A comparison with other web-based influenza surveillance platforms may also yield interesting insights into the cultural or geographic differences.

This study excluded reports where a hospital or emergency department visit occurred in the same reported week, and this may remove certain severe cases from the analysis. As explained in the methods, it is not possible to determine whether the influenza test was requested by the family doctor or by the other health care providers in the FluTracking system, and so must be excluded form this analysis. As this study has found the implied severity of illness to impact the probability of receiving an influenza test, bias may be introduced by excluding these occurrences and should be investigated in future work.

The marginal improvement in performance observed of the best ML algorithm (GB) indicates that logistic regression performs favourably against the ML algorithms as benchmarks. However, Bayesian logistic regression is able to provide insight into the predictors that influence whether an influenza test is received and inherently accounts for uncertainty in predictions. These factors are of importance to public health, and we argue should not be discarded for marginal improvements in predictive performance.

In this chapter we have continued the theme from Chapter 3 of deriving insights on behaviour in disease surveillance from a digital web surveillance system. In using Bayesian inference methods, we sacrificed very little in predictive performance compared to other ML classifiers, while maintaining the ability to discover how human behaviour has influenced the observation

of influenza through testing practises.

Chapter 5

Forecasting COVID-19 cases in Australia using social mobility metrics

In previous chapters we have examined how behaviour needs to be considered when examining digital data sets in epidemiology, and provided frameworks for determining their influence. Given the outbreak of the novel SARS-CoV-2 (COVID-19) viral pandemic as a major public health threat in early 2020, forecasting the spread of the disease amongst changing public health policies and public behaviour changes has become an imperative to assisting public health institutions in responding to the pandemic.

In this chapter we focus on using digital data, in the form of social mobility metrics, to forecast COVID-19 cases in all jurisdictions of Australia. Social mobility metrics can be used as a measure of the effect of public health policies and behaviour changes in forecasting models, adjusting for the time varying nature of the COVID-19 pandemic.

The effective reproduction number R_{eff} is an important latent parameter in epidemiology, representing the average number of secondary cases of an infected case. R_{eff} can be retrospectively estimated through examining the number of cases over time [27, 112], but to forecast cases using a mechanistic model, it must incorporate some estimate of the future transmission potential and/or arrival of infected cases. The relatively low number of cases in Australia also creates difficulties in utilising methods that rely on historical case incidence. Measures of mobility of each Australian jurisdiction provided by Google [49] and survey results of the public's behaviour in adhering to

personal distancing measures provides the ability to link these indicators to an estimate of the effective reproduction number in the near future. This allows for a mechanistic model to forecast cases.

The relatively few cases of local transmission in Australia, in conjunction with strict border control measures internationally and domestically, makes it natural to forecast the number of cases in each jurisdiction using a stochastic branching model. This generative model, using estimates from the literature for epidemiological parameters, can be paired with a model for the time varying effective reproduction number to forecast COVID-19 cases in Australian jurisdictions. This framework can adapt to changing public health policies and responses to the ongoing pandemic, particularly during small outbreaks and the irregular but frequent responses to outbreaks seen in Australia.

We develop a generative model of the dynamics of SARS-CoV-2 in Australia, and forecast COVID-19 cases by jurisdiction. The model links social isolation measures – captured via publicly available Google Mobility Indices[49], and an estimated "Micro-distancing" parameter – to the effective reproduction number of local infectious individuals, allowing us to produce forecasts under changes to Government-imposed isolation and social distancing measures.

Both this chapter and Chapter 6 use surveillance data reported through the Communicable Diseases Network Australia (CDNA) as part of the nationally coordinated response to COVID-19. We thank public health staff from incident emergency operations centres in state and territory health departments, and the Australian Government Department of Health, along with state and territory public health laboratories. We thank members of CDNA for their feedback and perspectives on the study results.

5.1 Methods

In this framework, social mobility metrics and indicators of public behaviour are incorporated into a model for the effective reproduction number, and calibrated to an established method [27, 112] that relies on historical case incidence. We obtain a posterior predictive distribution of the effective reproduction number over time, then use a branching process model to simulate and forecast the number of cases in each Australian jurisdiction, removing simulations that do not adequately reflect recent case incidence. An overview of the framework can be seen in Figure 5.1.

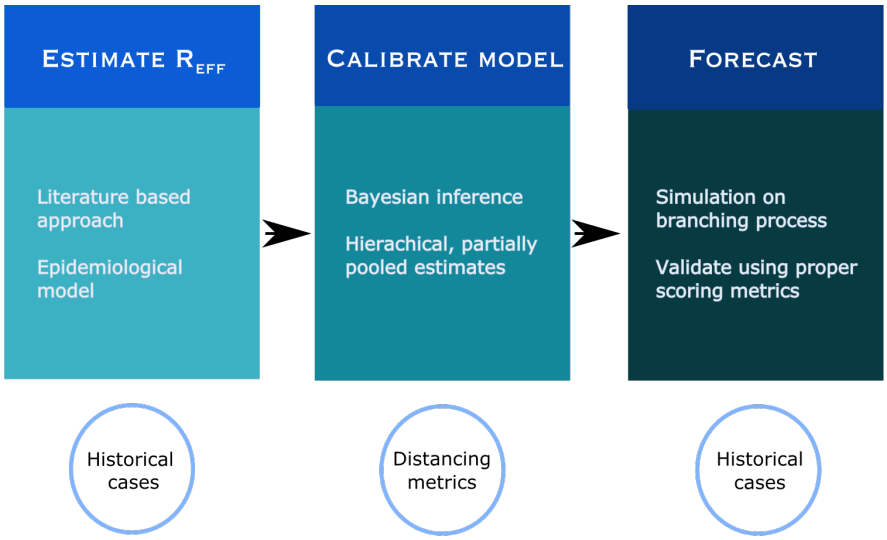

Figure 5.1: An overview of the pipeline for the model, from historical case data to estimating the effective reproduction number, calibrating the social mobility model, and then simulating and forecasting COVID-19 cases.

5.1.1 Data

Historical case incidence was obtained with permission from the National Notifiable Disease Surveillance System (NNDSS) of Australia, which provides the data of symptom onset and the place of acquisition, where known.

The Google Mobility Indices [49] are provided as Community Mobility Reports, and comprise of six mobility trends: Grocery & pharmacy, Parks, Transit stations, Retail & recreation, Residential, and Workplaces. All indices, except Residential, provide a measure of the number of visitors to locations of each category relative to a baseline in early 2020. The Residential index measures the relative change in duration, and therefore uses a different unit of measurement to the others. It can also be considered as the complement of the other categories for the majority of the population. For this reason, it is not used in this analysis. These trends represent proxies for the wider changes in social mobility, and is an indicator of the opportunities for transmission of COVID-19.

Individual behaviours, such as remaining 1.5 metres apart from non-household contacts, are surveyed originally by the Behavioural Economics Team of the Australian Government and currently by the Australian Department of Health to estimate social distancing at the individual level. This survey

is targeted to be representative of the Australian adult population, and is aggregated into a weekly summary of responses to determine the changing public response to transmission reduction behaviours. This measure gives an indication on the probability any given interaction will lead to an infection event.

5.1.2 Inferring effective reproduction numbers

We use the framework developed by Cori *et al.* (2013) [27] and Thompson *et al.* (2019) [112], where the number of local and imported cases observed each day are used to infer the distribution of the overall effective reproduction number, assuming fixed distributions of the incubation period and generation interval. We use an incubation period that is Gamma distributed with shape parameter 5.807 and scale parameter 0.948 [72]. Based on this incubation period, we follow [1] and use the generation interval defined by a Gamma distribution with mean 3.635 days and standard deviation 3.075 days, as described in [42]. We estimate a reporting delay distribution from the case data by finding the maximum likelihood estimates of the parameters of a Gamma distribution of the time between symptom onset, where known, to the notification date. This gave a shape parameter of 1.82, and a scale parameter of 2.88. The posterior distribution of the effective reproduction number R_{eff} from this process is estimated for each Australian State using case data, as seen in Figure 5.2. We assume a prior distribution of Gamma(1,2) and smoothing parameter $\tau = 4$.

This estimate will be referred to as the *historical reproduction number* R_{eff} from this point onwards.

5.1.3 Calibrating social mobility model

The estimates of R_{eff} in Figure 5.2 use historical case data, and cannot be used to forecast cases. In periods of low epidemic activity, the posterior distribution also reverts to the prior distribution, which in our implementation is uninformative when predicting the path of a possible outbreak of cases in the future. To address these issues, we combine social mobility measures and behavioural data to model and forecast the effective reproduction number in each Australian jurisdiction. However, instead of using the daily death rates and inferring the infection fatality ratio as in earlier work [39, 114], we calibrate this model to the historical estimates of R_{eff}. We do not rely

on inference from daily death rates here, due to the estimated high levels of detection of COVID-19 cases in Australia [95, 100].

Within Australia, there have been jurisdictional level differences in policy and response to social distancing, but the underlying culture and mobility patterns may have commonalities within the nation. As such, we employ a hierarchical model to partially pool information between jurisdictions, while allowing for inferred differences where they may occur.

These estimates of the effective reproduction numbers are assumed to be dependent upon the proportion of observed imported cases out of all observed cases (ρ), the impact of "macro" social isolation measures – captured via Google Mobility Indices ($\boldsymbol{\omega}(t)$) – and "micro-distancing" (M_d) behaviours – captured by the survey of the Australian public.

We link our estimates of the effective reproduction numbers with the latter features via the model described in equations 5.1 and 5.2, where $\hat{\mu}(t)$ represents the overall effective reproduction number of all cases, and $R_L(t)$ represents the time varying effective reproduction number from local cases:

$$\hat{\mu}(t) = \rho(t)R_I + (1 - \rho(t))\, R_L(t), \tag{5.1}$$

$$R_L(t) = R_{Li} M_d^{1_{\{\text{post-ban}\}}} \times 2 \times \text{logistic}\left(\boldsymbol{\beta}^T \boldsymbol{\Omega}(t)\right) \tag{5.2}$$

in which:

- $\rho(t)$ (inferred; jurisdictional level) is the proportion of imported cases (of all cases) on day t;
- R_I (inferred; national level) is the effective reproduction number of imported cases;
- R_{Li} (inferred; jurisdictional level) is the effective reproduction number of local cases at base levels of mobility for each jurisdiction i;
- M_d (inferred; national level) is a micro-distancing factor, which allows for the effective reproduction number of local cases to be reduced post 20/03/2020 ($1_{\{\text{post-ban}\}}$ is an indicator function of post that time point);
- $\boldsymbol{\beta}$ (inferred; national level) is a vector of parameters that links Google Mobility Indices to the effective reproduction number of local cases via the logistic function, which ranges between 0 and 1; and,
- $\boldsymbol{\Omega}(t)$ (jurisdictional level) is a vector comprising the Google Mobility Indices on day t.

The logistic function ranges from 0 to 1, and if the Google Mobility Indices are at baseline (the zero vector), then the logistic function will have value 0.5. To ensure at baseline there is no impact on the effective reproduction number, we multiply the logistic function by a multiple of 2.

20/03/2020 was chosen as the date at which micro-distancing takes effect, as this is the date many Australian government policies on social distancing were announced. We use a hierarchical model to infer R_{Li} via the following distributions, parameterised using shape and scale for Gamma distributions, and rate for Exponential distributions:

$$\begin{aligned} R_{Li} &\sim \text{Gamma}(R_{L0}^2/\sigma^2, \sigma^2/R_{L0}), \\ R_{L0} &\sim \text{Gamma}(1.8^2/0.01, 1.8/0.01), \\ \sigma^2 &\sim \text{Exponential}(20), \end{aligned}$$

where R_{L0} is the mean effective reproduction number at the national level at baseline levels of social distancing, and σ^2 the variance in the effective reproduction numbers across jurisdictions. Parameterised in this way, the hyperprior for R_{L0} and the prior for each R_{Li} has a mean of 1.8.

The micro-distancing factor M_d is modelled through the following relationship:

$$M_d = (1+\theta)^{-\gamma(t)}, \quad \theta, \gamma(t) > 0 \tag{5.3}$$

where $\gamma(t)$ is the proportion of individuals always complying with the recommended health policy of maintaining 1.5 metres of distance from non-household members on day t, taken from the behavioural survey.

We assume that (the likelihood)

$$\hat{\mu}(t) \sim \text{Gamma}(k(t), \theta(t)) \tag{5.4}$$

with $k(t) = \frac{E[R_{\text{eff}}(t)]^2}{Var(R_{\text{eff}}(t))}$ and $\theta(t) = \frac{Var(R_{\text{eff}}(t))}{E[R_{\text{eff}}(t)]}$, to calibrate the social mobility model with the posterior predictive distribution of R_{eff} estimates.

We perform inference in a Bayesian framework, using Hamiltonian Monte Carlo through the software package `pystan` [109]. Prior distributions for the

parameters are:

$$\begin{aligned}
\rho(t) &\sim \text{Beta}\left(1 + I(t), 1 + L(t)\right); \\
\boldsymbol{\beta} &\sim \text{Normal}(0, 1); \\
R_I &\sim \text{Gamma}(1.25, 0.4); \\
\theta &\sim \text{Lognormal}(0, 0.5); \\
\gamma(t) &\sim \text{Beta}\left(1 + A(t), 1 + N(t)\right),
\end{aligned}$$

where $I(t)$ and $L(t)$ are the number of imported and local cases on day t respectively, $\boldsymbol{\omega}(t)$ (state level) are the 7-day future moving average of Google Mobility Indices on day t, $A(t)$ is the number of survey respondents on day t who claim to always be complying to the recommended health policy of maintaining 1.5 metres of distance from non-household members, and $N(t)$ is the number of survey respondents who do not.

Calibrating the social mobility model (5.4) to time periods where the historical effective reproduction number R_{eff} has reverted to the prior is uninformative and will result in the model learning parameters that relate to the prior distribution. Small, localised outbreak clusters (<10 cases) may also not be brought about by population-level mobility, and the initial incursion should be excluded from the calibration. For this reason, the social mobility model is only calibrated to specific time periods in each jurisdiction.

We use case data for every state in Australia through the month of March, and case data from New South Wales and Victoria from 1st May to 10 days before the day the forecast begins to calibrate the social mobility model. We then generate the posterior predictive distribution of the effective reproduction numbers $R_L(t)$ and R_I for local and imported cases respectively in each jurisdiction over time. The marginal posterior densities of R_I and each R_{Li} in inferred from calibration can be seen in Figure 5.5, and indicates substantial variation in baseline reproduction numbers across jurisdictions. The low value of R_I is consistent with the overall success of the quarantine of imported cases. The posterior distribution of $\hat{\mu}(t)$ can be seen in Appendix C, and matches well to the reference distribution R_{eff}.

The Northern Territory (NT) and Australian Capital Territory observed very few cases overall, and so were not included in the inference at any stage. Their baseline effective reproduction number is assumed to follow the posterior distribution of R_{L0}.

5.1.4 Forecasts of mobility indices

While the effective reproduction number is an important epidemiological parameter to consider, the intensity of the initial public health response must consider the predicted number of COVID-19 cases in the near future, which require validated forecasts.

Our forecasts are produced by first forecasting the Google Mobility Indices using a random walk with drift [81]. Uncertainty in each index was introduced by using the standard deviation of the 7 data points used in the 7-day moving average. For all Google Mobility Indices in each jurisdiction, the differences in each successive day were assumed normally distributed and the mean and covariance matrix of the distribution estimated through a maximum likelihood estimator. This was applied to the previous 28 days from the last Google mobility index entry, which may be before the date the forecast is occurring on. The indices were then forecasted for at least the next 28 days (28 forecast days and any extra required days that Google has yet to provide) by successively adding a sample from the estimated multi-variate normal distribution each day, starting at the average value of the index in the preceding week to the forecast date. Each index is capped at a maximum of 0% of baseline or the historical maximum (whichever is higher), and -50% of baseline or the historical minimum (whichever is lower) to maintain reasonable estimates of the trend. These forecasts can be seen in Appendix C.

Using the posterior distribution of the parameters found in the inference of data from March and calibrated against the effective reproduction number R_{eff}, and the forecasted Google Mobility Indices, we then generate posterior predictive distributions of R_L over time using (5.2) for each jurisdiction.

5.1.5 Generative simulation model

We simulate the daily number of cases using a branching process based on the estimated reproduction number described above. The generative model contains three types of infectious individuals:

- Imported (I_I),
- Asymptomatic (I_A), and
- Symptomatic (I_S).

Given the index cases as initial conditions, we simulate the number of cases as a continuous-time branching process.

5.1.6 Secondary case distribution

Let α_S correspond to the contribution of transmissibility of symptomatic local cases and α_A correspond to the contribution of transmissibility of asymptomatic local cases. Each case is assumed to generate a number of cases drawn from a Negative Binomial distribution, with parameters k and, respectively, $R_I/(R_I+k)$, $\alpha_A R_L(t)/(\alpha_A R_L(t)+k)$ and $\alpha_S R_L(t)/(\alpha_S R_L(t)+k)$ for imported, asymptomatic and symptomatic local cases. The parameters R_I and $R_L(t)$ (the Effective Reproduction Numbers) are sampled from the posterior distributions described above. Using Equation (5.2), a single sample from the posterior generates $R_L(t)$ for any t and jurisdiction. Each simulation then uses a single sample of the posterior. The parameter k has been set equal to 0.1 in our analysis [35]. This value allows for heterogeneity in the transmissibility of individuals – so-called *super spreading* – in that the mean is realised with high variance.

The R_{eff} estimate inferred from the NNDSS case data does not readily distinguish between symptomatic and asymptomatic cases, and index cases in the initial outbreak are all assumed to be symptomatic. The effective reproduction number is the average number of secondary infections caused by an infected individual, and can be characterised as

$$R_{\text{eff}} = \frac{s_{t+1}}{s_t}, \tag{5.5}$$

where s_t is the number of detected symptomatic cases in generation t.

In order to correctly attribute the contributions to secondary cases between symptomatic and asymptomatic cases, we require

$$s_{t+1} = (S_t \alpha_S R_L + A_t \alpha_A R_L) p_S q_S, \tag{5.6}$$

where S_t is the actual number of local symptomatic cases, A_t is the actual number of local asymptomatic cases, p_S is the probability of being symptomatic and q_S is the probability of detecting a local symptomatic case. Using Equations (5.5) and (5.6), and for local cases where $R_{\text{eff}} = R_L$, we have

Table 5.1: The assumed detection probabilities of Symptomatic, Asymptomatic and Imported cases for each jurisdiction.

Jurisdiction	q_S	q_A	q_I
NSW	0.80	0.05	0.95
QLD	0.90	0.05	0.95
SA	0.70	0.05	0.95
TAS	0.40	0.05	0.95
VIC	0.75	0.05	0.95
WA	0.70	0.05	0.95
ACT	0.95	0.70	0.95
NT	0.95	0.70	0.95

$$\alpha_S p_S + \alpha_A(1 - p_S) = 1. \tag{5.7}$$

We set the probability of a case being symptomatic p_s to 0.7 and the relative infectiousness of asymptomatic cases α_A to 0.5 of symptomatic cases, consistent with earlier studies [15, 89, 100]. Equation (5.7) then determines that $\alpha_S = 1.18$ and $\alpha_A = 0.59$.

The generative model must also consider probabilities of observing infectious cases. Infectious individuals are detected, and hence become a case, with probabilities q_I, q_A and q_S respectively. See Table 5.1 for the values used in this forecast, and are chosen based on values determined in Price *et al.* (2020) [95].

We additionally assume a $\text{Poi}(\lambda_t)$ number of new imported infectious individuals on day t, where the mean arrival rate λ_t is a piecewise constant function inferred from data for each jurisdiction. Please see Appendix C for further details.

For each infected individual, we generate the time that they became infected by adding to the time of infection of their parent. For those that are detected and symptomatic, we also need to add on to their time of infection the delay until symptom onset. These times are sampled from the generation interval and incubation period defined earlier, and taken from [72] and [42], following [1].

5.1.7 Initialisation of model

Initialisation is based upon the *very early* stages of the outbreak. This is because we assume that in the early stages the cases observed make up a large proportion of initial infections, and hence considering these as an initial generation is reasonable.

The generative model is therefore initialised on 01/03/2020, where there were less than three local cases in any jurisdiction and transmission was largely driven by importations. The model is simulated up to the end of the forecast.

Specifically:

- Given n_S symptomatic local cases, and n_I imported cases with symptom onset times on 01/03/2020, we generate:
  - Undetected symptomatic individuals, $U_S \sim \text{NegBin}(n_S, q_S)$ and undetected imports $U_I \sim \text{NegBin}(n_I, q_I)$; and,
  - Asymptomatic individuals, $I_A \sim \text{NegBin}(I_S, p_S)$.
- Assign an infection time to the U_S, U_I and I_A individuals, by subtracting a sample from the incubation period distribution.

5.1.8 Conditioning on data

Events that are difficult to forecast precisely will occur, and did occur in Australia in 2020. These are typically initial cluster outbreaks localised to small regions that previously had little epidemic activity. However, once occurred, forecasts must be conditioned on these clusters occurring as observed. Examples in Australia include those in Tasmanian hospitals and in various settings in Victoria.

When such outbreaks occur, detected via the cumulative (over a moving 3-day period) NNDSS cases exceeding our forecasted cases by a threshold of ten times the number of cases occurring in the last three days of the simulation before the current time point, we add to our simulation state additional cases determined by performing the initialisation step immediately above on the day the threshold is exceeded. The additional n cases are distributed across the 3-day period by adding $n/6$, $n/3$ and $n/2$ to each corresponding day respectively, with the largest applied on the current day. This occurrence is simulation dependent, but will coincide with the initial cluster outbreaks observed in some jurisdictions.

The model also allows for the re-introduction of cases when there has been an extended period of no local cases. If the simulation does not have any infections occurring in the previous seven days, and the NNDSS data does on this particular day, then the missing cases are added to the simulation state as described above. This ensures that when active cases have ceased in the simulation, they can be reintroduced when observed.

Simulations are only permitted to re-initialise a maximum of ten times. If this is exceeded, the simulation is excluded from the forecast. This ensures parameters used in the forecast are reasonable and forecasts are not simply a product of endless re-initialisations resulting in rare realisations that are not representative of the parameters.

Where the number of cases in a simulation exceeds a threshold set for that jurisdiction, the simulation is excluded from the forecast. For convenience, we define the *data date* as the date at which observations from NNDSS are provided up to, and the date at which the forecast begins.

Before the data date, the maximum threshold in the most recent 60 days up to the data date is 100 local cases, or 4 times the number of local cases, whichever is higher. If this is not exceeded, there is a further condition for the most recent 14 days up to the data date to not exceed ten cases, or 1.5 times the number of local cases, whichever is higher.

If a simulation is excluded, either by exceeding the number of re-initialisations or exceeding the maximum number of cases, the corresponding sample from the posterior predictive distribution of $R_L(t)$ is then also excluded, and not included in the forecast. Simulations continue until at least 2000 simulations are not rejected

5.1.9 Forecast validation

The forecast is validated and its performance measured using the continuous probability ranked score (CRPS), which is a proper scoring rule for probabilistic forecasts and a generalisation of the mean absolute error [13, 60]. The forecast is also compared to a reference forecast of random walk with drift [81] using a skill score [86], which is trained on the most recent 14 days up to the data date.

5.2 Results

The whole pipeline was applied at multiple dates to examine the performance of the forecast at various stages of the epidemic: fade-out during lockdown, local outbreak, and ongoing transmission.

We will focus on the two most active jurisdictions, New South Wales (NSW) and Victoria (VIC), at 3 key dates: July 1st 2020, Aug 5th 2020, and Sept 2nd 2020. These dates mark the beginning, peak and end of the second wave in Victoria respectively, along with low levels of sustained transmission in New South Wales. For forecasts of the other jurisdictions, please see Appendix C.

5.2.1 Inference

The posterior distribution of the effective reproduction number using historical case data is shown in Figure 5.2. In jurisdictions where cases arrive regularly, we can see a clear relationship between the arrival of cases and the trend of the reproduction number, such as in NSW and VIC. However, as seen in TAS and SA, when there are no cases for an extended period of time, the estimate tends to revert to the prior distribution. Given not all cases have been observed in the most recent ten days, there is also a noticeable decline in the effective reproduction number in those ten days.

The marginal posterior distribution for the influence of each mobility metric on the effective reproduction number in the social mobility model can be seen in Figure 5.3. The mobility indicator with the greatest influence on the effective reproduction number is visits to retail and recreation locations. The effect of grocery and pharmacy is inferred to be negatively correlated to the effective reproduction number, and this might be explained by the various episodes of 'panic buying', where the public rush to purchase basic needs as a local outbreak is announced [30]. This effect is minor relative to the effect of visits to retail and recreation.

The general behaviour of individuals changes over time and in response to the localised outbreaks and sentiment towards outbreaks in other jurisdictions. Figure 5.4 depicts the inferred effect of micro-distancing behaviours on the effective reproduction number over time. After the initial lockdown in March, there is a waning effect over time as cases reduced. However, as VIC began to see a local outbreak develop into a second wave of cases in June, the jurisdiction was observed to increase micro-distancing behaviours, resulting in a reduction in the posterior estimate of the effective reproduction number.

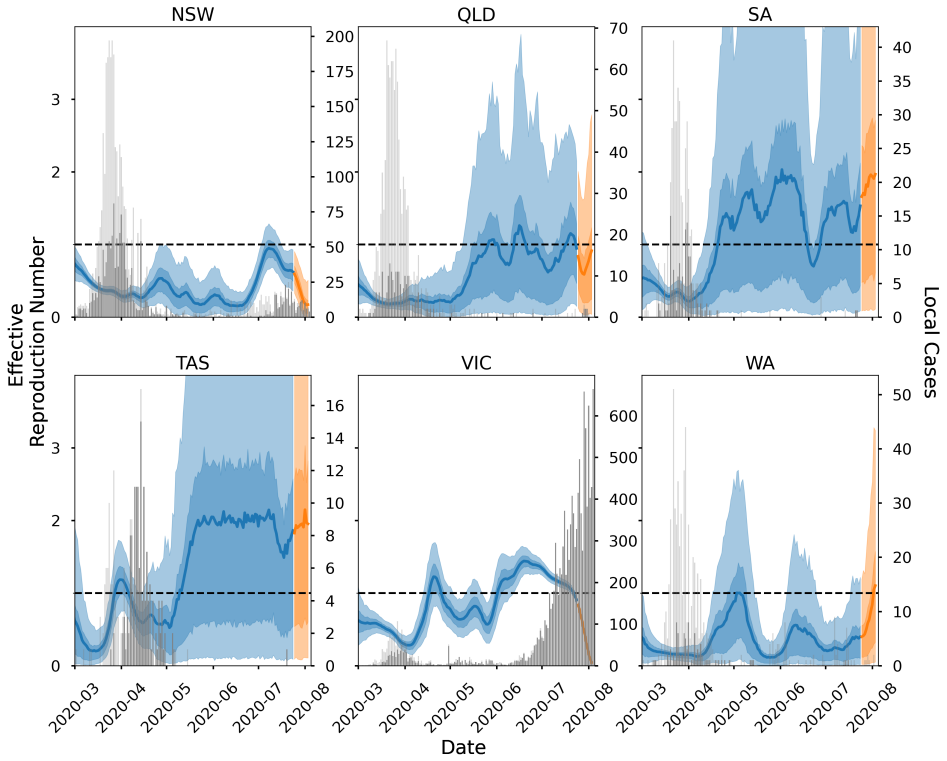

Figure 5.2: The posterior distribution (blue) of the effective reproduction number R_{eff} estimated using case data from each State in Australia as of August 5th 2020, with 50% and 90% credible intervals shaded. The most recent 10 days are shown in orange, and are excluded from calibrating the social mobility model as not all cases to occur on those dates have been observed yet. The number of local (dark grey) and imported cases (light grey) is shown plotted by symptom onset data.

See Figure C.6 for the proportion of surveyed individuals that observe micro-distancing behaviours over time.

The hierarchical model allows for insight into the differences in base level reproduction number in each jurisdiction R_{Li}. From Figure 5.5, VIC, TAS and NSW are inferred to have higher marginal posterior distributions of the baseline reproduction number. This can be attributed to the clusters that occurred in these two jurisdictions resulting in wider local outbreaks. Other jurisdictions have not seen localised cases lead to further outbreaks.

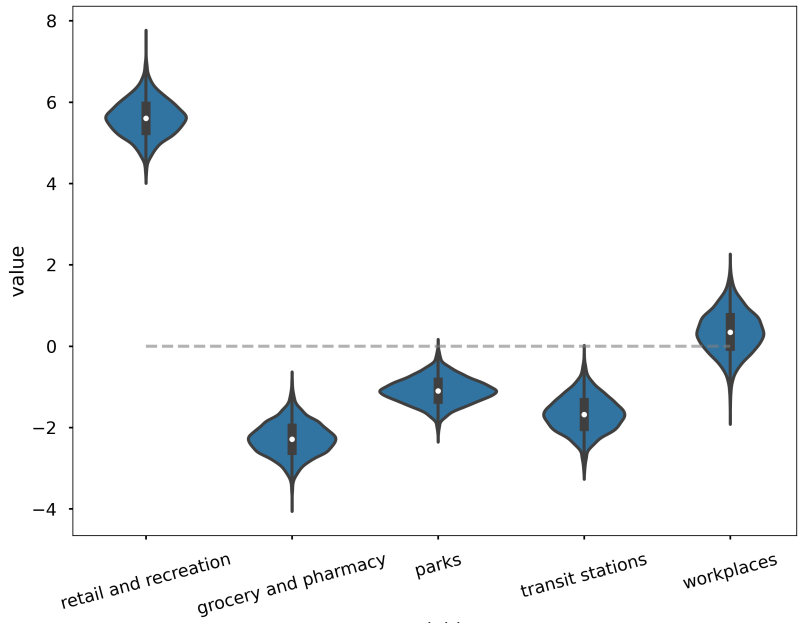

Figure 5.3: The marginal posterior distributions for the weight of the influence of social mobility metrics on their effect on the effective reproduction number, as of 2020-08-05. Positive values are positively correlated with the effective reproduction number.

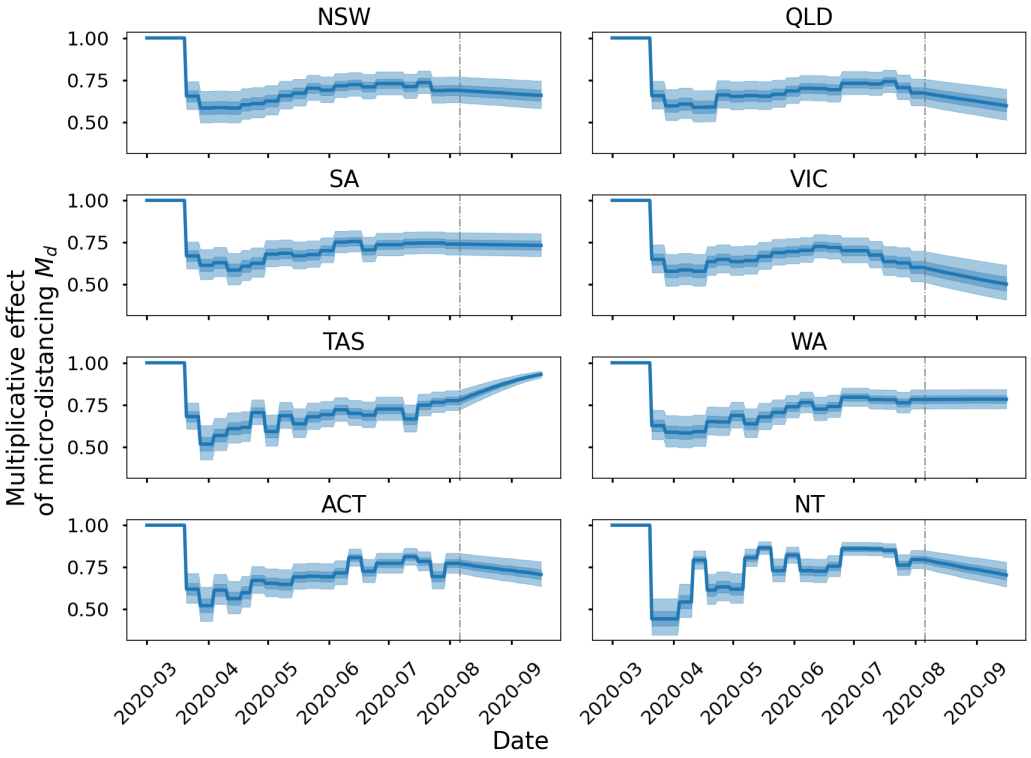

Figure 5.4: Posterior predictive distribution of the proportional effect of micro-distancing M_d on the effective reproduction number in all Australian jurisdictions over time, as of 2020-08-05 (Vertical dashed line. Solid line represents the median value, with dark and light shading the 50% and 90% credible intervals.

5.2.2 Forecast results

The CRPS of the forecast tends to increase, and therefore predictive accuracy reduces, over the forecast horizon, which is to be expected. However, when examining the forecast skill score, the forecast performs well over the entire forecast period (Figure 5.7).

The forecast at the data date of July 1st 2020 is the least predictive compared to the random walk, and this may be due to the sudden onset of the second wave in Victoria, where cases were initially driven by and highly concentrated around a very small geographical area of nine tower blocks of 3000 residents [105]. Disease transmission was not driven by the jurisdiction-wide trends in mobility that this model uses, but by local differences. The model is able to learn from this outbreak by the August 5th forecast (Figure 5.6),

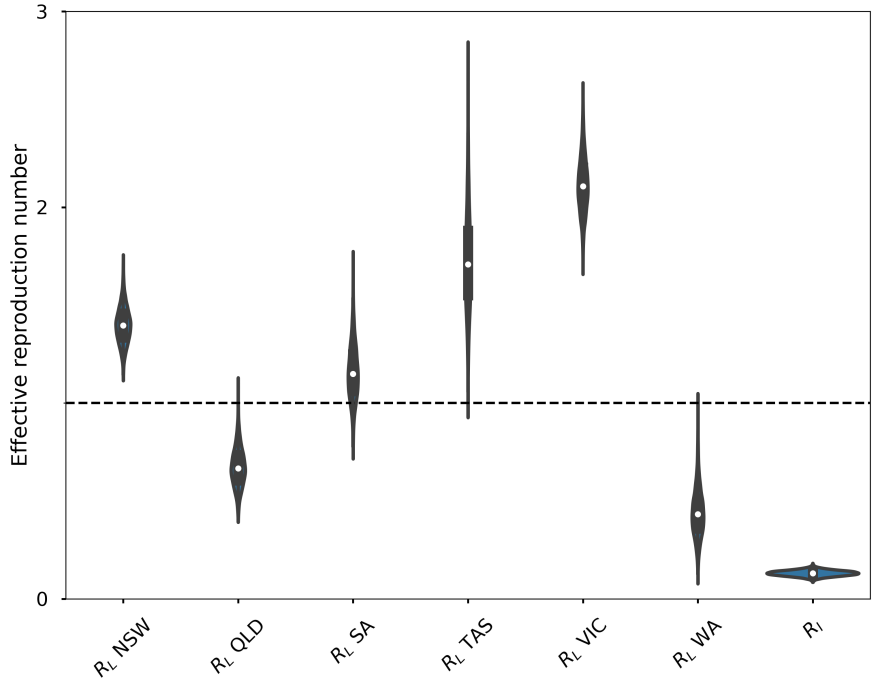

Figure 5.5: Marginal posterior distributions of the effective reproduction numbers of each state R_L and the effective reproduction number of imported cases R_I, as of 2020-08-05.

where we can see that it has captured the dynamics of transmission locally in Victoria, given a corresponding increase in the posterior prediction for the effective reproduction number. Note that throughout July, Victoria steadily increased its restrictions on social mobility, culminating in a state of disaster declaration in early August, where only essential food purchases and essential work were allowed, with residents to otherwise remain at home. The forecast model is able to capture these drastic and severe changes in social mobility.

For all forecasts, even during a highly localised event, we see that the model is at least as predictive as the reference model, and excepting the sudden onset of a local cluster, performs extremely well once sustained transmission occurs or after changes in mobility in response to an outbreak, either by policy or individual behaviours.

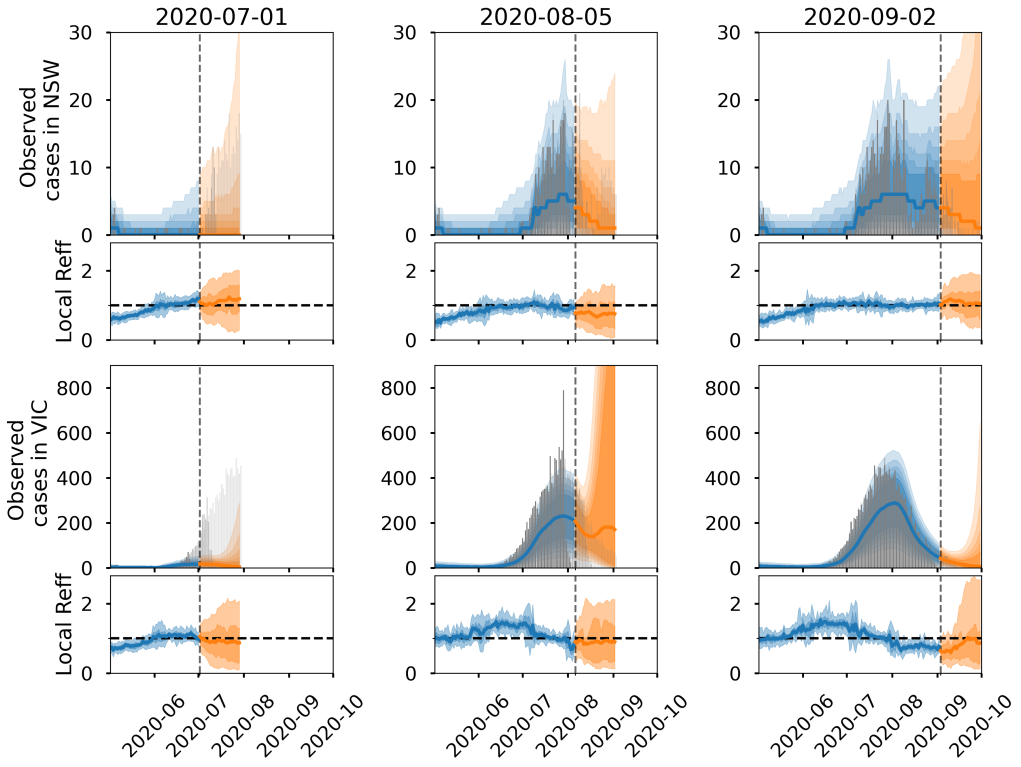

Figure 5.6: The nowcast (blue) and forecast (orange) of observed local cases and the local effective reproduction number in New South Wales and Victoria, using case data at three different dates. The shading represents the 90%, 80%, 70%, 60% and 50% quantiles. The dark grey bars represent the number of observed local cases known as of the data date, the medium grey bars represent the case data known 14 days after the data date, and the light grey bars represent the known cases as of November 2nd 2020.

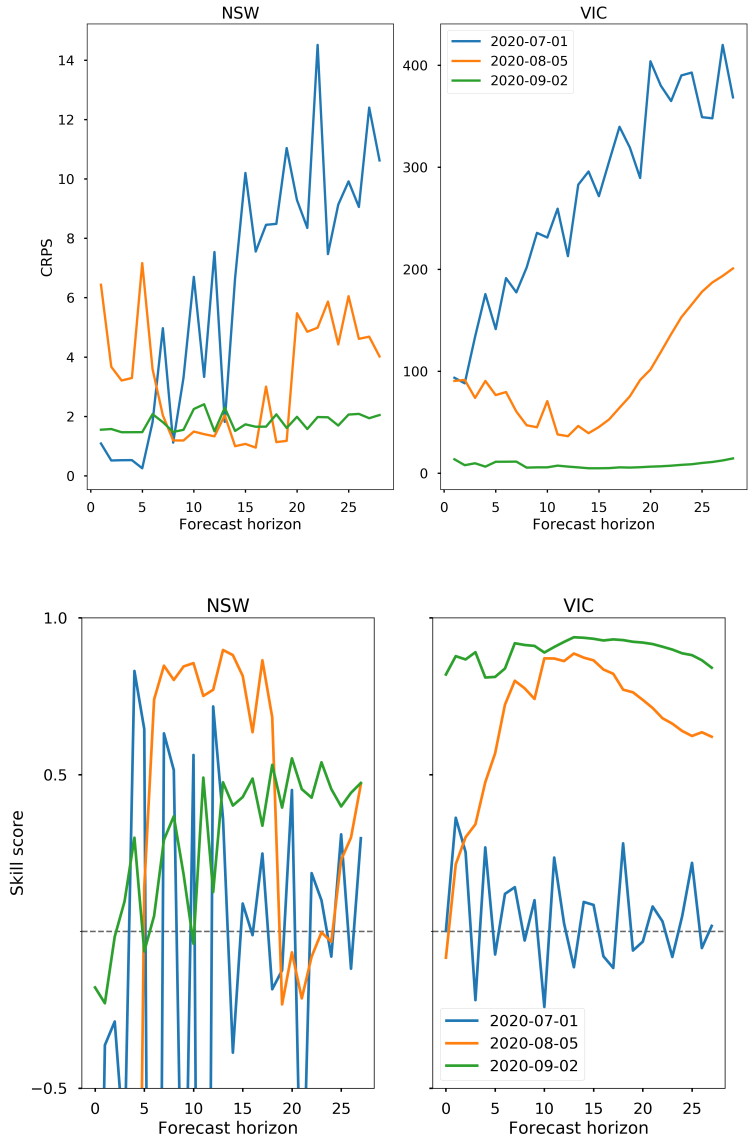

Figure 5.7: The continuous ranked probability score (CRPS) and the forecast skill of the New South Wales (NSW) and Victoria (VIC) forecast performance for each of the data dates of interest - July 1st, August 5th, and September 2nd.

5.3 Discussion

The most recent ten days up to the data date is excluded from calibrating the social mobility model in Equation (5.2), which reduces the information used in this framework. This is plotted in orange in Figure 5.2. Inference on the historical effective reproduction number R_{eff} relies on an unbiased estimate of the number of cases observed. Using this estimate near the data date results in a biased, lower estimate of the reproduction number, as not all cases infected near the data date have been observed by the data date. Not using the most recent data and adjusting for this censoring contributes to this model's poorer performance during the onset of an outbreak. Once observed, the model is able to update and appropriately forecast the occurrence of an outbreak. Methods for accounting for the censoring of observations near the data date is left to further work, and is a current area of research [50].

The social mobility metrics used in the model are aggregates across entire jurisdictions, and may not be representative of those currently infected. There may be a large amount of heterogeneity in individual mobility patterns, resulting in poor predictions when given a low number of active cases. However, if cases rise to a larger number, then transmission begins to resemble the jurisdictional level dynamics. This can be seen in the model performance when comparing the July 1st forecast to the August 5th and September 2nd forecasts.

Simulations are initialised with the number of cases observed on the first day, and undetected cases simulated given these initial cases. This analysis used an initial date where there were very few local cases to limit the influence of the initial case distribution on the simulation path. This enables consistency with the initial epidemiology being driven by importations from international arrivals. The influence of the assumptions around the initial cases is yet to be determined. Inference on these assumptions, as well as a number of others, could be done through an Approximate Bayesian Computation (ABC) framework. This may be computationally challenging, but allows further inference in fitting to observed cases.

The model was compared to a random walk with drift [81] in this analysis. Comparisons to more sophisticated epidemiological and non-mechanistic models is possible with CRPS. These comparisons are left for further work.

Code access

All code can be accessed at the repository

`https://github.com/tdennisliu/covid19-forecasting-aus`.

Chapter 6

Evaluating the effectiveness of contact tracing strategies and performance metrics for COVID-19

During any outbreak in the COVID-19 epidemic the purpose of testing is to identify cases and initiate public health responses to reduce onwards spread of infection. The most common approach to reducing transmission during incursions is to identify and isolate close contacts before they transmit COVID-19 further into the community. If this can be achieved in a timely and rigorous manner, then overall transmission may be reduced to a controllable level or potentially eliminated.

Identification of close contacts can vary by the period of time to trace back over and the number of generations to trace. For example, tracing two generations of close contacts identifies close contacts of the close contacts of the index case. However, it is not known for COVID-19 whether increasing the period of time to trace back or increasing the number of generations of close contacts traced provide greater value for effort in minimising outbreak size or elimination probability in a capacity constrained environment. Given the need for both high levels of detection and timely isolation of cases to ensure control [56, 70, 101], the aforementioned aspects of contact tracing must be balanced appropriately. Which aspects to sacrifice may be dependent on the epidemic stage, or the varying timeliness of different public health responses.

In Chapter 5, we constructed a framework to forecast COVID-19 cases in each

Australian jurisdiction. In this chapter, we extend this work to an assessment of the public health response, and to see how variation in the extent and timeliness of case isolation and contact tracing influence transmission. We then examine, as a case study, performance metrics utilised by the Australian Government on determining the extent of control on current local outbreaks.

To help determine if local outbreaks are under control and appropriately addressed by each jurisdiction in Australia, the Common Operating Picture (COP) [5] was established to aid in consistent comparisons of performance across the different populations in Australian jurisdictions. Here we will focus on the performance metrics as they relate to contact tracing and test turnaround time, where rapid detection and early isolation is targeted to aid in disease control.

The content of this chapter forms part of a research contract with the Commonwealth of Australia to inform their COVID-19 response.

6.1 Model overview

Here, we adapt the branching process model (see Section 5.1.5) to evaluate the effectiveness of contact tracing strategies, and the use of various performance metrics as a case study. The performance metrics used here are utilised by the Department of Health of the Australian Government at the time of writing. We compare tracing efforts of differing rigour and evaluate the utility of current green/amber/red Common Operating Picture [5] indicators of public health response capacity to predict the likely effectiveness of interventions[1]. Our branching process model is used to simulate the course of an epidemic with theoretically infinite population size. Chains of disease transmission are tracked over a period of 30 days for each simulation.

Three epidemic scenarios are considered, relating to a minor incursion, a major incursion, and established community transmission. Table 6.1 shows the conditions used to initialise each scenario in the model. The primary goal is elimination for the incursion scenarios. In the case of active community transmission, epidemic suppression is the primary goal, but elimination remains desirable.

For each epidemic scenario, different contact tracing assumptions are explored, with a range of values considered for the routine detection probabil-

[1] At time of writing, we acknowledge that these indicators are currently in review and may change in the future.

Table 6.1: Epidemic scenarios and corresponding initial conditions examined.

Scenario	**Initial cases on day 0**	**Initial undetected cases**
Minor incursion	6	No
Major incursion	12	No
Active community transmission	60	Yes

ity (0.3, 0.5 and 0.8 of true cases identified by spontaneous presentation to services). Mild and asymptomatic case detection probabilities are set to 1/5 of the symptomatic detection probability. Contact tracing efficiency, defined as the probability a close contact of an index case who is positive will test positive and be identified by contact tracing at any time within the duration of backtracing, is assumed to be 0.9 for all results presented here. Minor changes to this probability did not significantly alter the results.

The branching process model (Figure 6.1) is used to model the transmission process of COVID-19, as developed in Chapter 5. The proportion of asymptomatic cases and their relative transmissibility is unchanged from that work. Initial, detected index cases are seeded and observed on day 0, and their symptom onset and infection dates drawn from samples of the incubation period (T_S) and the generation interval (T_G) used previously in the probabilistic forecasting model. 20% of initial cases are asymptomatic. The number of offspring of each infected individual follows a negative binomial distribution with mean equal to the effective reproduction number, which is sampled for each individual from a distribution with mean 1.2 (Figure D.1). The effect of the length of time prior to symptom onset that is contact traced (τ_k), or backtracing, is two days at baseline assumptions.

The delay between symptom onset and contact tracing action was modelled as four stages:

1 Presentation T_1 (time between symptom onset and time of swab);

2 Test turnaround T_2 (time between swab and time of laboratory result);

3 Notification T_3 (time between laboratory result and time of health authority notification);

4 Contact tracing action T_A (time between health authority notification and time contact traced).

Unlike the previous chapter, index cases are identified with probability equal to the routine detection probability, and then contact tracing initiated on

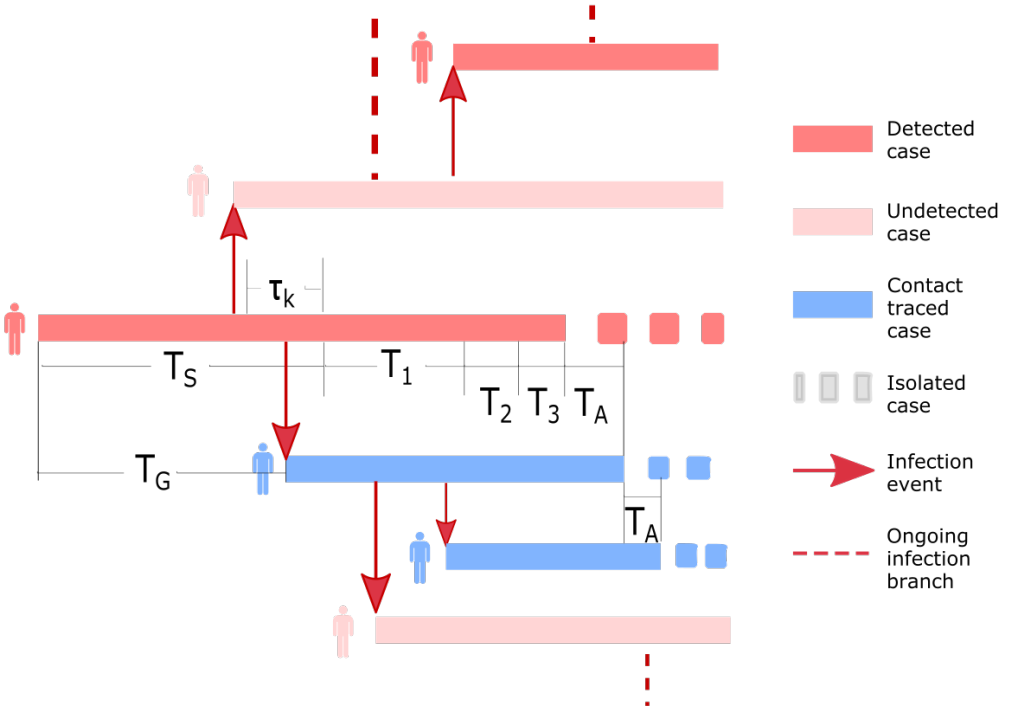

Figure 6.1: Diagram of branching process model, where each horizontal bar is a positive case proceeding through time. In this diagram we show two generations being traced (contacts of contacts). The time for an index case to present for a test (T_1), test turnaround time (T_2), and notification of the index case (T_3) all contribute to delays towards action and isolation. The action of notifying and isolating of close contacts occurs after a minor period of time (T_A). The generation interval (T_G), the incubation period (T_S), and length of contact tracing ($\tau_k + T_1 + T_2 + T_3$), and isolation time (after T_3) are all marked for the index case. Positive cases may be missed by being outside the contact tracing window (upper pink bar), but there is a probability that positive cases may be detected via routine detection and surveillance (upper red bar). Close contacts may also be missed due to the contact tracing efficiency (lower pink bar).

Table 6.2: Summary of gamma distribution parameters for delay distributions used, their performance targets, and modifications to the gamma distributions to align with the performance targets.

Delay	**Inferred Gamma Parameters**	**Green COP Target**	**Modifications**
1 Presentation	shape = 1.03 rate = 0.42	Not applicable.	Condition: each sampled value is less than 7 days.
2 Test turnaround	shape = 7.91 rate = 4.86	2 and 3: 90% within 48 hours. 3: 100% within 24 hours.	Scaled by a factor of 0.75.
3 Notification	shape = 0.45 rate = 1.82		Scaled by a factor of 0.6
4 Contact tracing action	Not applicable.	100% within 48 hours.	Parameters selected as shape = 0.45, rate = 1.82.

the index case. Close contacts within the tracing window ($\tau_k + T_1 + T_2 + T_3$) traced with probability equal to the contact tracing efficiency.

The time distributions for delays 1, 2 and 3 were informed using data from the National Notifiable Diseases Surveillance System (NNDSS) and adjusted to match the green traffic light targets from the Common Operating Picture (COP)[5]. To ensure that the distributions reflect current operations, the NNDSS dataset was limited to include only cases occurring after 31/08/20, at the end of the second wave in Victoria. In the absence of individual records for delays in contact tracing action (Delay 4), a gamma distribution was assumed with parameters selected to align with both the green traffic light COP target and the New South Wales COVID-19 Weekly Surveillance Reports [88]. The resulting distributions and the COP targets assessed here are listed in Table 6.2.

The distribution of time for presentation for a test was truncated to 7 days. This is based on the assumption that the majority of individuals who will choose to present for a test will do so within 7 days of symptom onset, and those who did not were mildly symptomatic and were identified through contact tracing. These latter individuals would not have presented for a test without being identified through contact tracing. The NNDSS data does not differentiate whether a case was identified through routine surveillance or through contact tracing.

The model operates with the following assumptions:

- Cases identified through routine detection are assumed to isolate after the public health unit (PHU) issues the notification of a positive case.
- Cases identified through contact tracing are assumed to isolate after the PHU conducts the contact tracing interview of the parent case and notifies the close contacts.
- A case that is missed through contact tracing may still be picked up in routine detection (they are independent processes).
- The effective reproduction number is distributed around a mean of 1.2, with 90% credible interval [1.02, 1.43] (Figure D.1).
- Testing and tracing capacities, where applied in simulations, are set such that for every subsequent case on the same day of notification, they are pushed back to the next day.
  - Every identified (traced or routine detected) positive case adds one to the testing queue, with capacity set to 500 per day.[2]
  - Every identified positive case adds either five close contacts for each generation traced, or double their number of positive infections, whichever is larger, to the contact tracing queue.

6.2 Strategies

Using the delay distributions and the branching process model outlined above, we simulate the epidemic for each choice of routine detection probability, contact tracing efficiency, and epidemic scenario. Each choice is simulated 10,000 times, with the epidemic trajectory recorded for 30 days. In each set of 10,000 simulations, we prescribe various contact tracing assumptions:

1. **Baseline:** A baseline contact tracing intensity of first order close contacts made within 48 hours prior to symptom onset of the index case,
2. **Second order contacts:** As in 1, but second order close contacts (close contacts of close contacts) are also traced,

[2]While no information has been provided on capacity limits in jurisdictions of Australia, the State of Victoria was notably struggling with case load at approximately 500 cases per day in late July 2020, and so this is chosen as a reasonable limit for larger Australian jurisdictions.

3. **Increased backtracing:** As in 1, but an additional 24 hours of contact tracing of first order close contacts, totalling 72 hours prior to symptom onset of the index case, and
4. **High intensity tracing:** Both 2 and 3.

In the initial comparison of these four contact tracing assumptions, we set no capacity constraints, so that the number of cases has no influence on the delay distributions.

The four contact tracing assumptions above are also applied to the following four scenarios:

A) **Delayed presentation scenario:** For each of the four assumptions, the initial cases present for testing an additional 7 days later than in delay distribution 1. All other cases present for testing as described in delay distribution 1.

B) **Delayed presentation with capacities:** As in A), but where the testing and tracing capacities are set to 500 cases per day.

C) **Testing at capacity:** For each of the four contact tracing assumptions, the simulations are initialised with testing already at capacity, and hence the simulation will be in the amber zone of the COP metric for test turnaround times, before the initial cases are observed.

D) **Tracing at capacity:** For each of the four contact tracing assumptions, the simulations are initialised with contact tracing already at capacity, and hence the simulation will be in the amber zone of the COP metric for contact tracing notification times, before the initial cases are observed.

E) **Sensitivity to generation interval:** The generation interval is increased to match the distributions found in Ferretti *et al.* 2020 [38], and the sensitivity of the results for each of the four contact tracing assumptions to this distribution examined, with capacity limits active.

The relevance of the COP performance targets for capacity are then examined in the baseline contact tracing assumptions (1) for C), D) and E). The performance targets are defined in relation to the delay distributions in Table 6.2.

The total number of cases, COP performance metrics, and whether epidemic elimination occurred are recorded for each simulation. The elimination probability is defined as the proportion of simulations for each choice of parameters where no further infections occurred after 30 days, signalling the end of the outbreak within 30 days.

6.3 Results

6.3.1 Contact tracing strategies

Given no capacity constraints, we examine the relative change in elimination probability relative to the current tracing effort (assumption 1) with tracing assumptions 2, 3, and both 2 and 3 in Figure 6.2. Given no capacity limits, we find that tracing second order contacts is more effective than increased duration of testing, and that doing both has synergistic effects. However, it should be noted that the absolute improvement is marginal.

The marginal improvement is partially due to high variance in the number of secondary cases of COVID-19. There is a high probability of cases producing no offspring, but small probability of few cases producing many. This can lead to elimination (through stochasticity) regardless of choice of tracing policy.

We undertake a sensitivity analysis on the influence of delays to initial case detection, by comparing this level of contact quarantine intensity for cases presenting for tests within 7 days of symptom onset (Figure 6.2), or beyond 7 days (Figure 6.3 and 6.4) for initial cases, before reverting to previous assumptions on presentation times for tests of subsequent cases. However, there are only minor changes to the improvement in elimination probability across the tracing strategies when initial cases present at least 7 days after their symptom onset date.

When capacity constraints are active, as seen in Figure 6.4, the policy of tracing second order contacts can be detrimental, as seen in the Active Community Transmission scenario, as increased tracing effort leads to tracing capacity reached earlier and delays impeding control and isolation. In the current implementation of the base assumptions, tracing capacities are usually reached before testing capacities.

6.3.2 Systems under stress

We consider the impact of capacity constraints on the likely effectiveness of public health responses of different intensity, represented as a shift in the COP indicators from green to amber. It is plausible that testing capacity and the ability to notify cases or trace contacts may be differentially stressed in the epidemic response. Key indicators are therefore varied one at a time to assess their overall public health response effectiveness.

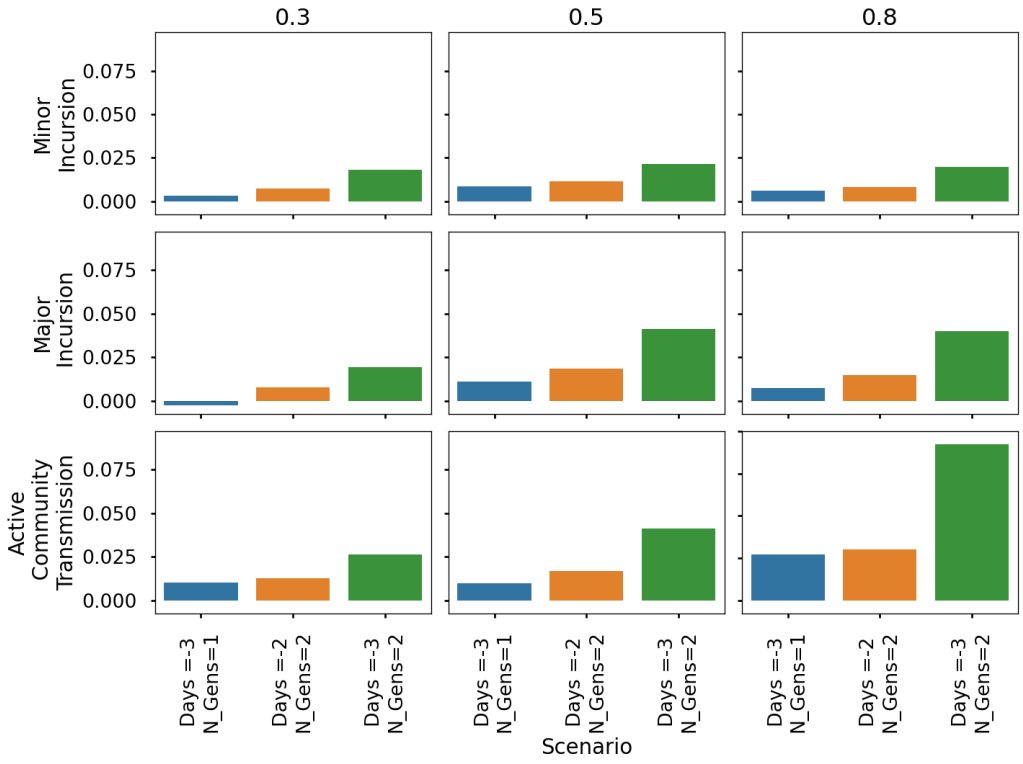

Figure 6.2: The marginal improvement in elimination probability after 30 days relative to first order contact tracing up to the preceding two days to symptom onset of the index case, given no capacity limits, at routine detection probabilities of 0.3, 0.5 and 0.8. The assumptions examined are increased duration of backtracing up to three days prior to symptom onset (Days=-3, Assumption 3, blue), tracing an additional generation of close contacts (N_Gens=2, Assumption 2, orange), and both of the former (Assumptions 2 and 3, green).

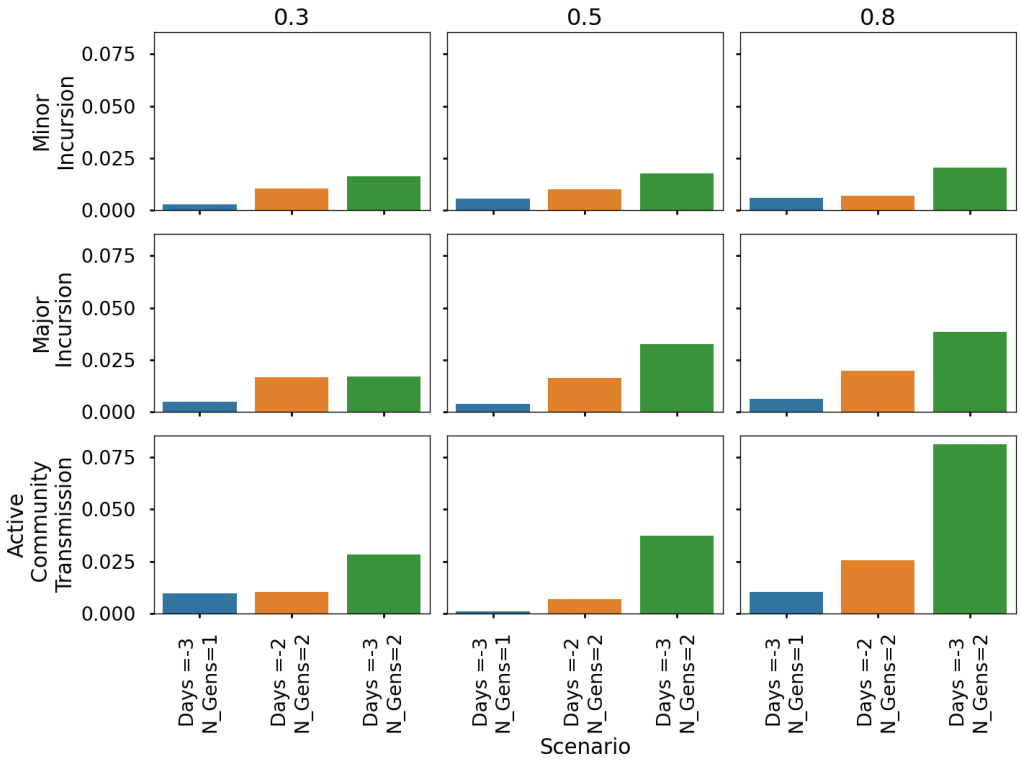

Figure 6.3: The marginal improvement in elimination probability given initial cases present for testing >7 days since symptom onset and given no capacity limits, relative to first order contact tracing up to the preceding two days to symptom onset of the index case, at routine detection probabilities of 0.3, 0.5 and 0.8. The assumptions examined are increased duration of backtracing up to 3 days prior to symptom onset (Assumption 3, blue), tracing an additional generation of close contacts (2, orange), and both of the former (2 and 3, green).

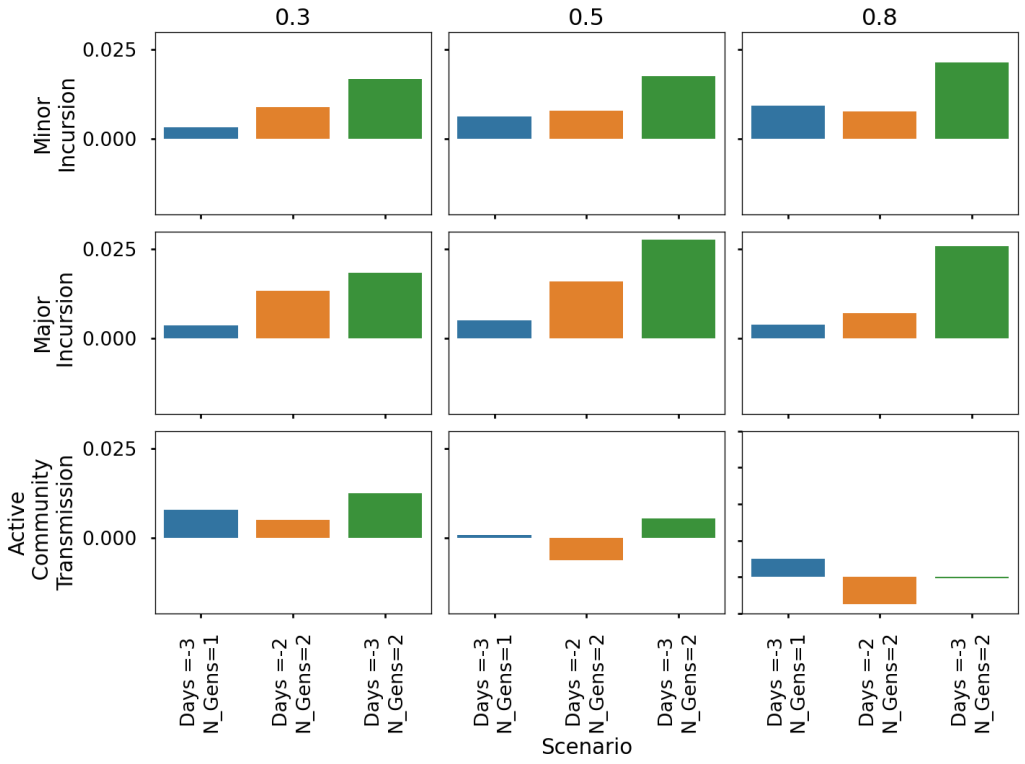

Figure 6.4: The marginal improvement in elimination probability given initial cases present for testing >7 days since symptom onset and with capacity limits imposed, relative to first order contact tracing up to the preceding two days to symptom onset of the index case, at routine detection probabilities of 0.3, 0.5 and 0.8. The assumptions examined are increased duration of backtracing up to 3 days prior to symptom onset (Assumption 3, blue), tracing an additional generation of close contacts (2, orange), and both of the former (2 and 3, green).

When testing is initialised at capacity, we see that while initially many simulations are in either amber or red alert, in both minor and major incursions the test turnaround COP metric is able to recover to green over time (Figure 6.5). During active community transmission, the large number of cases continues to escalate, resulting in higher and high proportions of simulations in the amber and red zones. The COP target for contact tracing also deteriorates over time as the epidemic grows. Reaching the amber and red zones at least once for the test turnaround time target is indicative of greater number of cases within 30 days, but not for the contact tracing target in the minor and major incursions.

If a simulation reaches a red alert, it will often indicate that the outbreak will reach a larger final size, but this is not always the case as seen in Figures 6.11 and 6.13. This unintuitive result can be explained when there are fewer cases.

For example, using a data driven approach from the NNDSS case data as in this study, the delay distribution for testing turnaround times is defined to have 0.9 probability to be less than 48 hours (see Figure D.3). If there are 3 cases in the week, and no issues with capacity, then each of the 3 cases has an independent 0.1 probability of exceeding the COP metric. The probability of at least 1 of the 3 cases exceeding the COP metric can be determined from a binomial distribution giving a 0.27 probability of occurring. In a week with 3 cases that are well under control, there is a 0.27 probability of giving a red alert and 0 probability of giving an amber alert. With 20 cases in a week, the red alert is reached when 3 or more cases exceed 48 hours. The binomial probability of this occurring is 0.04.

We see similar effects when instead of testing we have contact tracing initialised at capacity. Loss of public health response efficacy is seen in outbreak response settings when contact tracing times increase, demonstrated by amber status. Delays in testing exert their greatest impacts in the community transmission scenario, particularly when contact tracing capacity is stretched, as the ability to detect cases beyond tracing efforts is reduced, resulting in fewer constraints on transmission (see Figure 6.12).

We consider the sensitivity of the total outbreak size after 30 days on the presentation time of the initial cases. We initialise simulations using initial cases with presentation times of an additional 7 days plus a sample from the presentation delay distribution. In Figure 6.13, we see that the average number of cases across all scenarios are higher than in Figure 6.10, particularly in scenarios where control of the outbreak is lost.

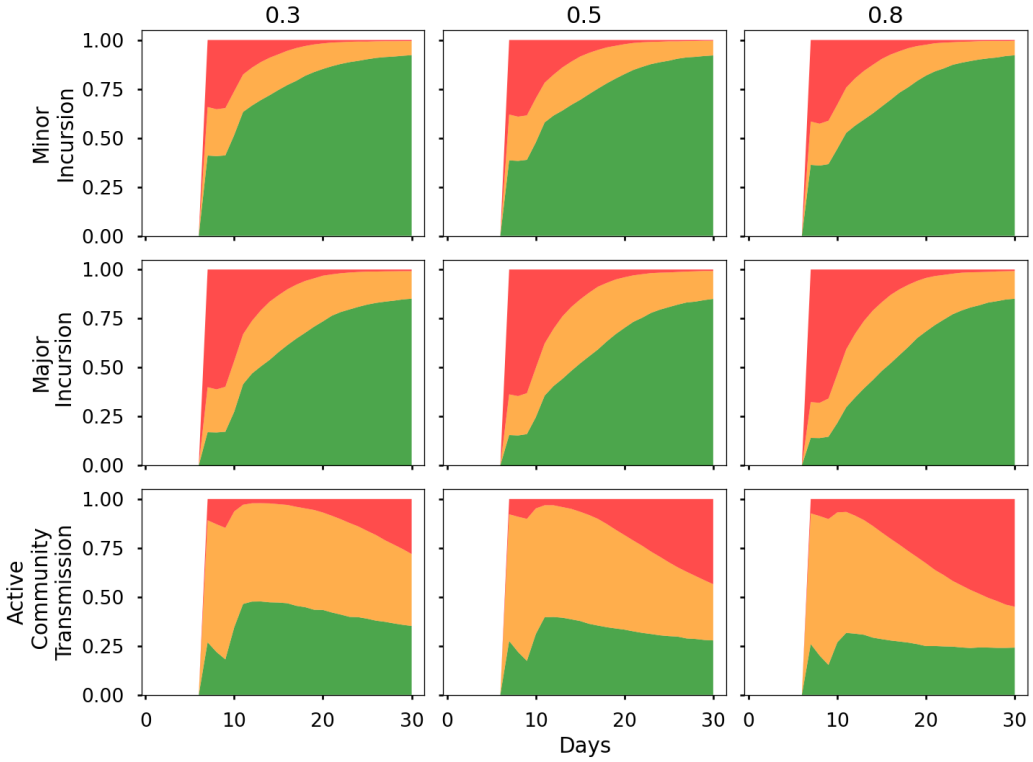

Figure 6.5: The proportion of the simulations at green, amber and red of the COP metric for providing test results of positive cases within 48 hours, given routine detection probabilities of 0.3, 0.5 and 0.8. Simulation initialised with tests to complete at capacity, while assuming first order contact tracing up to the preceding two days of symptom onset of the index case.

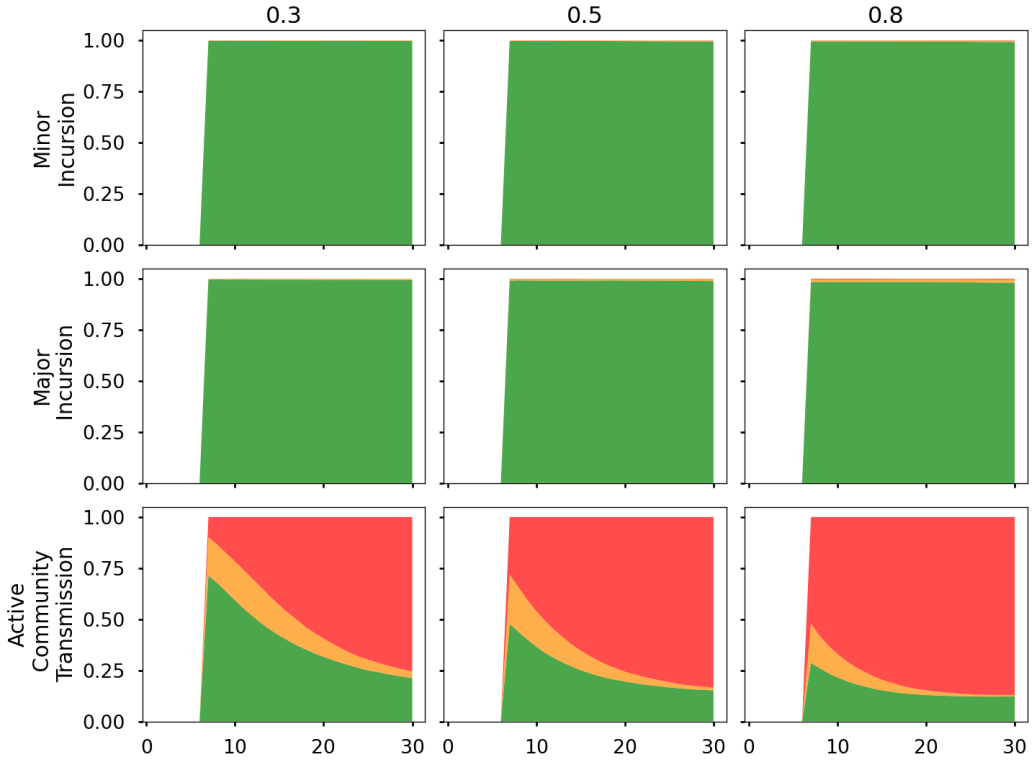

Figure 6.6: The proportion of the simulations at green, amber and red of the COP metric for contacting close contacts within 48 hours, given routine detection probabilities of 0.3, 0.5 and 0.8. Simulation initialised with tests to complete at capacity, while assuming first order contact tracing up to the preceding two days of symptom onset of the index case.

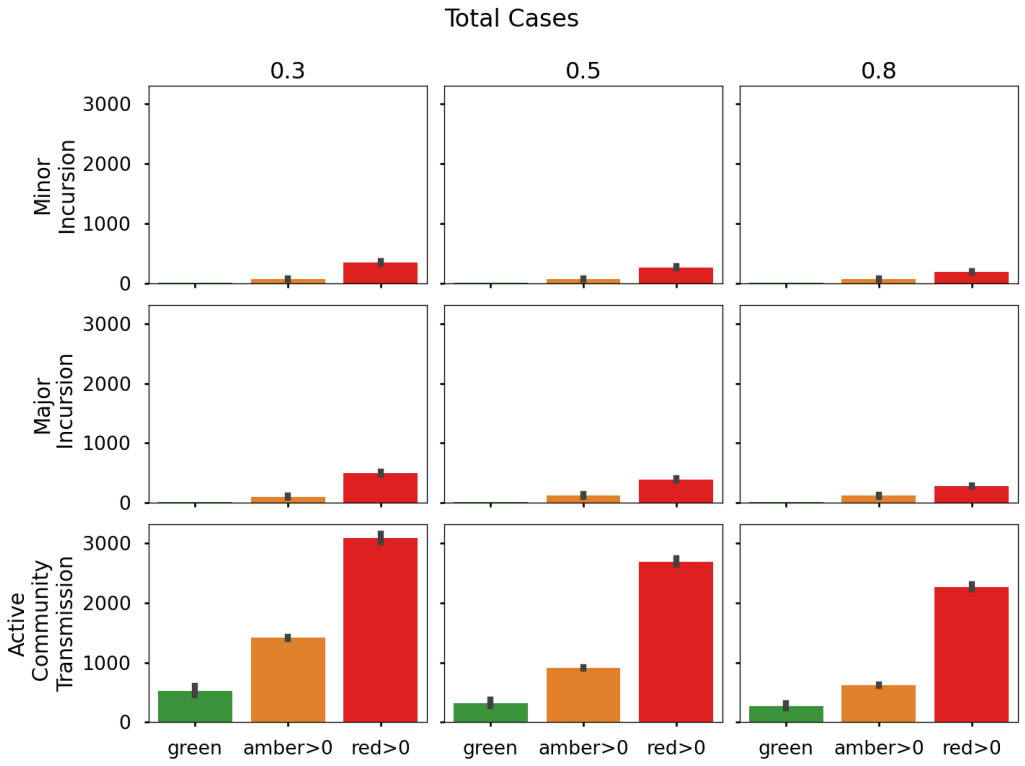

Figure 6.7: The mean total number of cases in simulations that never exceeded the COP metric for testing turnaround delays (green), exceeded at least once into amber alert but never red (amber >0) and exceeded at least once into red alert, given the simulation begins with testing at capacity and routine detection probabilities of 0.3, 0.5 and 0.8. Simulations assume first order contact tracing up to the preceding two days of symptom onset of the index case.

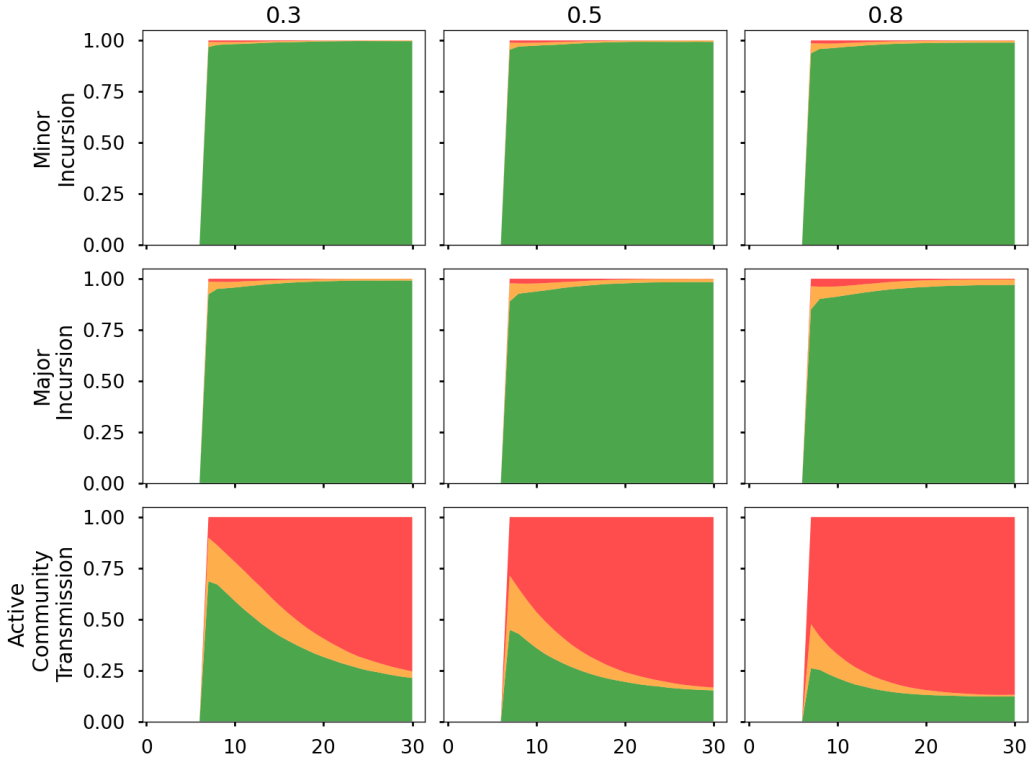

Figure 6.8: The proportion of the simulations at green, amber and red of the COP metric for contacting close contacts within 48 hours, given routine detection probabilities of 0.3, 0.5 and 0.8. Simulation initialised with tracing to complete at capacity, while assuming first order contact tracing up to the preceding two days of symptom onset of the index case.

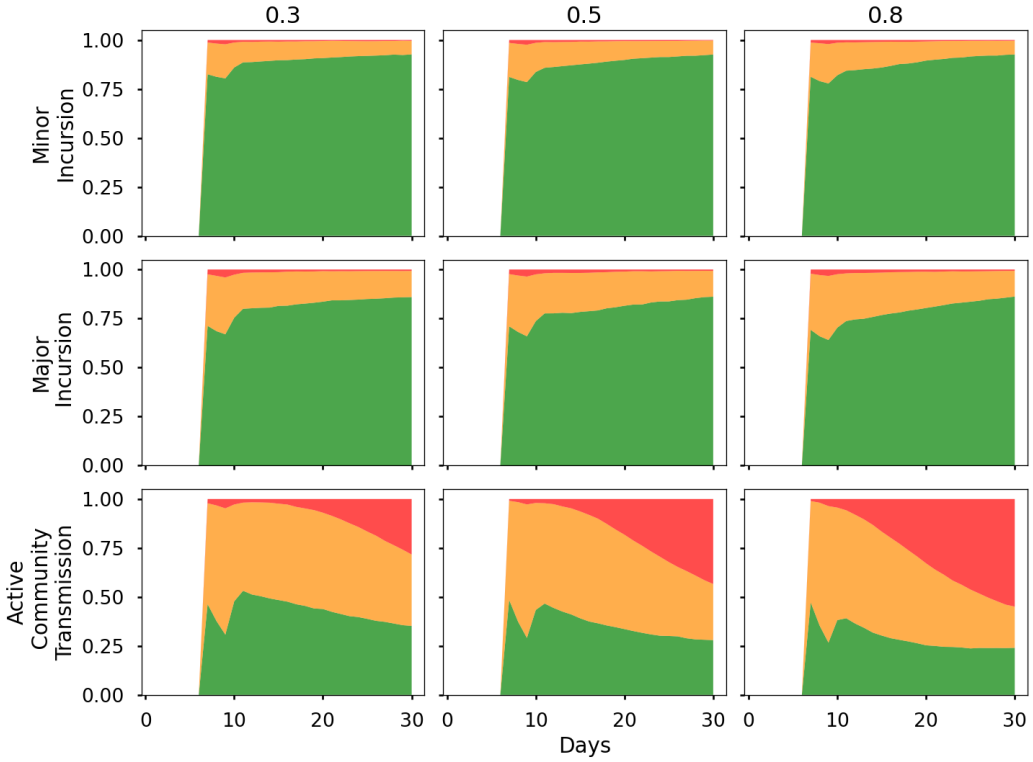

Figure 6.9: The proportion of the simulations at green, amber and red of the COP metric for testing delays within 48 hours, given routine detection probabilities of 0.3, 0.5 and 0.8. Simulation initialised with tracing to complete at capacity, while assuming first order contact tracing up to the preceding two days of symptom onset of the index case.

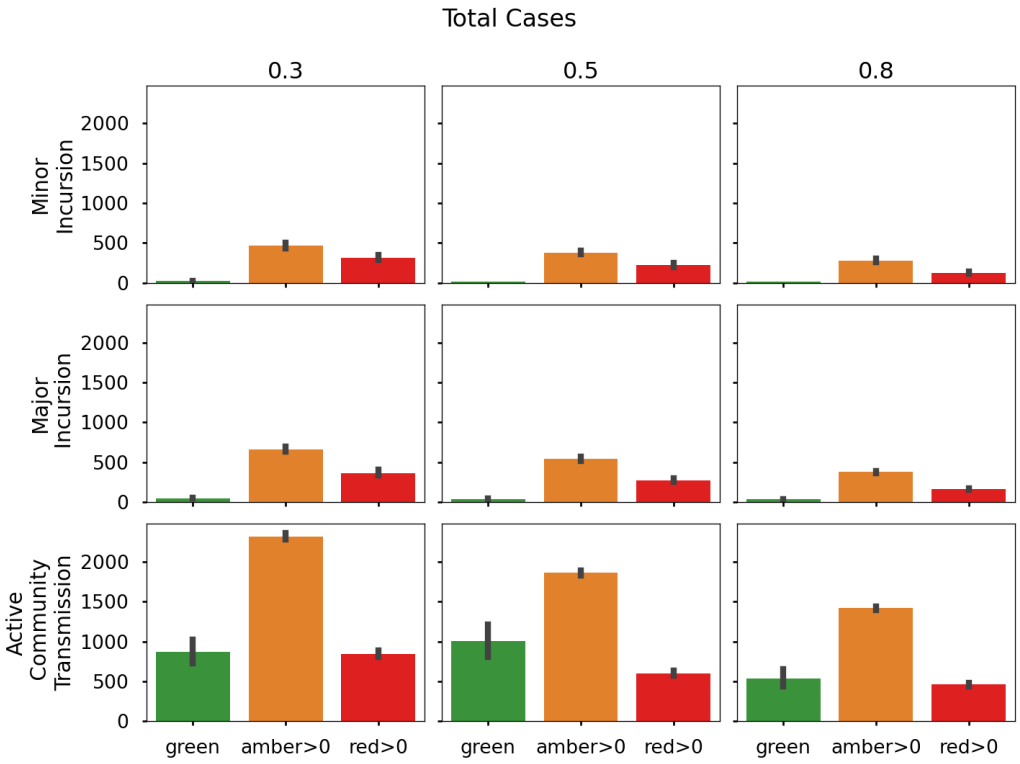

Figure 6.10: The mean total number of cases in simulations that never exceeded the COP metric for tracing delays (green), exceeded at least once into amber alert but never red (amber >0) and exceeded at least once into red alert, given routine detection probabilities of 0.3, 0.5 and 0.8. Simulations use first order contact tracing up to the preceding two days of symptom onset of the index case.

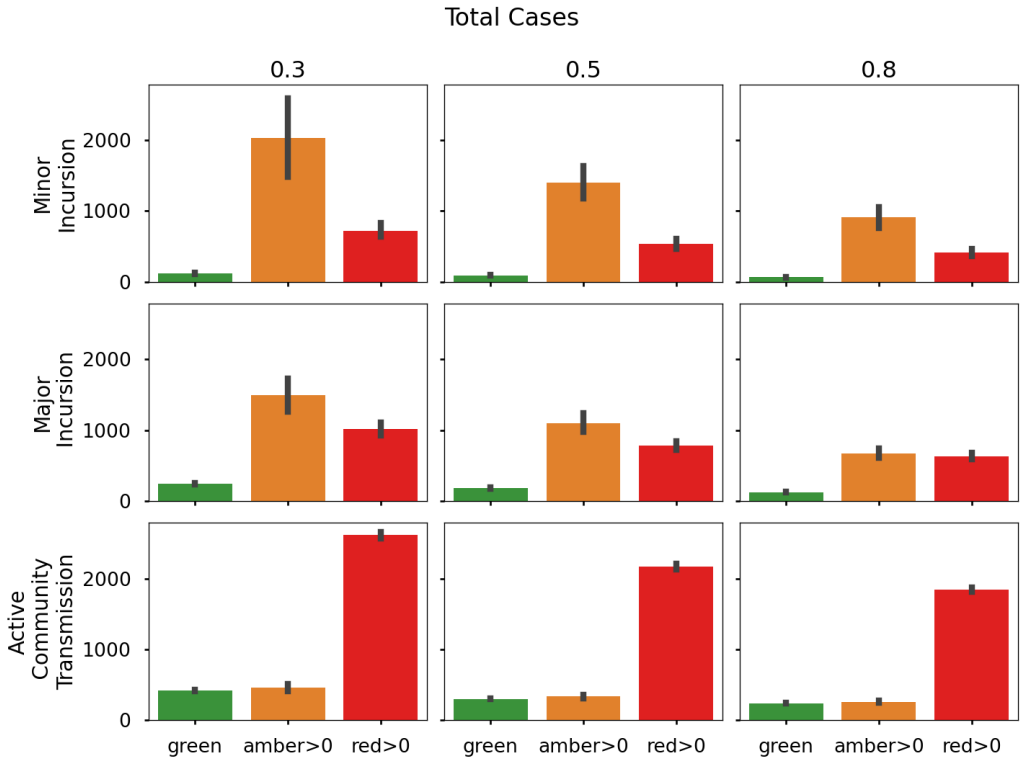

Figure 6.11: The mean total number of cases in simulations that never exceeded the COP metric for tracing delays (green), exceeded at least once into amber alert but never red (amber >0) and exceeded at least once into red alert, given routine detection probabilities of 0.3, 0.5 and 0.8. Simulation initialised with tracing to complete at capacity, and assume first order contact tracing up to the preceding two days of symptom onset of the index case.

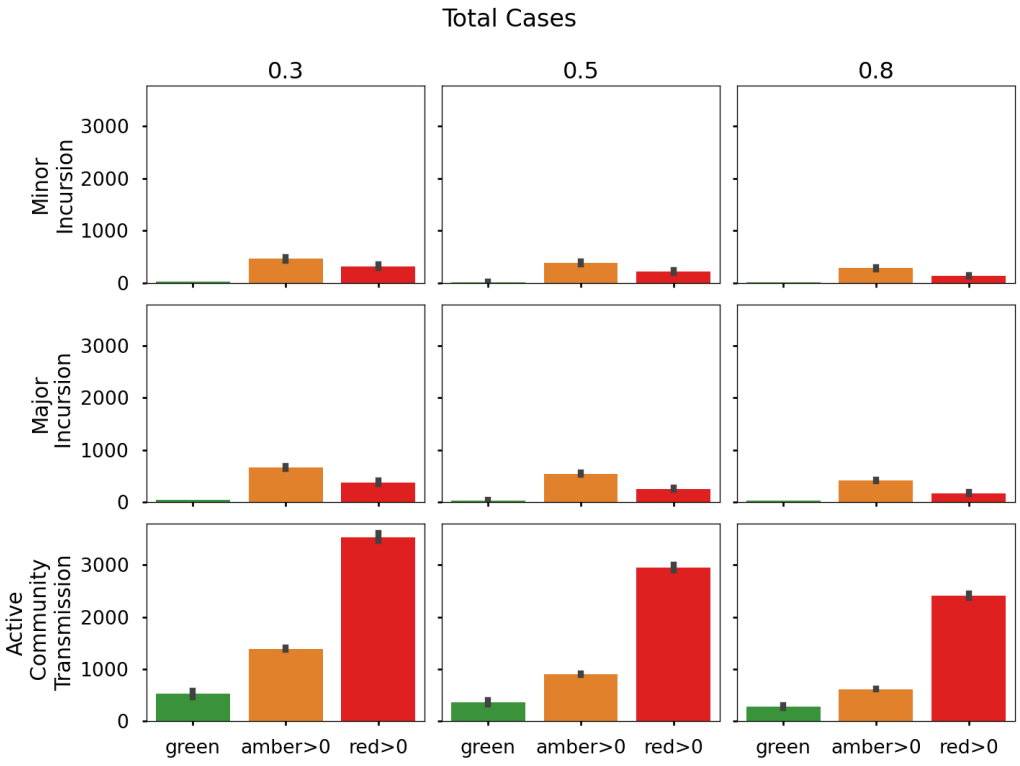

Figure 6.12: The average number of total cases after 30 days after reaching amber for testing delays at least once, compared to when never reaching amber, given routine detection probabilities of 0.3, 0.5 and 0.8. Simulations initialised with tracing to complete at capacity, and assume first order contact tracing up to the preceding two days of symptom onset of the index case.

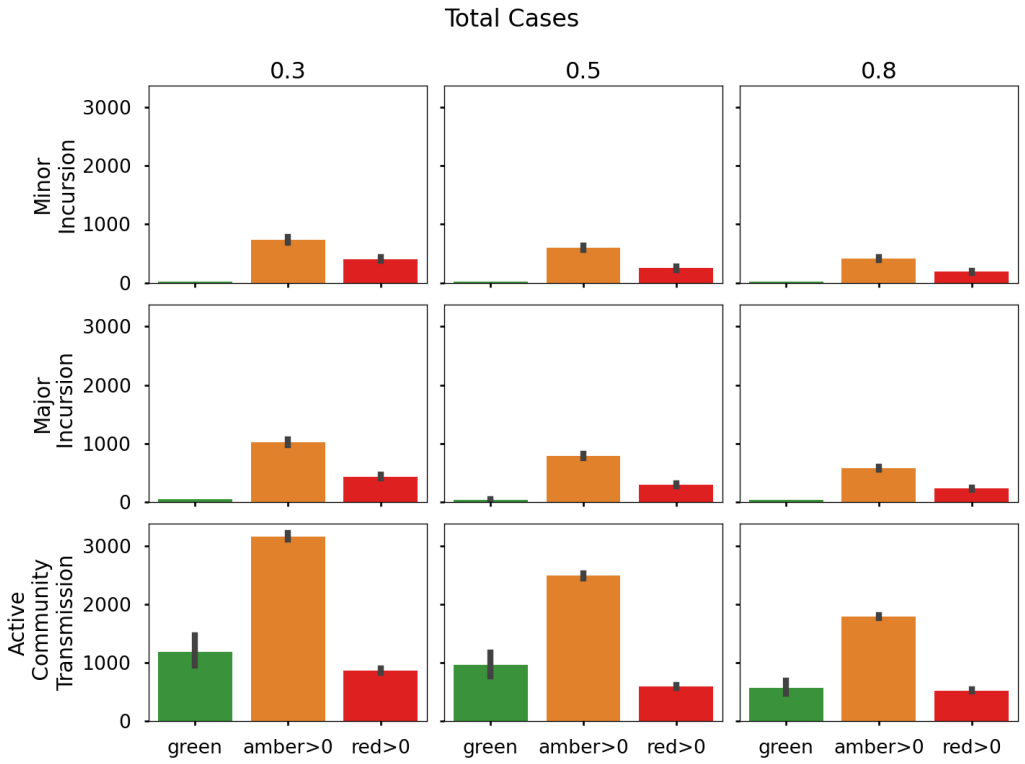

Figure 6.13: The average number of total cases after 30 days after reaching amber for testing delays at least once, compared to when never reaching amber, given routine detection probabilities of 0.3, 0.5 and 0.8. Simulations initialised with the index cases presenting for swabbing an additional seven days plus the sampled presentation time.

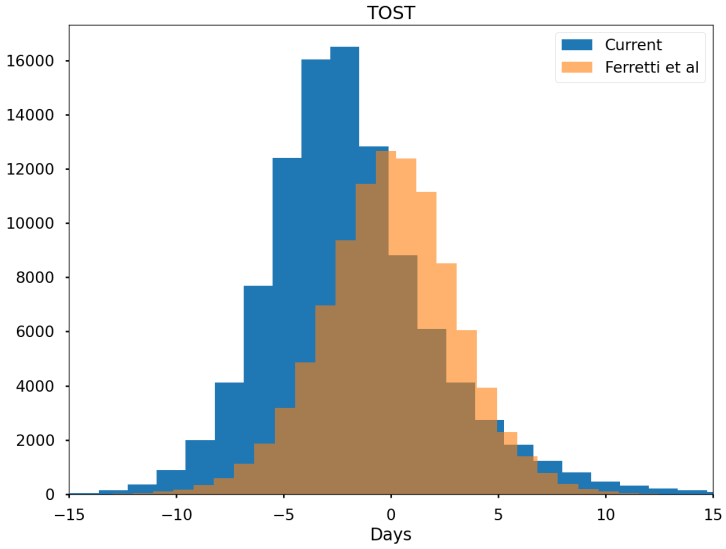

Figure 6.14: The sampled distribution of the time in days of transmission relative to symptom onset of the infector (TOST) used in this analysis (blue), and an alternative similar to that found in Ferretti *et al.* 2020 (orange). 40% of infections will occur 3 days before symptom onset, and 80% of infections by 1 day after symptom onset. In the Ferretti *et al.* distribution, 40% of infections occur 0.70 days before symptom onset, and 80% of infections occur by 2.5 days after symptom onset.

6.3.3 Sensitivity to generation interval

For a number of reasons, case isolation alone has limited effectiveness for controlling the spread of COVID-19. A substantial proportion of the infectious period occurs prior to symptom onset [38, 56, 58, 111], and this is also the case in the choice of generation interval and incubation period distributions used here where 40% of infections occur 3 days before symptoms onset. The sampled distribution of transmission relative to symptom onset of the infector used here can be seen in Figure 6.14. We examine the sensitivity of our results to this assumed distribution by comparing them with simulations incorporating a longer generation interval to align with the distributions proposed by Ferretti *et al.* (2020)[38]. The results of this comparison can be seen in Figure 6.15.

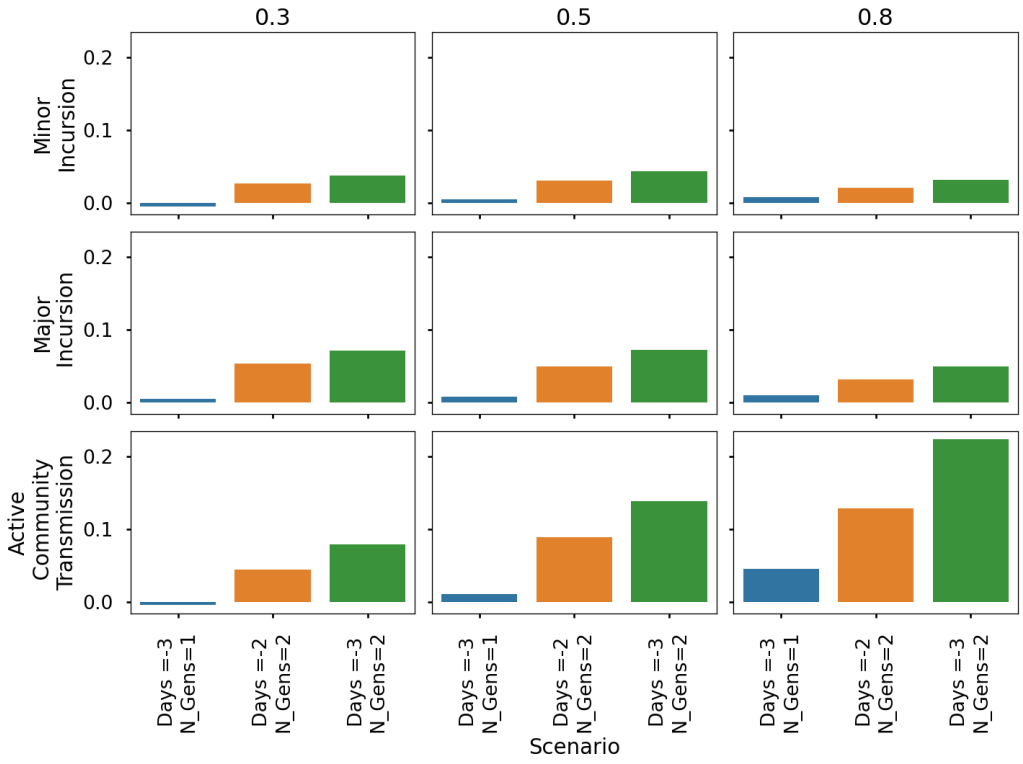

Figure 6.15: The marginal improvement in elimination probability using the approximate generation interval and TOST used in Ferretti *et al.* (2020), relative to first order contact tracing up to the preceding two days to symptom onset of the index case, with capacity limits and given routine detection probabilities of 0.3, 0.5 and 0.8. The assumptions examined are increased duration of backtracing up to 3 days prior to symptom onset (Assumption 3, blue), tracing an additional generation of close contacts (2, orange), and both of the former (2 and 3, green).

6.4 Discussion

The marginal improvement in elimination probability is low across all scenarios if the proportion of pre-symptomatic transmission is high. When a lower pre-symptomatic transmission is assumed, the marginal improvement in elimination probability is much higher, as seen when comparing Figures 6.2 and 6.15. Differing settings and contexts will change the generation interval distribution, and the appropriate assumption may lead to alternative conclusions to those presented here. The sensitivity to the generation interval presented here is a significant consideration when intending to generalise this work to other settings. However, the model described in this chapter is flexible to changes in the generation interval, and can be adapted to other contexts.

In both the minor and major incursion, the absolute elimination probability at baseline is already high, largely due to the small value of the dispersion $k = 0.1$, as described in Section 2.3.1 and Section 5.1.6. This leads to a high probability of very few offspring from the initial cases, which can lead to elimination, but small probability of a very large number of offspring that may be difficult to contain. In the latter outcome, the mean number of cases after 30 days may be a better metric to examine, and this is presented in some scenarios here. However, it is likely that presented with a super spreading event, public health departments may choose to initiate wider scale social distancing policies, such as lock downs and large scale limitations in mobility of the public. This can be modelled in this framework, where the overall reproduction number $R_L(t)$ may be reduced according to the model in Chapter 5 (see Equation (5.4)), triggered by observing more than a threshold of cases. The complexity of choosing a threshold, which will vary among the many scenarios and assumptions examined here, is left for future work.

In this chapter, we have adapted a branching process model to examine the effectiveness of different contact tracing strategies, and considered the relevance of performance metrics in examining contact tracing effectiveness, using targets taken from an Australian context. We found that with no capacity constraints, tracing second order contacts is more effective than pursuing an increased duration of backtracing from 2 days prior to symptom onset up to 3 days, but there are synergistic effects of implementing both. However, if capacity constraints are exceeded, tracing second order constraints may lead to detrimental outcomes as additional delays occur. Care must also be given when using metrics to assess the performance of contact tracing, to avoid misleading signals when there are few cases.

Chapter 7

Conclusion

This thesis utilises digital data to provide insight to social behaviours that may drive disease transmission and observations. We have examined how behaviours may influence the observations of an epidemic, and how the response of individuals and public health may change the course of the epidemic, using Bayesian inference methods to learn the influence of behaviour on epidemiological parameters, and mathematical models to simulate probabilistic outcomes.

In Chapter 3 we address how the behaviour of individuals in a participatory health surveillance system is influenced by acquisition of the disease we wish to observe, resulting in a biased estimate of disease prevalence. Using a Bayesian logistic regression model, we arrive at a substantially different estimate of disease prevalence after adjusting for individual behaviours. We see that individuals are much more likely to participate when they themselves have symptoms, but this effect decays over time. This is extended in Chapter 4 where we apply a similar approach to determine the probability of individuals being associated with an influenza test from a primary care physician. Compared to other machine learning classifiers, our model is among the best performing algorithms, while inherently providing estimates of uncertainty, which is critical for public health policy. The framework used in both Chapter 3 and Chapter 4 can be used wherever participation is of concern in digital health surveillance systems, and emphasises the influence of the behaviour of individuals in observing epidemics.

Chapter 5 then developed a framework to understand and forecast the spread of COVID-19 in Australian jurisdictions, applying a model incorporating social mobility metrics and individual behaviour to an established method

of inferring the effective reproduction number [27, 112] to better adapt to changing behaviours and response to an ongoing epidemic, without relying solely on historical case incidence. A hierarchical model is applied to allow for differences across jurisdictions, while also partially informing parameters from jurisdictions with greater observations to jurisdictions with very few. The forecasted effective reproduction number is then used to simulate the number of cases into the future. This framework is validated across a number of forecasts that cover large changes in epidemic activity and social mobility from individual behaviours and government policy. The branching process model of this framework was expanded to assess the effectiveness of contact tracing strategies in Chapter 6, where different public health responses in capacity constrained environments were compared. Using targets set by the Australian Government's Department of Health as a case study, we assess the efficacy of certain performance metrics in signalling control of an outbreak.

Throughout this thesis, we have emphasised the ability of digital data sets in providing insight into the influence of behaviour in epidemics that is typically unavailable in traditional settings. When carefully utilised, epidemic models incorporating such insights and behaviour provide a greater understanding of disease transmission that are robust to changing circumstances. The frameworks developed in this thesis can be applied in a variety of contexts and settings wherever digital data sets can provide insight into behaviour in epidemics.

Appendix A

Supplementary material for Chapter 3: User behaviour in digital systems

Methods

Table A.1 presents all predictors used in the analysis. Figure A.1 depicts the choice of prior distribution for the regression coefficients.

Results

All figures are presented in reverse chronological order, as the more recent years have the most directly comparable populations of participants to 2017, the year that was primarily discussed in the main paper. Figures A.2 to A.7 compare the naïve estimate to model posterior distributions of ILI prevalence in vaccinated and unvaccinated individuals across all years studied. 2017 results are presented in the main article. Figures A.8 to A.12 show the marginal posterior densities of the log odds ratio of all regression coefficients for predicting the probability of an individual reporting on-time across all years examined. Figure A.11 shows some of the bivariate kernel densities of certain parameters of the posterior, displaying the lack of correlation between parameters in the posterior. Figures A.13 to A.19 compare the model predictions against summary statistics of the test sets across all years examined.

Table A.1: List of predictor variables used and their definitions. Asterisks indicate variables in which the value is divided by the number of weeks in the season to scale the predictor between 0 and 1. School holidays breaks are sourced from the Australian Government [3].

Predictor	Description
HH has reported having ILI previously	Has the household (HH) reported having ILI previously this year, exclusive of the master.
HH has reported having symptoms	Has the HH reported having either fever or cough, but not both, previously this year, exclusive of the master.
Master works with patients	Is the master in a patient facing occupation.
Report is during a State/Territory school holiday	Does the current survey week cover days that are in the master's state school holiday period.
Master is reporting on behalf of others	Is the master reporting on behalf of others.
Master reported having ILI previously	Has the master reported having ILI previously this year.
Master reported having a symptom previously	Has the master reported having either fever or cough, but not both, previously this year.
Master is vaccinated	Has the master reported being vaccinated for influenza this year.
Proportion of reports so far	Proportion of reports submitted on-time so far this year.
Number of weeks since HH reported having ILI*	The number of weeks since the HH, but not the master, was reported to have ILI.
Number of weeks since HH reported having symptoms*	The number of weeks since the HH, but not the master, was reported to have a fever or a cough, but not both.
Number of weeks since reporting ILI*	The number of weeks since the master reported having ILI.
Number of weeks since reporting symptom*	The number of weeks since the master reported having a fever or a cough, but not both.
Week	Categorical predictor for the week of the survey. Not used in predicting whether a report will be on-time or late given a report has been submitted.
Intercept	Intercept variable to capture base level reporting rate. Not used in predicting whether a report will be on-time or late given a report has been submitted.

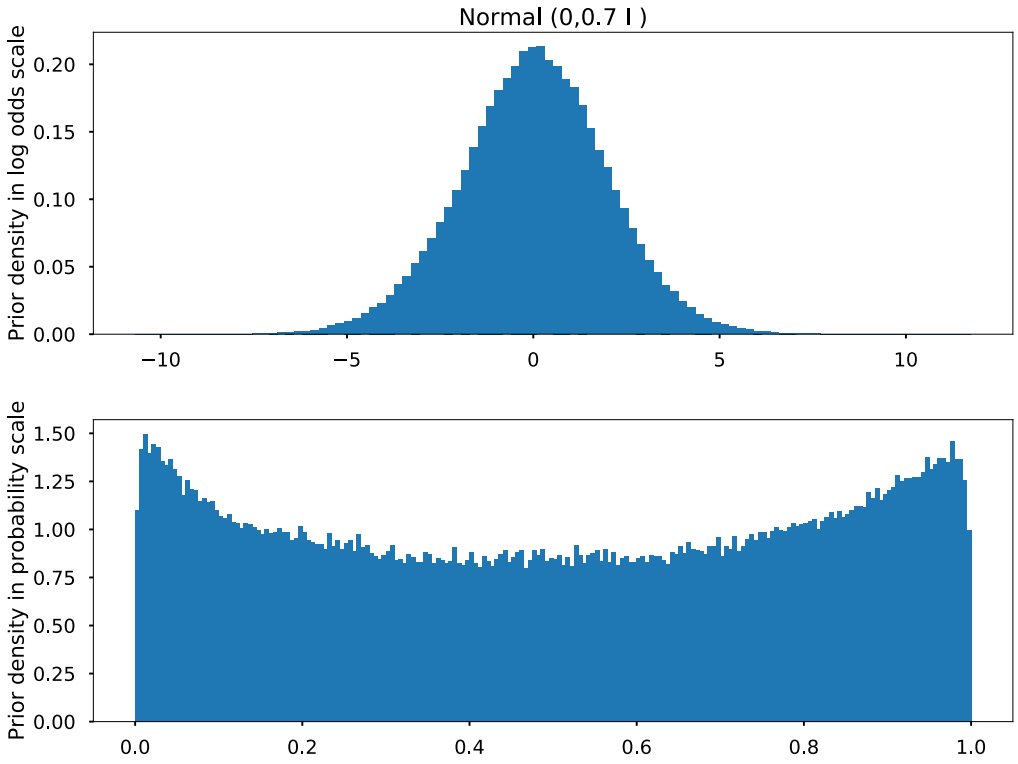

Figure A.1: The prior distribution of the regression coefficients transformed from the log odds scale to the probability scale via the logit function, after multiplication with 1000 samples from the training set predictors of 2017. The covariance matrix $0.7\ \mathcal{I}$ was chosen for the model as it was somewhat uniform across the probability scale, with some skewness towards the boundaries.

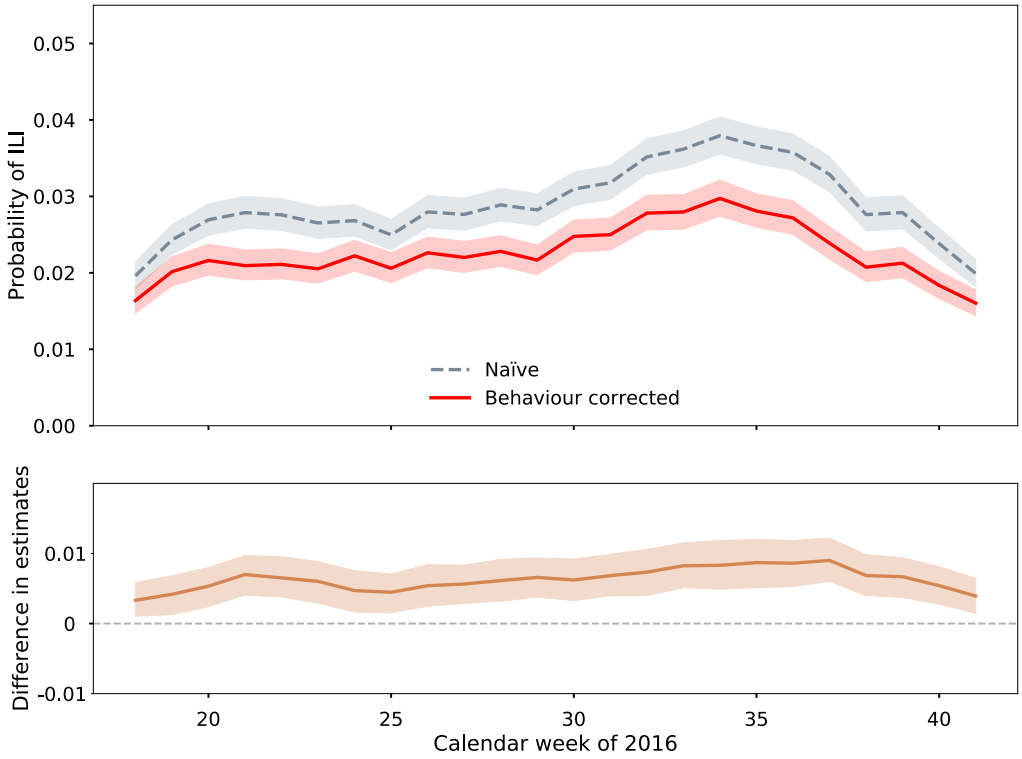

Figure A.2: Comparison of model posterior estimate of the prevalence rate of ILI in the population to the naïve estimate for the winter of 2016, and the difference between the two distributions over time. Lines represent the median of the distribution and shaded regions are 95% credible intervals of the posterior distributions.

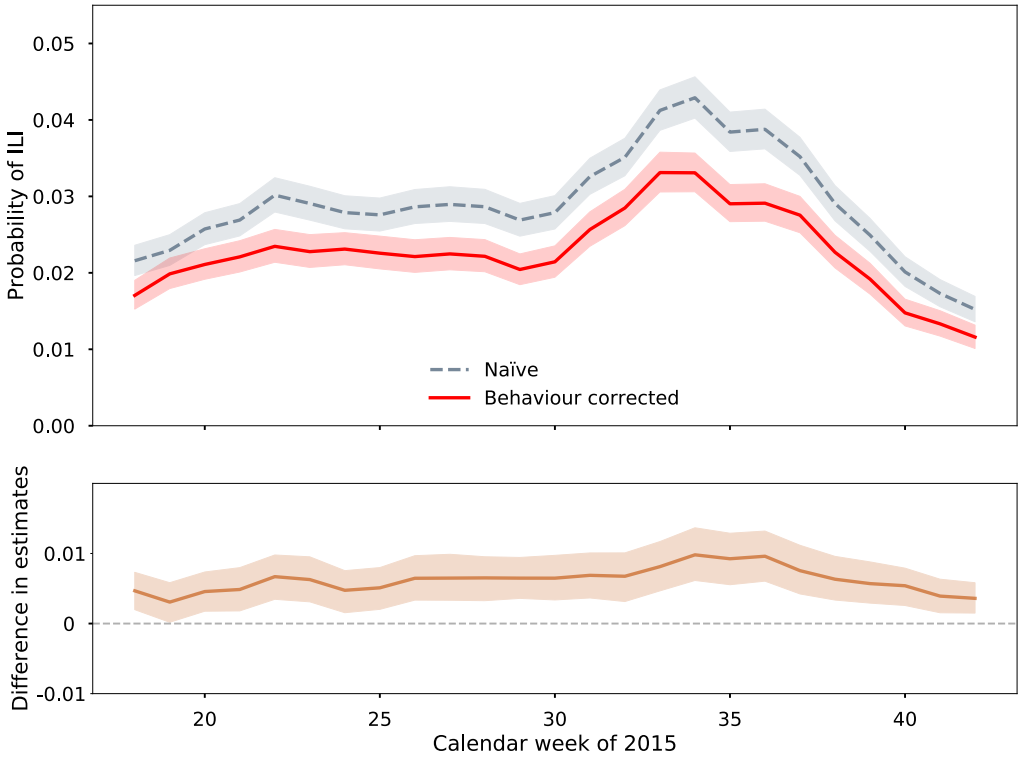

Figure A.3: Comparison of model posterior estimate of the prevalence rate of ILI in the population to the naïve estimate for the winter of 2015, and the difference between the two distributions over time. Lines represent the median of the distribution and shaded regions are 95% credible intervals of the posterior distributions.

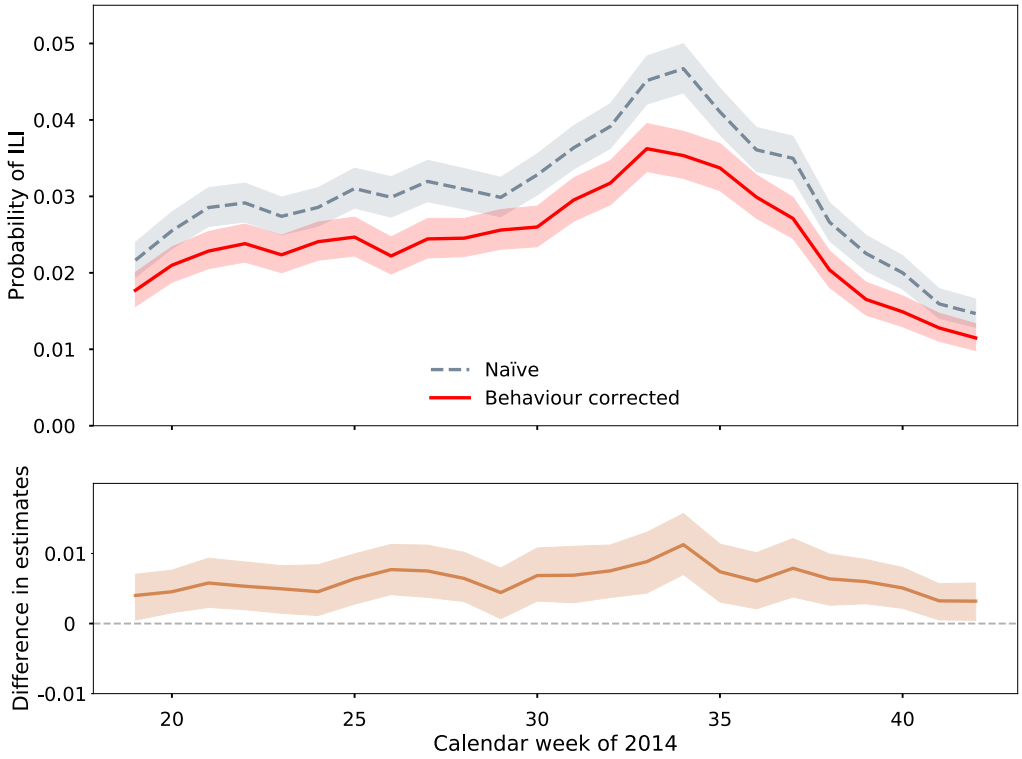

Figure A.4: Comparison of model posterior estimate of the prevalence rate of ILI in the population to the naïve estimate for the winter of 2014, and the difference between the two distributions over time. Lines represent the median of the distribution and shaded regions are 95% credible intervals of the posterior distributions.

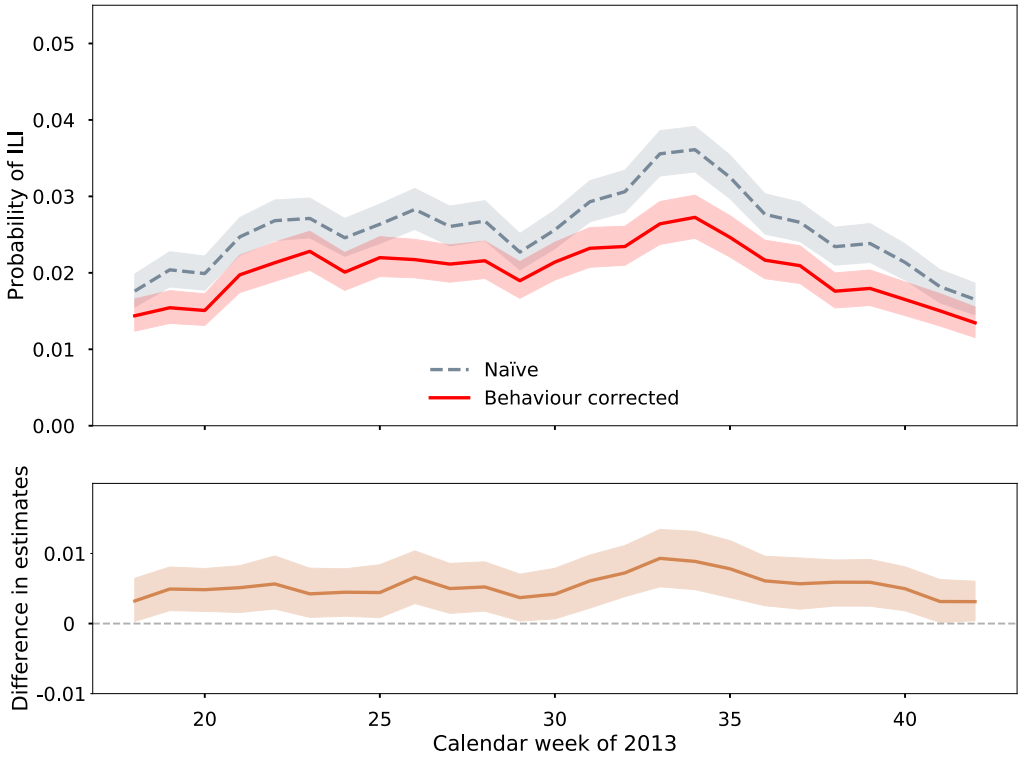

Figure A.5: Comparison of model posterior estimate of the prevalence rate of ILI in the population to the naïve estimate for the winter of 2013, and the difference between the two distributions over time. Lines represent the median of the distribution and shaded regions are 95% credible intervals of the posterior distributions.

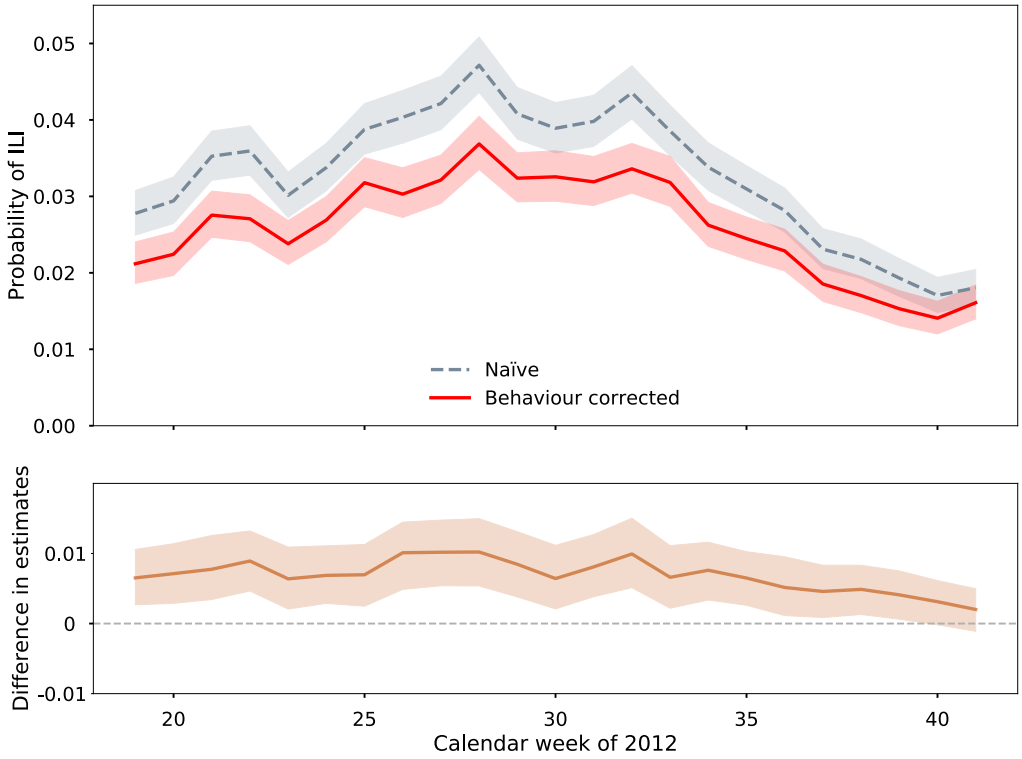

Figure A.6: Comparison of model posterior estimate of the prevalence rate of ILI in the population to the naïve estimate for the winter of 2012, and the difference between the two distributions over time. Lines represent the median of the distribution and shaded regions are 95% credible intervals of the posterior distributions.

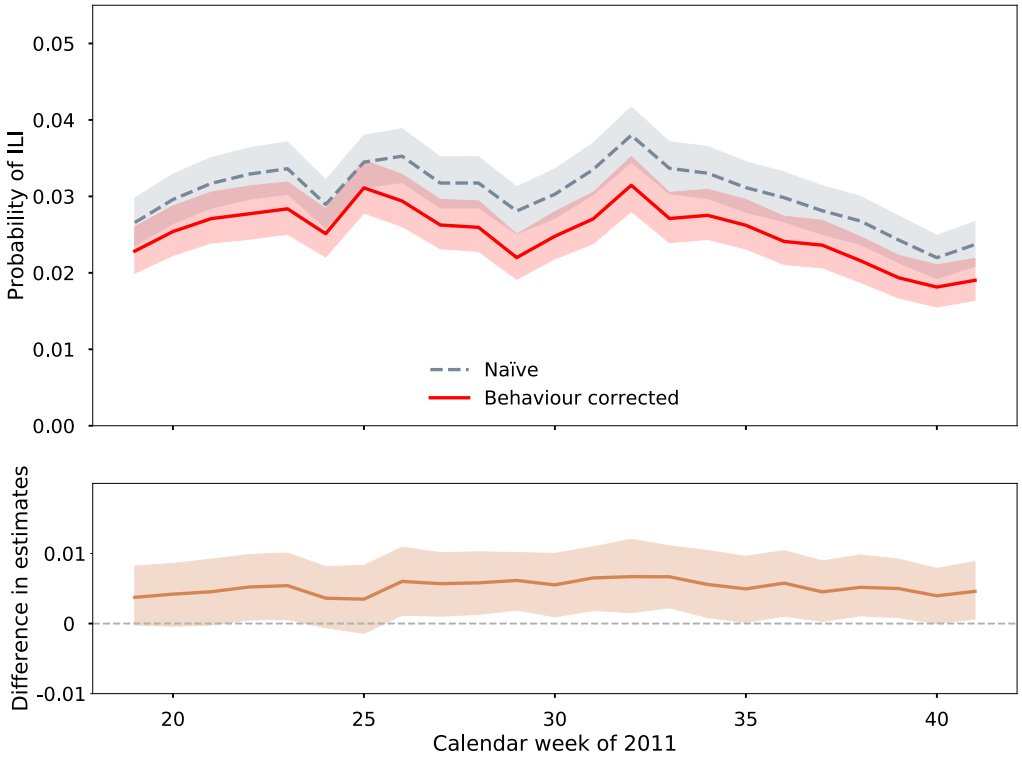

Figure A.7: Comparison of model posterior estimate of the prevalence rate of ILI in the population to the naïve estimate for the winter of 2011, and the difference between the two distributions over time. Lines represent the median of the distribution and shaded regions are 95% credible intervals of the posterior distributions.

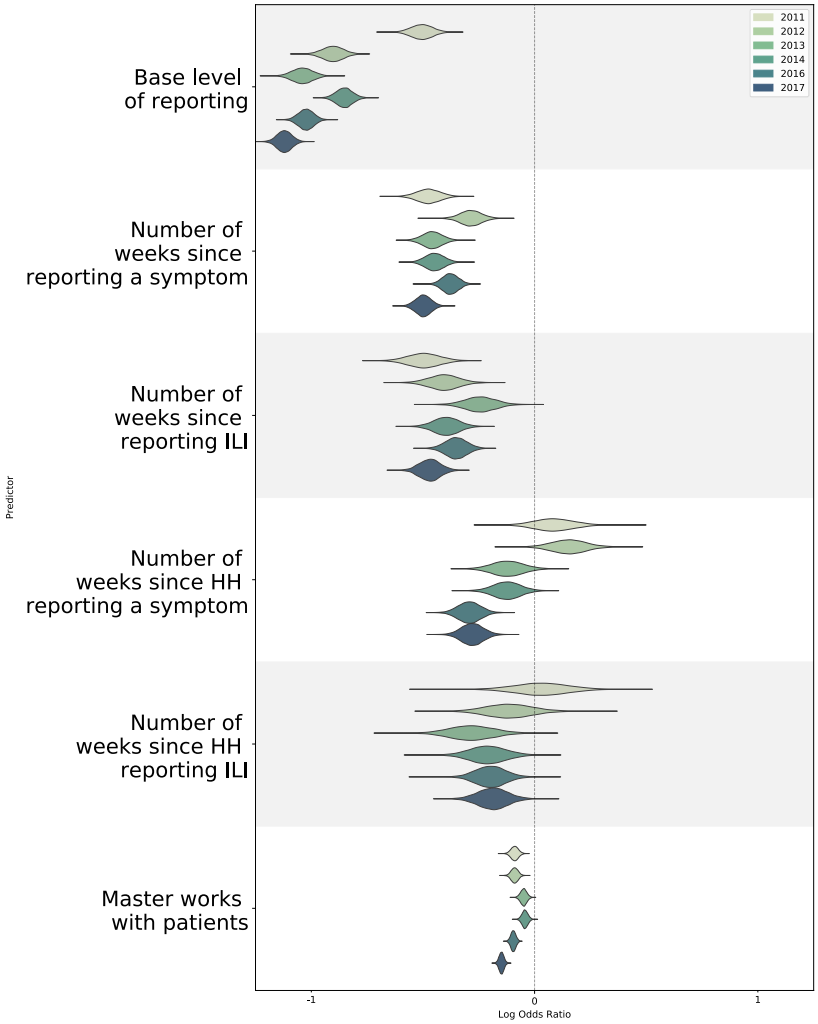

Figure A.8: The marginal posterior densities of the log odds ratio of the first half of non-chronological regression coefficients for predicting the probability of an individual reporting for all years.

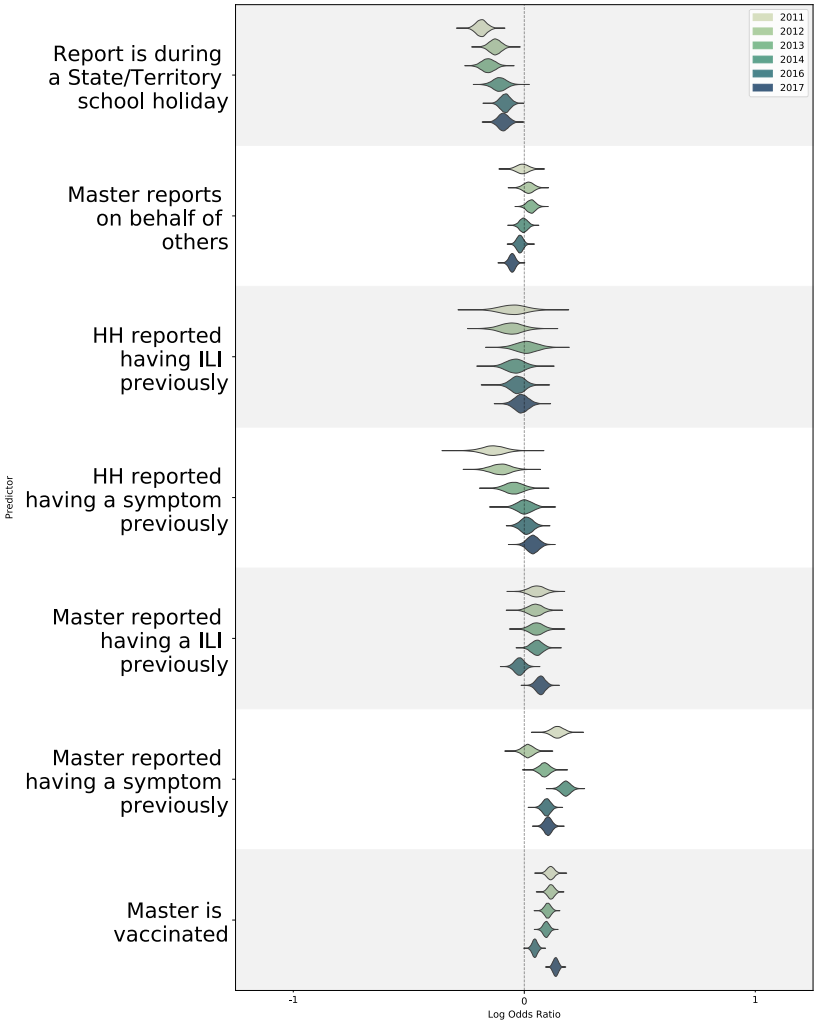

Figure A.9: The marginal posterior densities of the log odds ratio of the second half of non-chronological regression coefficients for predicting the probability of an individual reporting for all years.

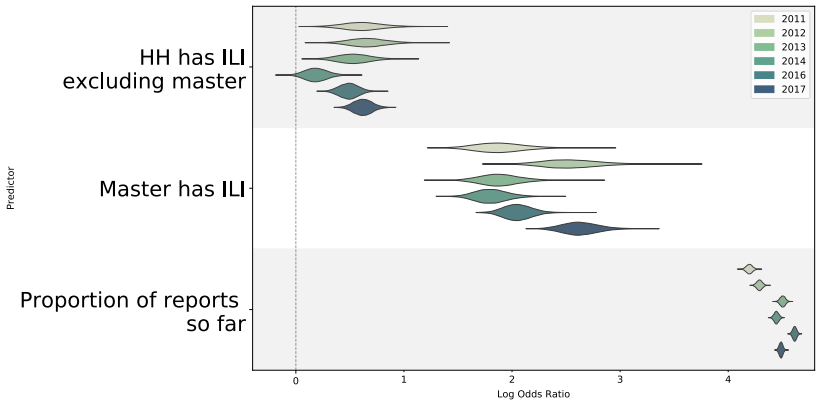

Figure A.10: The marginal posterior densities of the log odds ratio of the largest regression coefficients for predicting the probability of an individual reporting for all years. Presented here are coefficients for if a member of the household has ILI and the proportion of reports submitted on-time

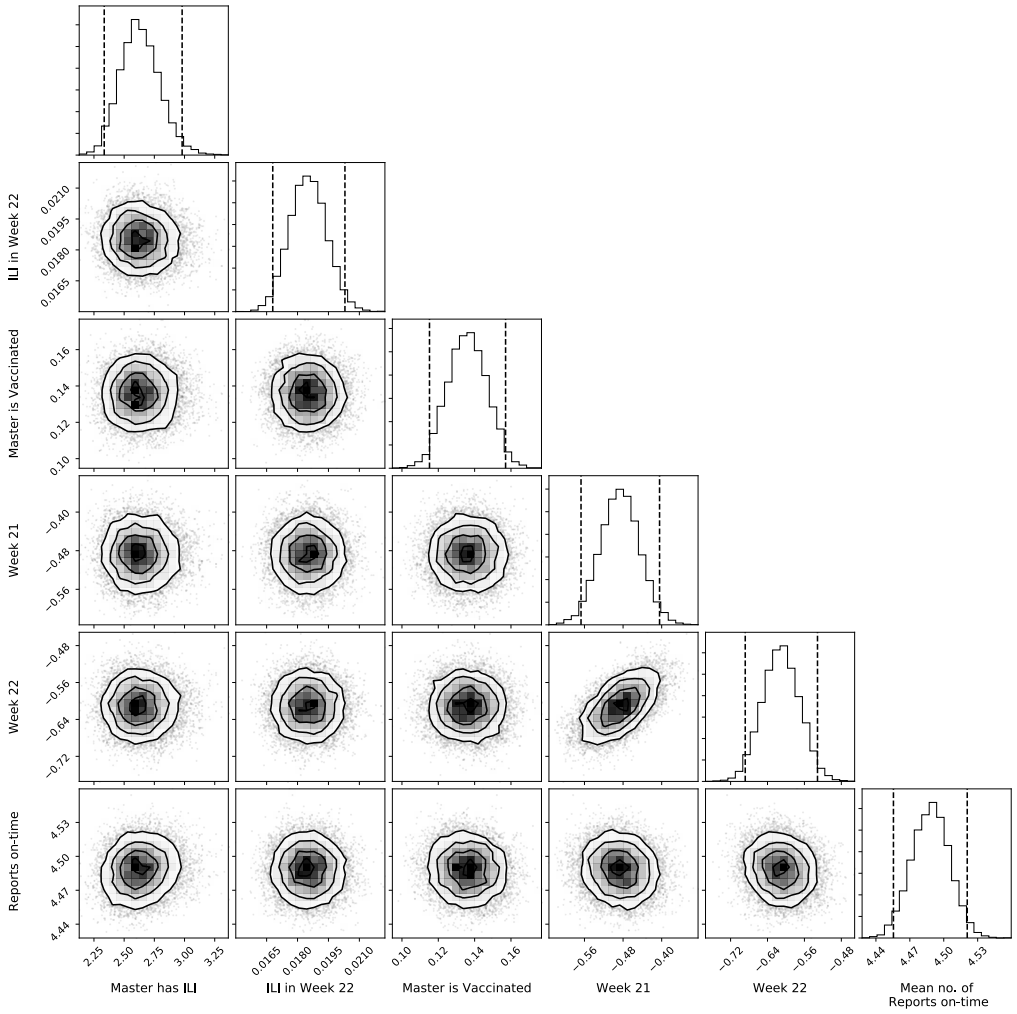

Figure A.11: The bivariate kernel density of samples from the posterior for some parameters inferred from the 2017 data. Bivariate plots show little correlation between parameters, with some correlation in the chronological regression coefficients (Week 21 and Week 22).

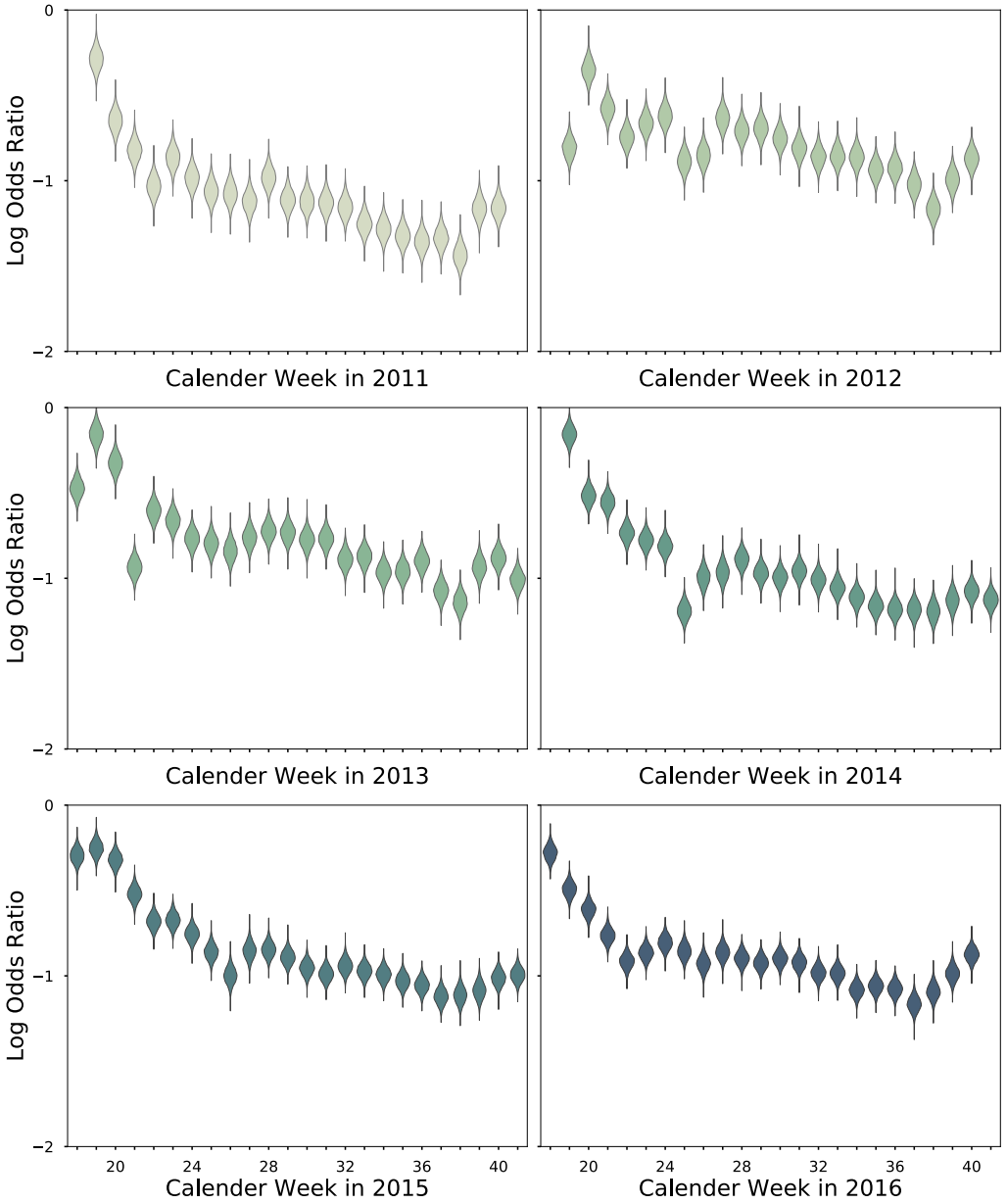

Figure A.12: The marginal posterior densities of the log odds ratio of chronological regression coefficients for predicting the probability of an individual reporting in a given week for years 2011 to 2016.

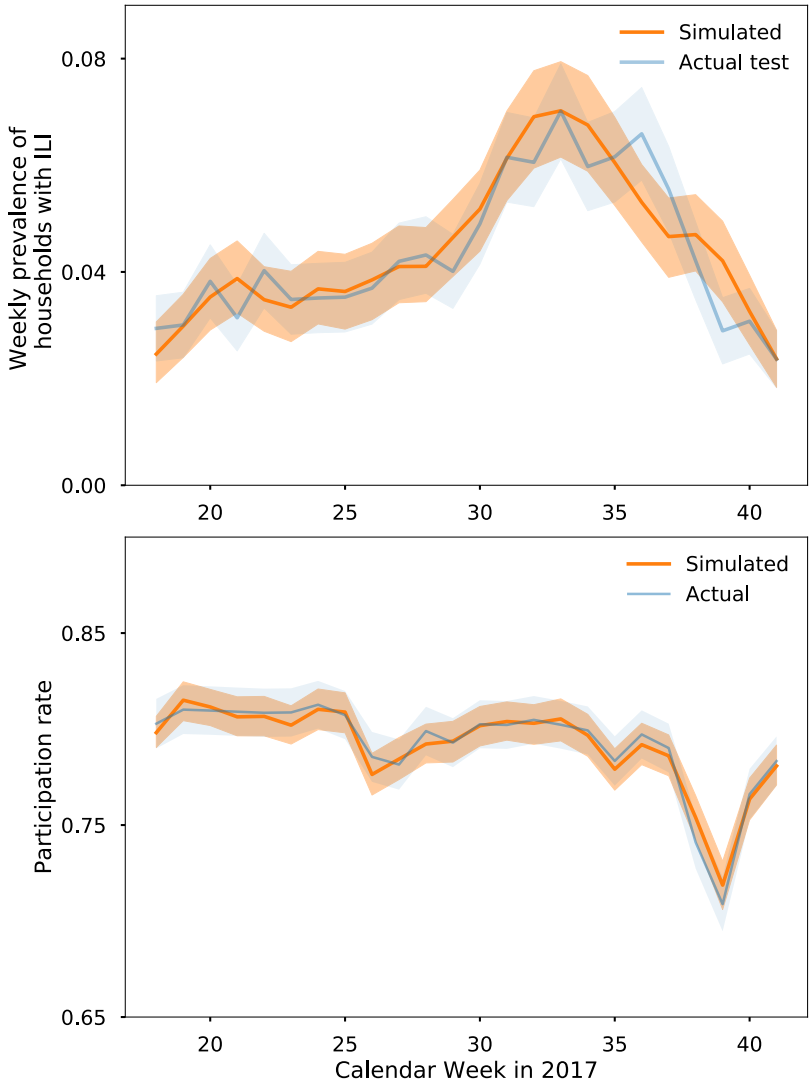

Figure A.13: Cross validation of model predictions with actual outcomes of test set for the 2017 season. Lines represent the median and shaded regions the 95% credible intervals. The model is able to generate the data observed in the test set with high probability.

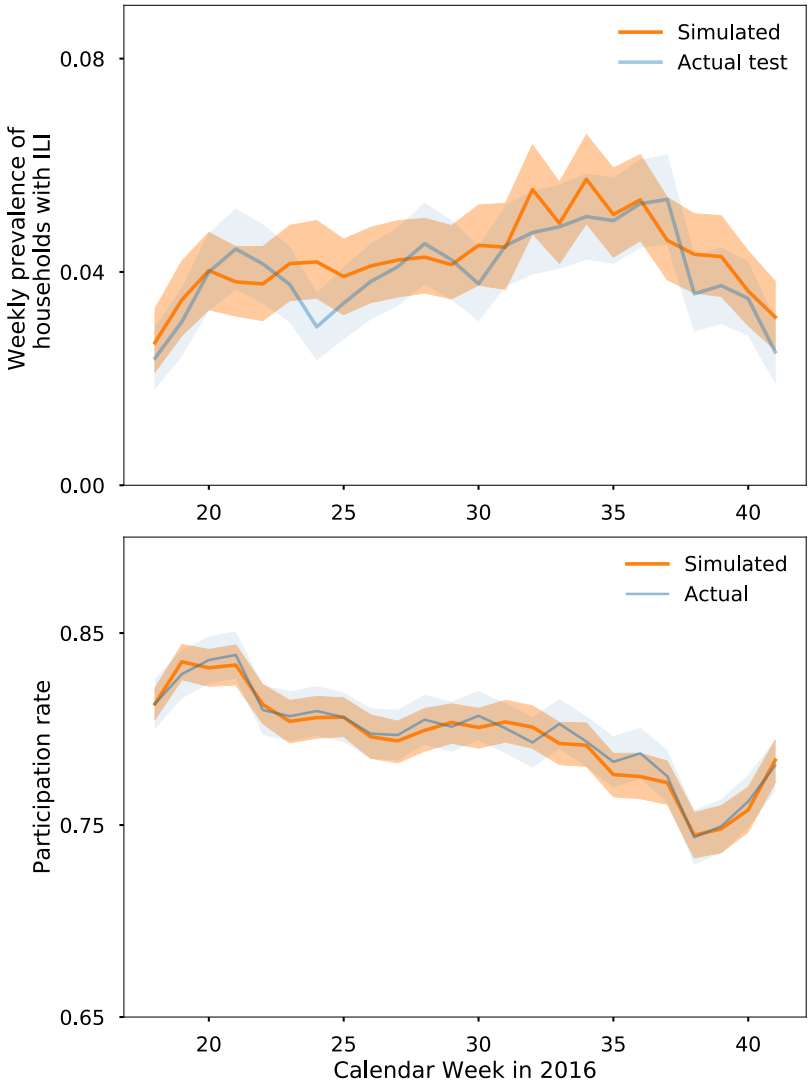

Figure A.14: Cross validation of model predictions with actual outcomes of test set for the 2016 season. Lines represent the median and shaded regions the 95% credible intervals. The model is able to generate the data observed in the test set with high probability.

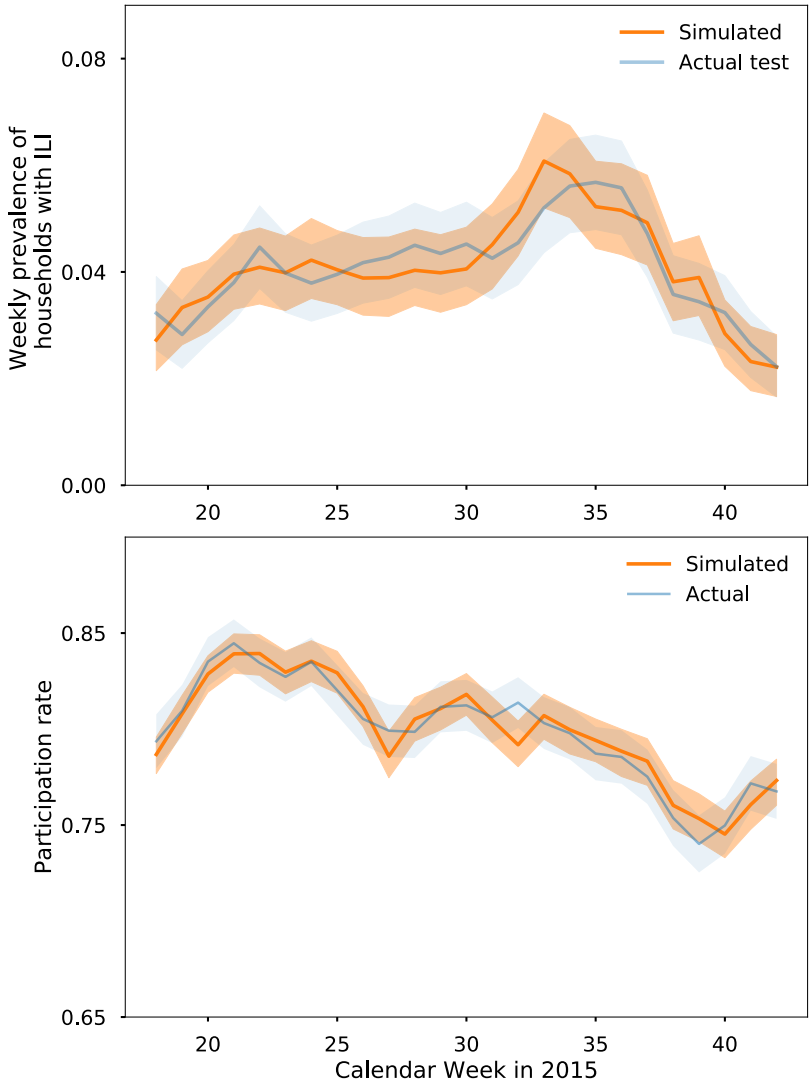

Figure A.15: Cross validation of model predictions with actual outcomes of test set for the 2015 season. Lines represent the median and shaded regions the 95% credible intervals. The model is able to generate the data observed in the test set with high probability.

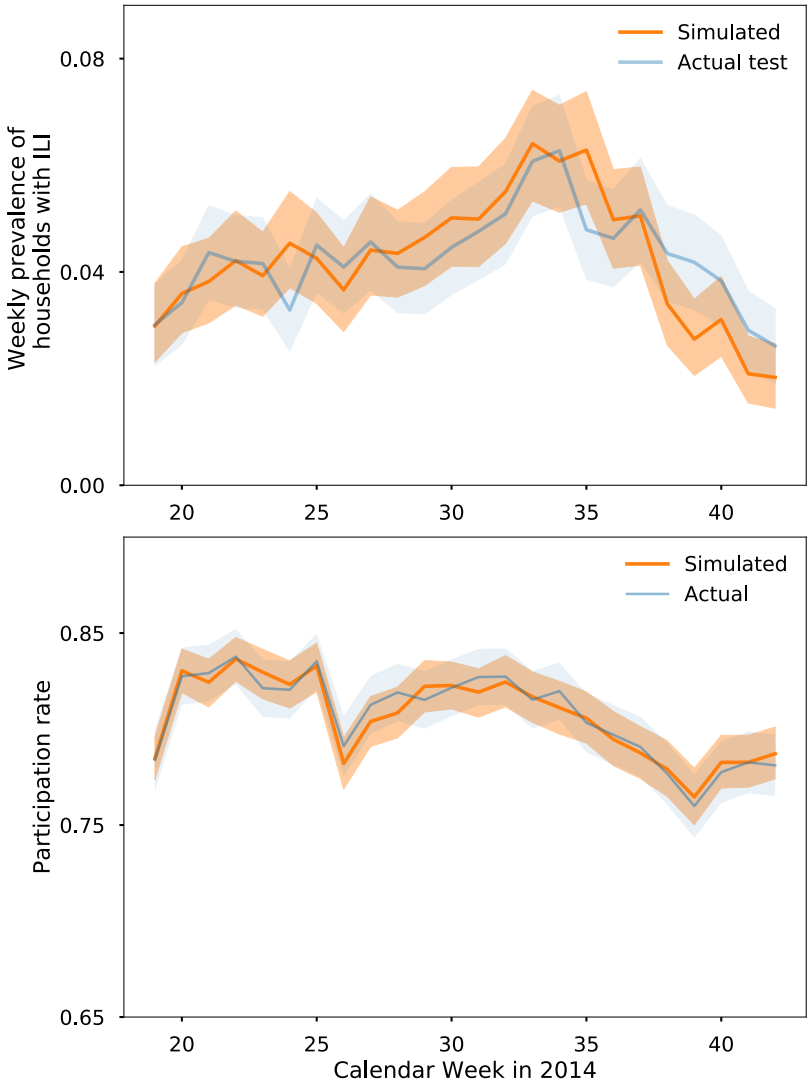

Figure A.16: Cross validation of model predictions with actual outcomes of test set for the 2014 season. Lines represent the median and shaded regions the 95% credible intervals. The model is able to generate the data observed in the test set with high probability.

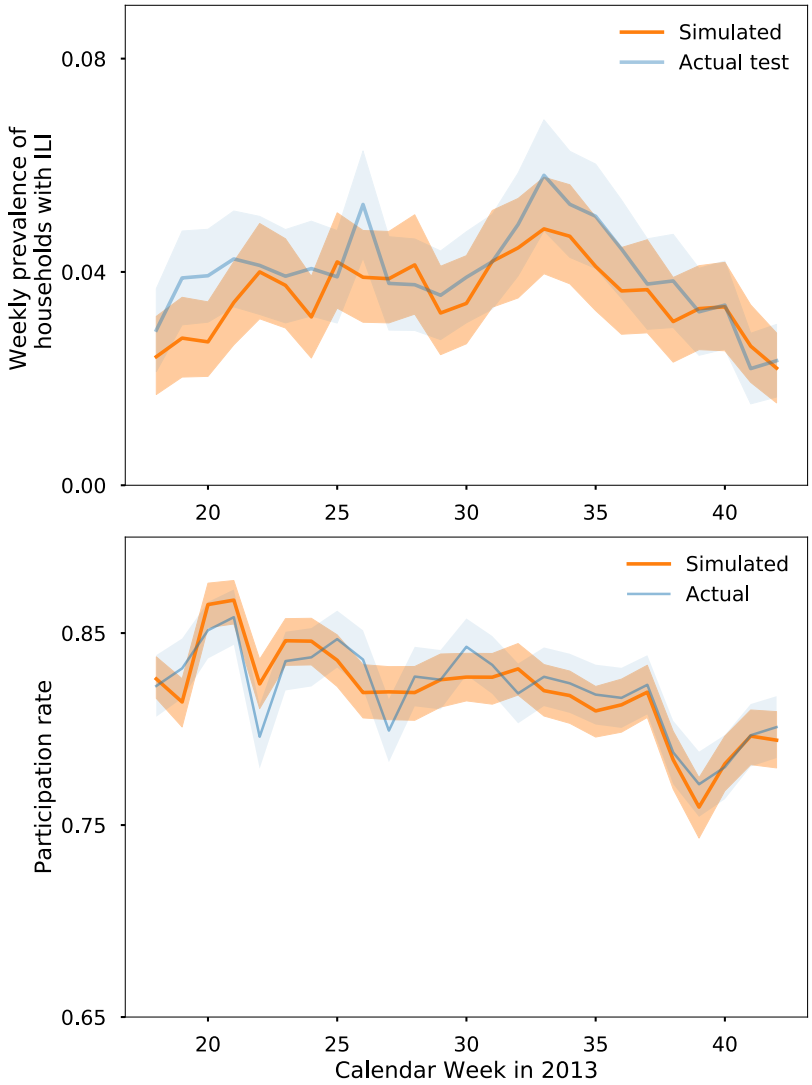

Figure A.17: Cross validation of model predictions with actual outcomes of test set for the 2013 season. Lines represent the median and shaded regions the 95% credible intervals. The model is able to generate the data observed in the test set with high probability.

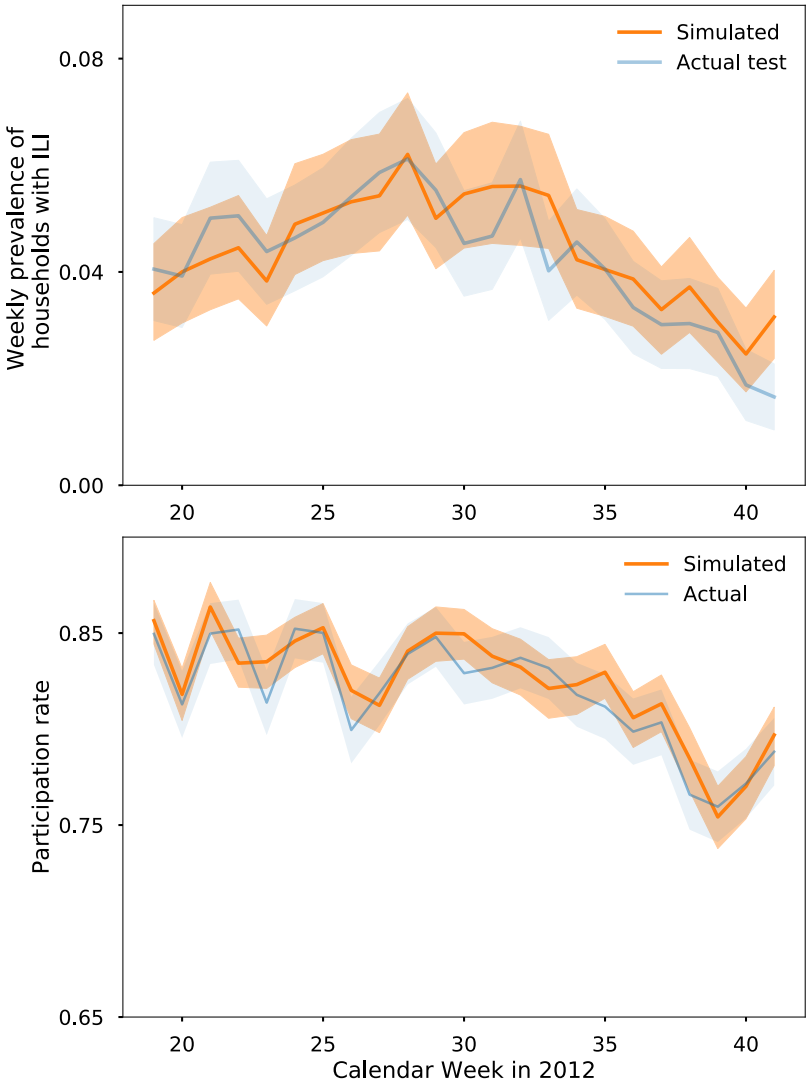

Figure A.18: Cross validation of model predictions with actual outcomes of test set for the 2012 season. Lines represent the median and shaded regions the 95% credible intervals. The model is able to generate the data observed in the test set with high probability.

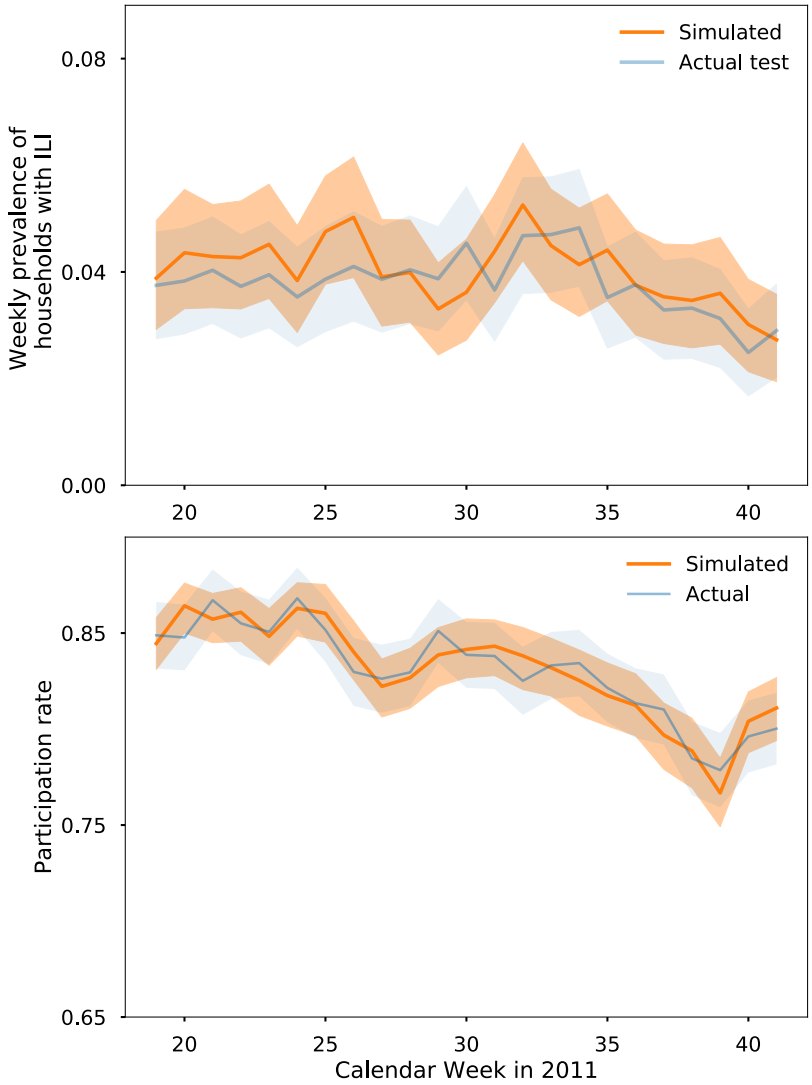

Figure A.19: Cross validation of model predictions with actual outcomes of test set for the 2011 season. Lines represent the median and shaded regions the 95% credible intervals. The model is able to generate the data observed in the test set with high probability.

Appendix B

Supplementary material for Chapter 4: Behaviour in testing

B.1 Predictors

The complete set of predictors used in the model and their definitions can be seen in Table B.1. These were derived from the variables in Table 3.1.

B.2 Supplementary results

Figure B.1 displays the log odds ratio of the all the non-chronological regression coefficients for predicting the probability of receiving an influenza test when presenting to the family doctor, ordered by the median of the their marginal posterior distribution.

Table B.1: Predictors used as input to classification algorithms.

Predictor	Explanation
Days of absence	The number of days the patient is absent from normal duties and work in the reporting week.
Healthcare worker	Is the patient in a patient facing occupation.
HH has children	Are any of the patient's reported household (HH) members under 15 years of age.
Patient reported having ILI	Has the patient reported having ILI previously in the year.
HH reported having ILI previously	Have members of the patient's HH reported having ILI previously in the year.
Patient is elderly	Is the patient over the age of 65 years.
Patient is child	Is the patient under 5 years of age.
Patient is youth	Is the patient between 5 and 15 years of age.
Second visit	Has the patient previously reported visiting the family doctor earlier in the year.
Patient is vaccinated	Has the patient reported being vaccinated for influenza this year.
HH has elderly	Are any of the patient's reported household members over the age of 65 years old.
HH currently ILI	Does another member of the patient's household currently have ILI.
Year	Categorical predictor for the year the survey has been submitted for.
Intercept	Base level testing influenza rate.

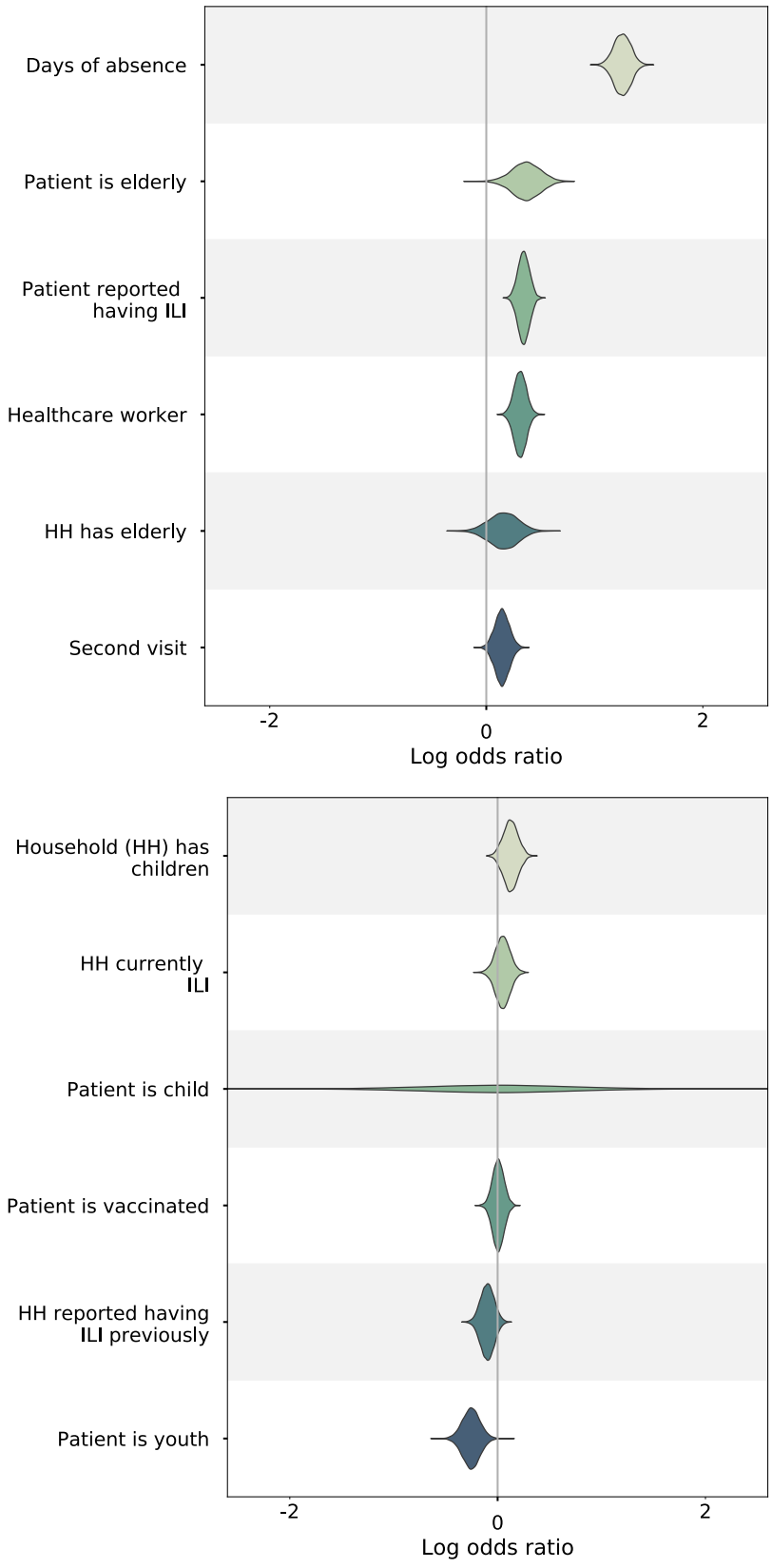

Figure B.1: The marginal posterior densities of the log odds ratio of the regression coefficients for predicting the probability of receiving an influenza test when presenting to the family doctor.

Appendix C

Supplementary material for Chapter 5: Forecasting COVID-19 using social mobility

C.1 Methods

C.1.1 Importation model

We use six time periods ($i = 1, \ldots, 6$) corresponding to:

- 01/03/20 to 06/03/20;
- 07/03/20 to 13/03/20;
- 14/03/20 to 18/03/20;
- 19/03/20 to 23/03/20;
- 24/03/20 to 14/04/20; and,
- 15/04/20 onwards.

Imported cases, and their dates of appearance in the NNDSS, were determined by:

- if known to be acquired in Australia, unknown, or missing, then it was assumed to be locally acquired, and is otherwise an import; and,
- True Onset Date is used as the primary date; but Notification Date - 5 is used as required, due to missing data. If Notification Date is missing, then Notification Receive Date -6 is used.

To assign these to the period in which they are likely to have arrived (as infectious), we have subtracted 4 days from their symptom onset date, for fitting purposes. We assume, within each state ($j = 1, \ldots, 8$) and in each period, i, a Poisson number of imports which are subsequently detected, $N_{i,j}$. That is, $N_{i,j} \sim \text{Poi}(\lambda_{i,j})$ independently each day.

Assuming *a priori* $\lambda_{i,j} \sim \text{Gamma}(\alpha, \beta)$, we have *a posteriori* that $\lambda_{i,j} \sim \text{Gamma}(a_{i,j}, b_i)$ where

$$a_{i,j} = \begin{cases} \alpha + n_{i,j} & \text{if } i \neq 4, \\ \alpha + 1.3n_{i,j}, & \text{if } i = 4 \end{cases}$$
$$b_i = \beta + m_i,$$

in which $n_{i,j}$ is the total number of detected imported cases in period i in State j, and m_i is the number of days in the period i ($m = (m_1, m_2, m_3, m_4, m_5, m_6) = (6, 8, 4, 5, 22, 152)$). This could be generalised to have different priors in each State and/or period if imported cases is of significant interest in future work.

Imported infectious individuals that are to be subsequently detected are simulated, each day in period i in State j, $D_{i,j}$, from its posterior predictive distribution,

$$D_{i,j} \sim \text{NegBin}(a_{i,j}, 1/(b_i + 1)).$$

The parameters of the prior distribution have little impact on the posterior distribution. They are set to $\alpha = 1$ and $\beta = 1/5$.

We then need to simulate for each of these individuals a time from the symptom onset distribution, which then determines the day upon which they are observed, to fit the dates of appearance of imported cases in the NNDSS data.

After generating $d_{i,j}$ for each day in period i in State j, we sum up the total number of such cases, $s_{i,j}$, and simulate the total number of undetected imported cases for the period i and State j, $U_{i,j}$, as

$$U_{i,j} \sim \text{NegBin}(s_{i,j}, q_I),$$

where q_I is the detection probability of imported cases. The undetected imported cases are allocated to the days in the period i with equal probability.

This process is simulated first, before seeding with any local cases and simulating forward.

C.1.2 Results

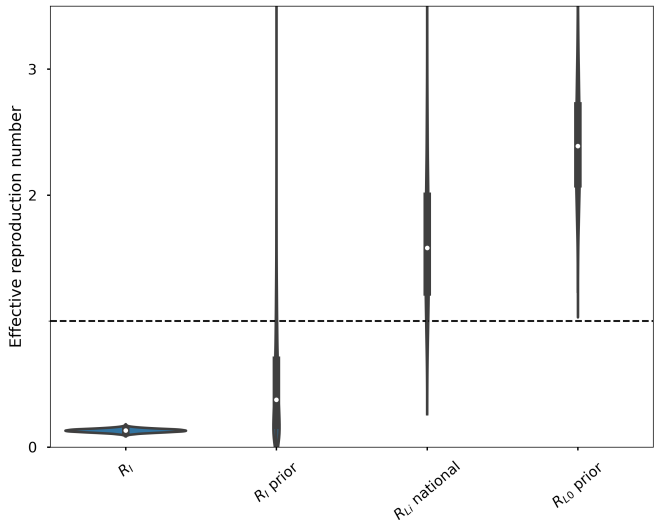

Figure C.1: Marginal posterior distributions of the effective reproduction number of imported cases R_I and the distribution of the effective reproduction numbers of all states R_{Li} national compared to their respective prior distributions.

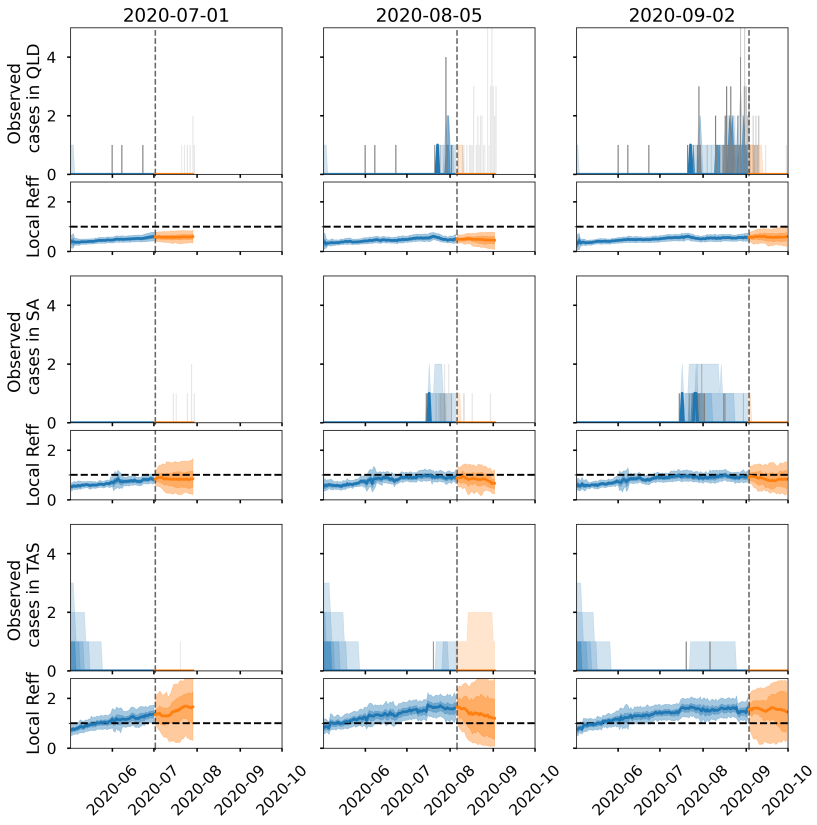

Figure C.2: The nowcast (blue) and forecast (orange) of observed local cases and the local effective reproduction number in Queensland, South Australia and Tasmania, using case data at 3 different dates. The shading represents the 90%, 80%, 70%, 60% and 50% quantiles. The dark grey bars represent the number of observed local cases known as of the data date, the medium grey bars represent the case data known 14 days after the data date, and the light grey bars represent the known cases as of November 2nd 2020.

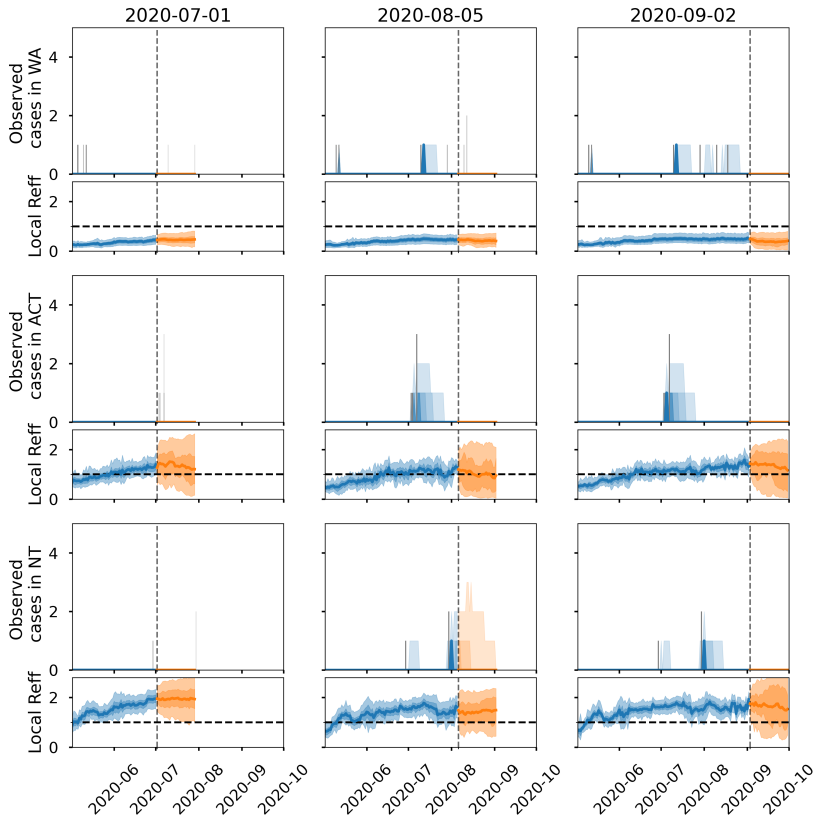

Figure C.3: The nowcast (blue) and forecast (orange) of observed local cases and the local effective reproduction number in Western Australia, Australian Capital Territory and the Northern Territory, using case data at 3 different dates. The shading represents the 90%, 80%, 70%, 60% and 50% quantiles. The dark grey bars represent the number of observed local cases known as of the data date, the medium grey bars represent the case data known 14 days after the data date, and the light grey bars represent the known cases as of November 2nd 2020.

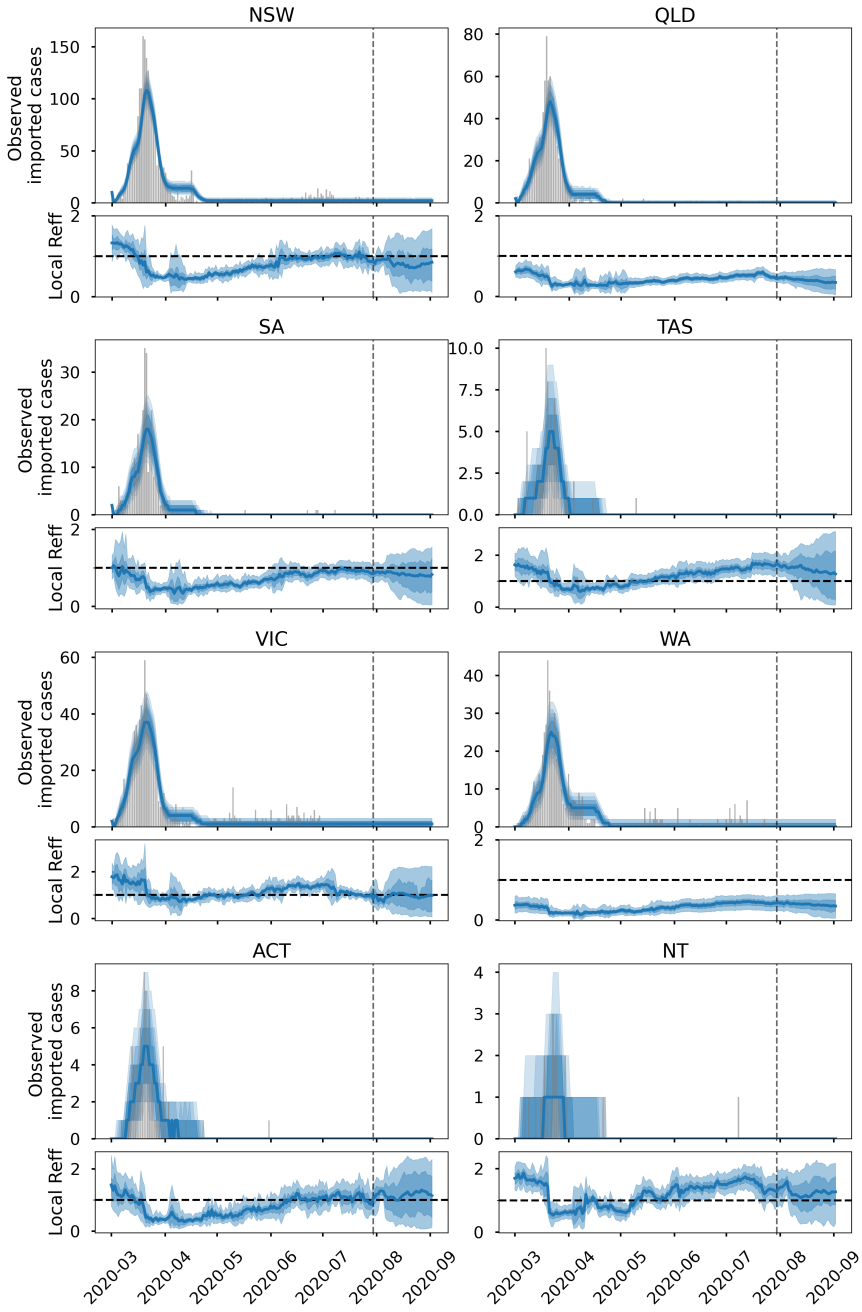

Figure C.4: Model forecasted versus observed of imported cases by State. Forecasted until 02/09/20. The vertical dashed line is the date from which forecasts begin.

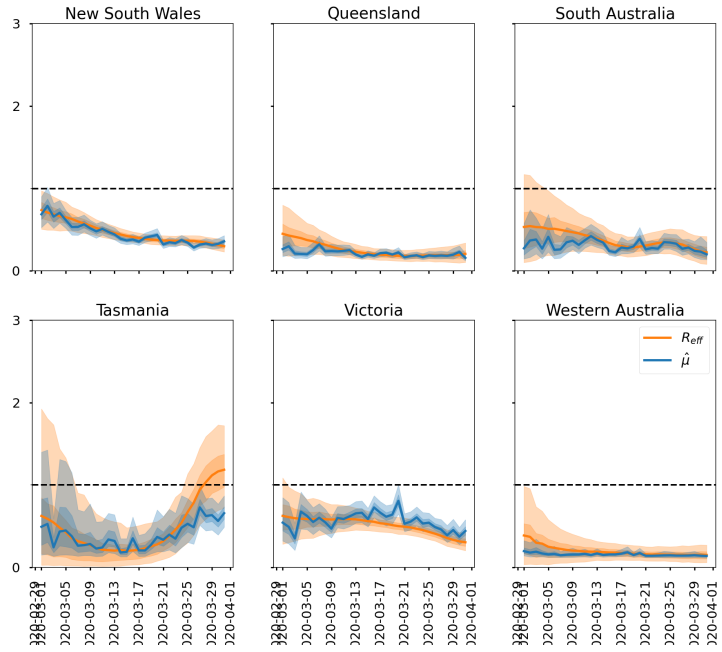

Figure C.5: The posterior predictive distribution of the effective reproduction number of all cases from the social mobility model $\hat{\mu}(t)$ (blue) compared to R_{eff} (orange) estimated using case data as of August 5th 2020. Solid line represents the median value, with the darker and lighter shade representing the 50% and 90% credible interval respectively.

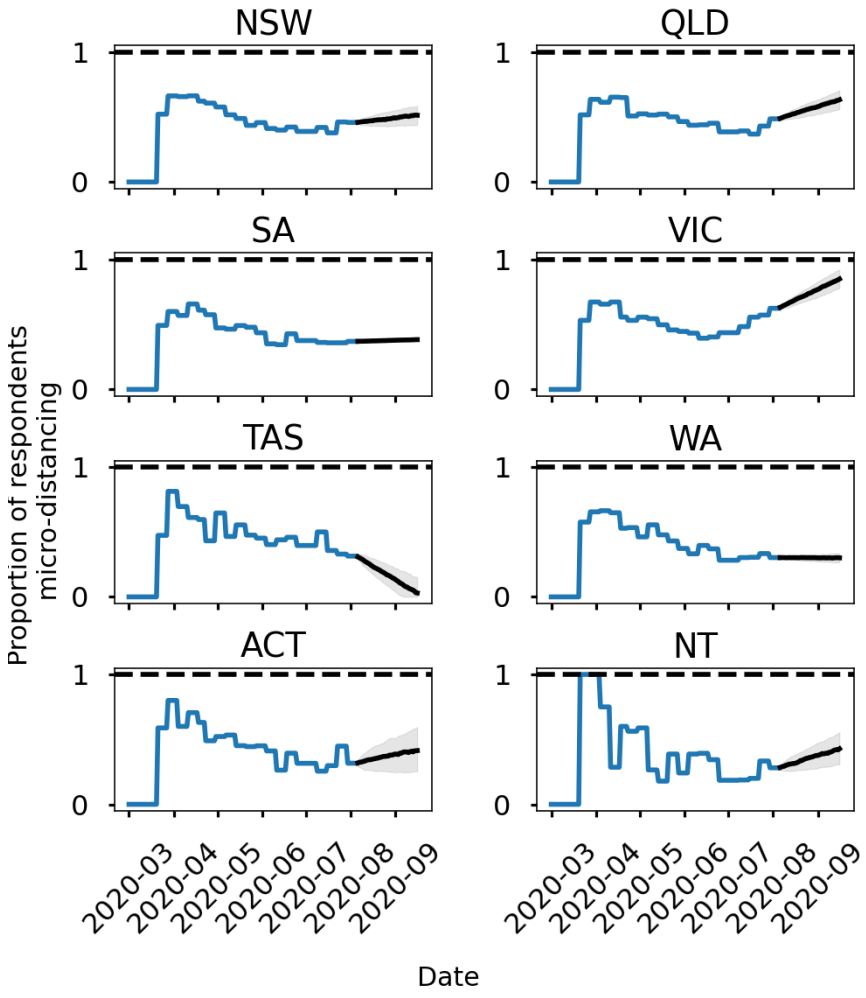

Figure C.6: The historical trend for the proportion of surveyed individuals observing micro-distancing behaviour (blue) and the forecasted values (black) as of 2020-08-05. Solid lines represent the mean value while intervals are the 25% and 75% quantiles of the sampled distributions.

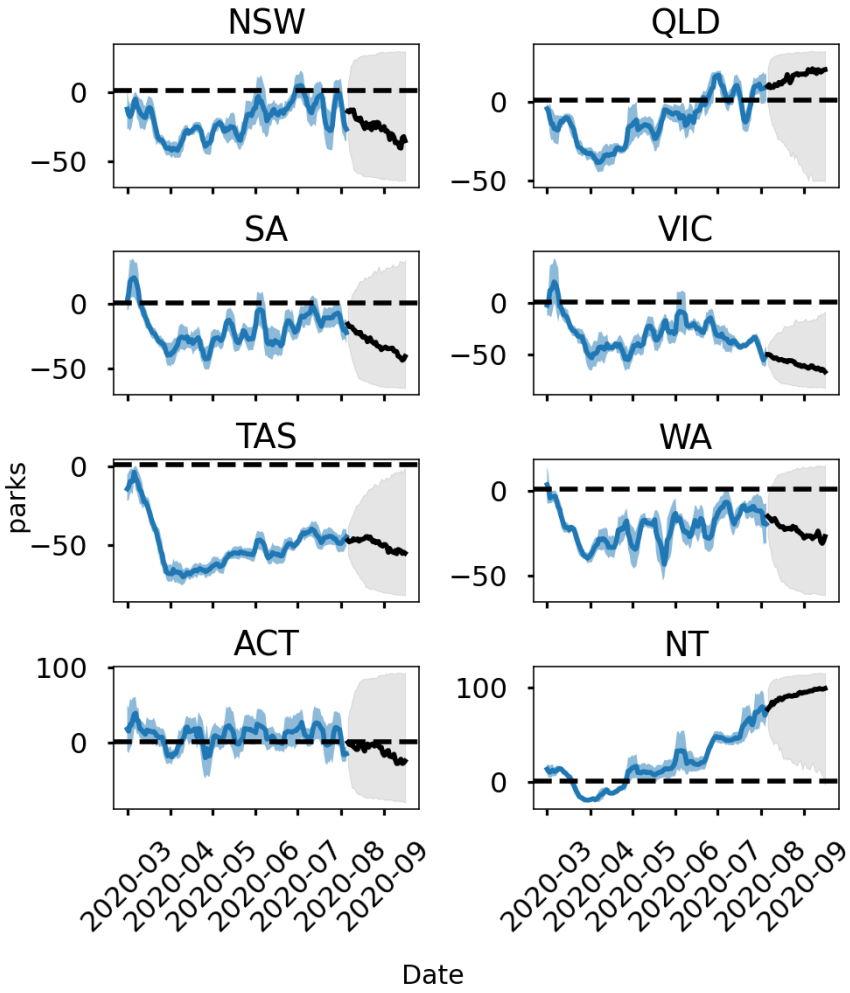

Figure C.7: The historical Google mobility index for visits to parks (blue) and the forecasted values (black) as of 2020-08-05. Solid lines represent the mean value while intervals are the 25% and 75% quantiles of the sampled distributions.

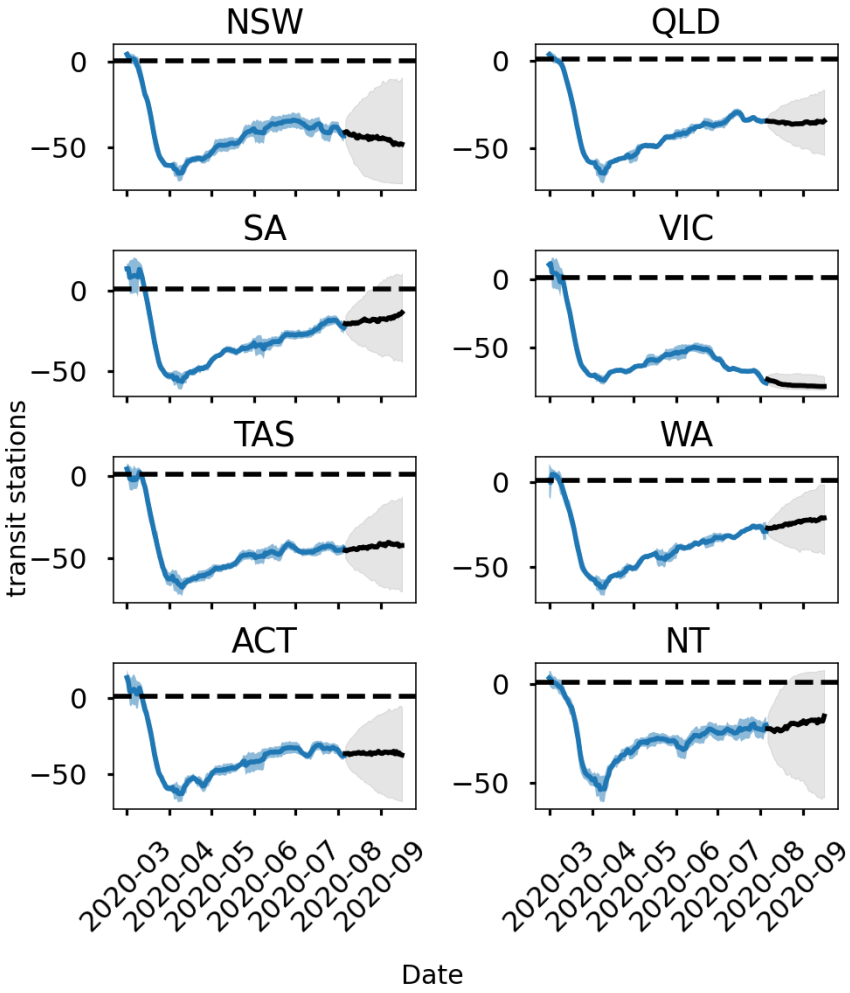

Figure C.8: The historical Google mobility index for visits to transit stations (blue) and the forecasted values (black) as of 2020-08-05. Solid lines represent the mean value while intervals are the 25% and 75% quantiles of the sampled distributions.

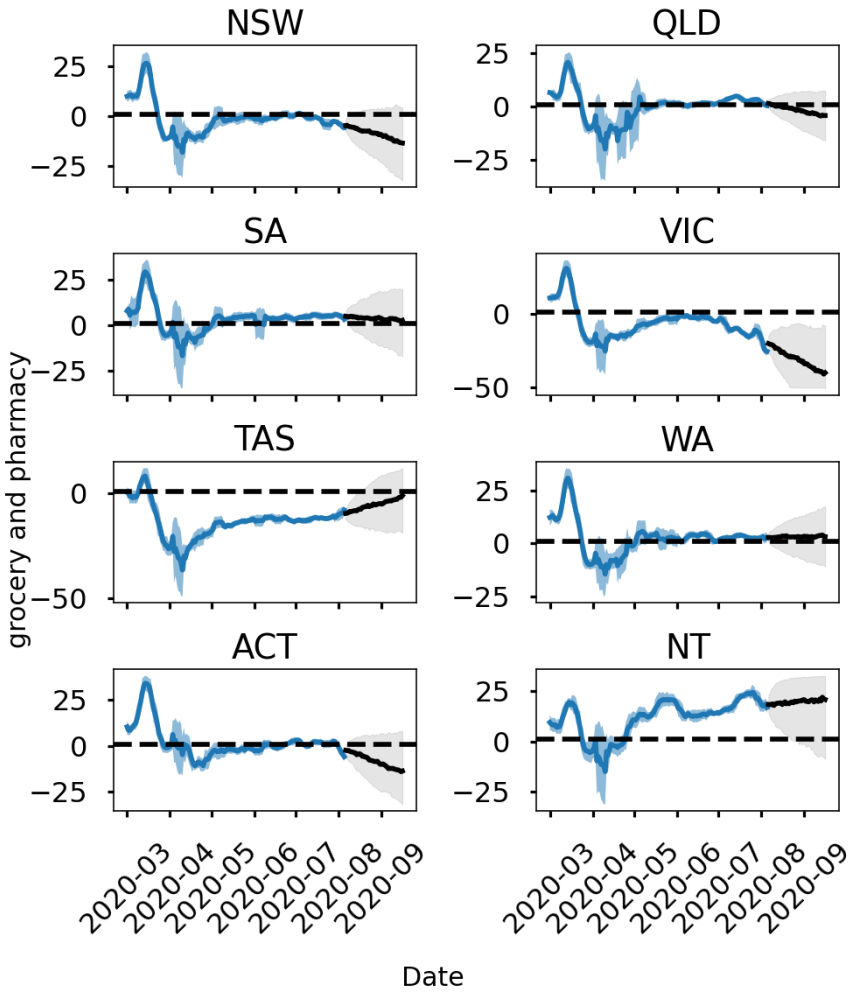

Figure C.9: The historical Google mobility index for visits to grocery and pharmacy locations (blue) and the forecasted values (black) as of 2020-08-05. Solid lines represent the mean value while intervals are the 25% and 75% quantiles of the sampled distributions.

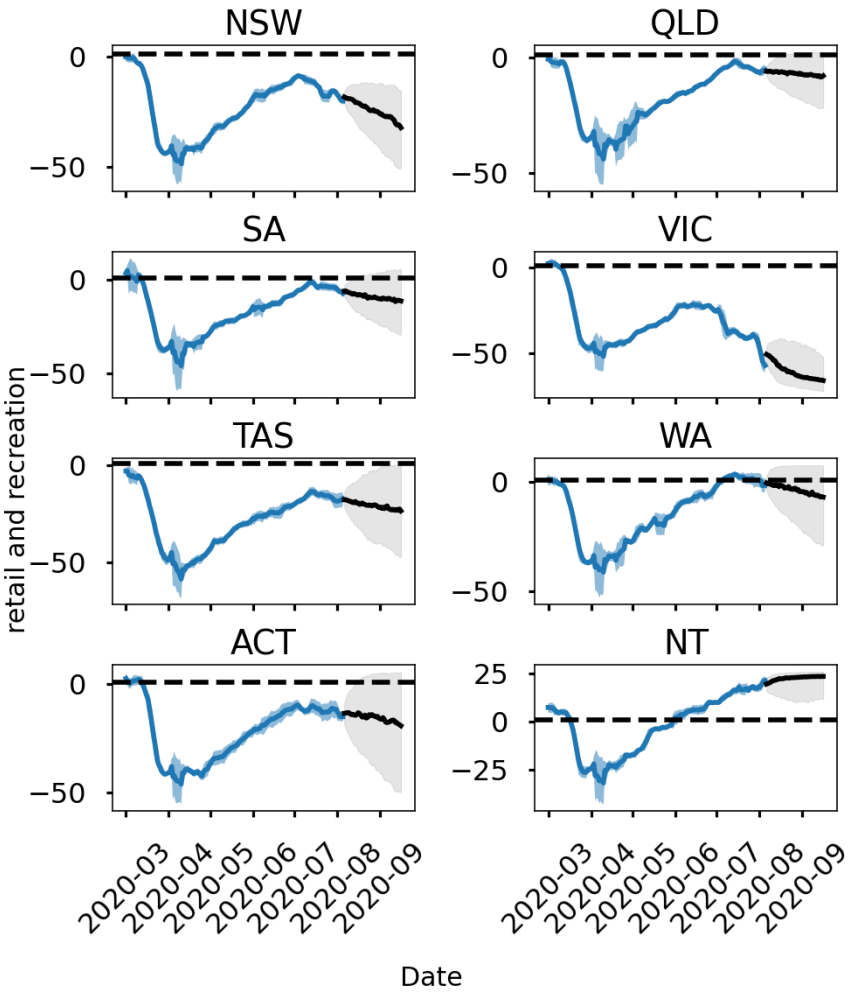

Figure C.10: The historical Google mobility index for visits to retail and recreation locations (blue) and the forecasted values (black) as of 2020-08-05. Solid lines represent the mean value while intervals are the 25% and 75% quantiles of the sampled distributions.

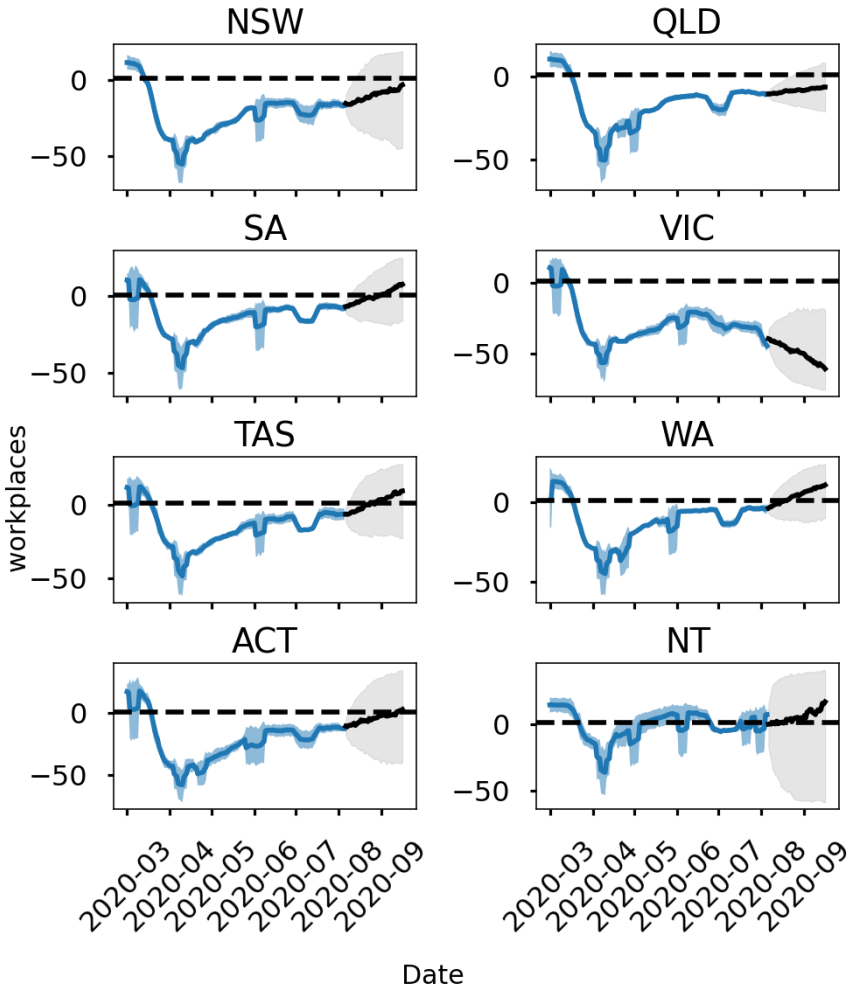

Figure C.11: The historical Google mobility index for visits to workplaces (blue) and the forecasted values (black) as of 2020-08-05. Solid lines represent the mean value while intervals are the 25% and 75% quantiles of the sampled distributions.

Appendix D

Supplementary material for Chapter 6: Evaluating contact tracing strategies

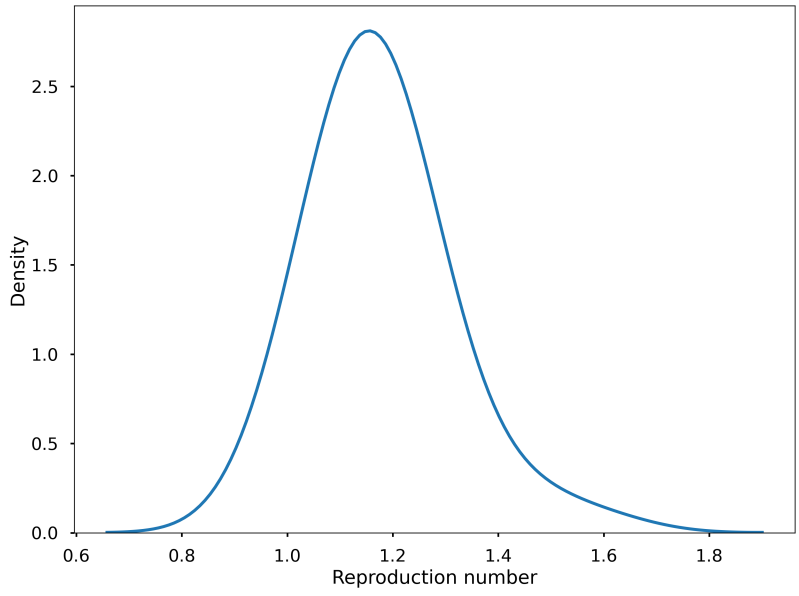

Figure D.1: The distribution of the reproduction number, and the mean of the negative binomial offspring distribution. This distribution has mean 1.20 and 90% credible interval 1.02 - 1.43.

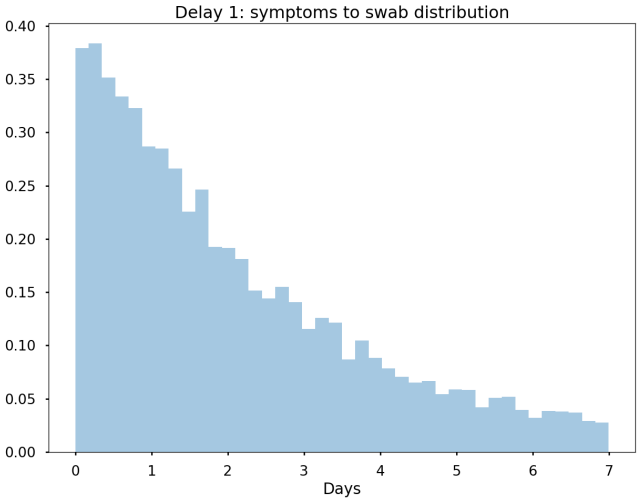

Figure D.2: The distribution of the delay between symptom onset to presentation for a test. Derived from NNDSS data and truncated at 7 days.

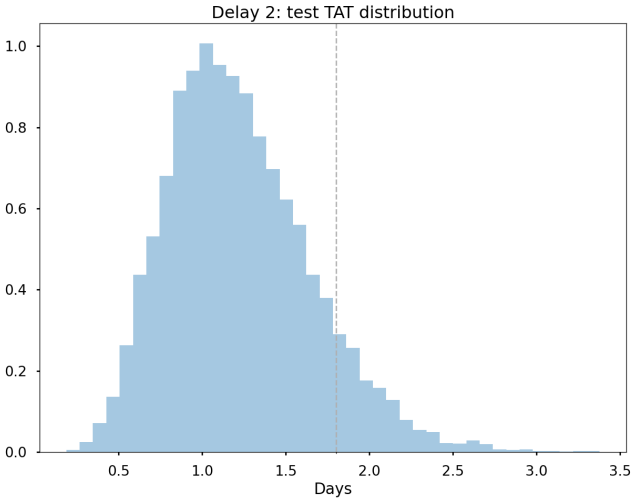

Figure D.3: The distribution of the delay between presentation for a test and notification of positive result to the PHU. The dashed line represents the 90th percentile, aligning with the COP indicator of less than 10% exceeding 48 hours. Derived from NNDSS data fields SPECIMEN COLLECTION DATE and NOTIFICATION DATE. This delay may be impacted by exceeding limits on testing capacity.

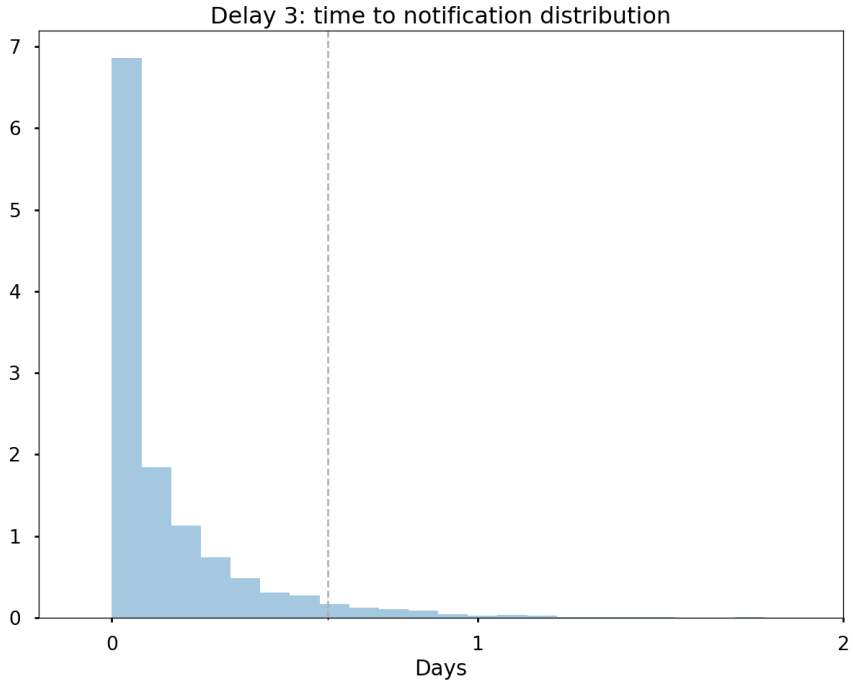

Figure D.4: The distribution of the delay between test completion of a positive case and the PHU notifying the index case. The dashed line represents the 99th percentile, aligning with the COP indicator of 0 cases exceeding 24 hours. Derived from NNDSS data fields NOTIFICATION DATE and NOTIFICATION RECEIVE DATE. This distribution is identical to the distribution of delay from conducting case interviews and notifying close contacts (Delay 4), and occurs after the index case is notified. Delay 4 may be impacted by exceeding limits on contact tracing capacity.

Appendix E

Statement of authorship

Title: "Elucidating user behaviours in a digital health surveillance system to correct prevalence estimates"

Publication status: Published

Publication details: Liu, D., Mitchell, L., Cope, RC., Carlson, SJ., Ross, JV. Elucidating user behaviours in a digital health surveillance system to correct prevalence estimates. Epidemics, 33, (2020)

Principle Author

Name of principal author: Dennis Liu

Contribution to the paper: Conceptualised, designed, and executed the methodology and analysis. Wrote the manuscript.

Overall percentage: 70%

Certification: This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Signed: ..... Date: .....

Secondary Author

Name of secondary author: Lewis Mitchell

Contribution to the paper: Conceptualisation, project direction and critically reviewed the process and manuscript.

Signed: ..... Date: .....

Secondary Author

Name of secondary author: Robert C. Cope

Contribution to the paper: Conceptualisation, project direction and critically reviewed the process and manuscript.

Signed: ..... Date: .....

Secondary Author

Name of secondary author: Sandra J. Carlson

Contribution to the paper: Curated and validated the data, and critically reviewed the process and manuscript.

Signed: ..... Date: .....

Secondary Author

Name of secondary author: Joshua V. Ross

Contribution to the paper: Conceptualisation, project direction and critically reviewed the process and manuscript.

Signed: ..... Date: .....

Bibliography

[1] Sam Abbott, Joel Hellewell, Robin Thompson, Katelyn Gostic, Katharine Sherratt, Sophie Meakin, James Munday, Nikos Bosse, Joe Hickson, Paul Mee, Peter Ellis, and Sebastian Funk. EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters. 2020.

[2] Richard Ambrosino, Bruce G. Buchanan, Gregory F. Cooper, and Michael J. Fine. The use of misclassification costs to learn rule-based decision support models for cost-effective hospital admission strategies. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pages 304–308, 1995.

[3] Australian Government. Special dates and events: School term dates, 2018.

[4] Australian Government Department of Health. 2018 Influenza Season in Australia: A summary from the National Influenza Surveillance Committee. (November):1–5, 2018.

[5] Australian Government Department of Health. Coronavirus (COVID-19) common operating picture, 2020.

[6] Paolo Bajardi, Alessandro Vespignani, Sebastian Funk, Ken T.D. Eames, W. John Edmunds, Clément Turbelin, Marion Debin, Vittoria Colizza, Ronald Smallenburg, Carl E. Koppeschaar, Ana O. Franco, Vitor Faustino, Annasara Carnahan, Moa Rehn, and Daniela Paolotti. Determinants of follow-up participation in the internet-based european influenza surveillance platform influenzanet. *Journal of Medical Internet Research*, 16(3):1–15, 2014.

[7] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J. Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of

infectious diseases: The global epidemic and mobility computational model. *Journal of Computational Science*, 1(3):132–145, 2010.

[8] Kristin Baltrusaitis, John S Brownstein, Samuel V Scarpino, Eric Bakota, Adam W Crawley, Giuseppe Conidi, Julia Gunn, Josh Gray, Anna Zink, and Mauricio Santillana. Comparison of crowd-sourced, electronic health records based, and traditional health-care based influenza-tracking systems at multiple spatial resolutions in the United States of America. *BMC infectious diseases*, 18(1):1–8, 2018.

[9] Kristin Baltrusaitis, Mauricio Santillana, Adam W Crawley, Rumi Chunara, Mark Smolinski, and John S Brownstein. Determinants of Participants' Follow-Up and Characterization of Representativeness in Flu Near You, A Participatory Disease Surveillance System. *JMIR Public Health and Surveillance*, 3(2):e18, 2017.

[10] Shweta Bansal, Gerardo Chowell, Lone Simonsen, Alessandro Vespignani, and Cécile Viboud. Big data for infectious disease surveillance and modeling. *Journal of Infectious Diseases*, 214(Suppl 4):S375–S379, 2016.

[11] Michael Betancourt, Simon Byrne, Sam Livingstone, and Mark Girolami. The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli*, 23(4A):2257–2298, 2017.

[12] Aleta B Bonner, Kathy W Monroe, Lynya I Talley, Ann E Klasner, and David W Kimberlin. Impact of the Rapid Diagnosis of Influenza on Physician Decision- Making and Patient Management in the Pediatric Emergency Department: Results of a Randomized, Prospective, Controlled Trial. *Pediatrics*, 112(2):363–367, 2003.

[13] Jochen Bröcker. Evaluating raw ensembles with the continuous ranked probability score. *Quarterly Journal of the Royal Meteorological Society*, 138(667):1611–1617, 2012.

[14] David A. Broniatowski, Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate. *American Journal of Public Health*, pages e1–e7, 2018.

[15] Diana Buitrago-Garcia, Dianne Egli-Gany, Michel J. Counotte, Stefanie Hossmann, Hira Imeri, Aziz Mert Ipekci, Georgia Salanti, and

Nicola Low. Occurrence and transmission potential of asymptomatic and presymptomatic SARSCoV-2 infections: A living systematic review and meta-analysis. *PLoS Medicine*, 17(9):1–25, 2020.

[16] Pietro Cantarelli, Marion Debin, Clément Turbelin, Chiara Poletto, Thierry Blanchon, Alessandra Falchi, Thomas Hanslik, Isabelle Bonmarin, Daniel Levy-Bruhl, Alessandra Micheletti, Daniela Paolotti, Alessandro Vespignani, John Edmunds, Ken Eames, Ronald Smallenburg, Carl Koppeschaar, Ana O. Franco, Vitor Faustino, Annasara Carnahan, Moa Rehn, and Vittoria Colizza. The representativeness of a European multi-center network for influenza-like-illness participatory surveillance. *BMC Public Health*, 14(1):1–17, 2014.

[17] Sandra J Carlson, Daniel Cassano, Michelle T Butler, David N Durrheim, and Craig Dalton. Flutracking weekly online community survey of influenza-like illness annual report, 2016. *Communicable Diseases Intelligence*, 43, 2019.

[18] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 2017.

[19] Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. Intelligible Models for HealthCare. pages 1721–1730, 2015.

[20] Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky, and Jure Leskovec. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 2020.

[21] Vivek Charu, Scott Zeger, Julia Gog, Ottar N. Bjørnstad, Stephen Kissler, Lone Simonsen, Bryan T. Grenfell, and Cécile Viboud. Human mobility and the spatial transmission of influenza in the United States. *PLoS Computational Biology*, 13(2):1–24, 2017.

[22] Rumi Chunara, Susan Aman, Mark Smolinski, and John S. Brownstein. Flu Near You: An Online Self-reported Influenza Surveillance System in the USA. *Online Journal of Public Health Informatics*, 5(1):2579, 2013.

[23] Hazel J Clothier, James E Fielding, and Heath A Kelly. An evaluation of the Australian Sentinel Practice Research Network (ASPREN)

surveillance for influenza-like illness. *Communicable Diseases Intelligence*, 29(3):231–247, 2005.

[24] Shannon Collinson, Kamran Khan, and Jane M. Heffernan. The effects of media reports on disease spread and important public health measurements. *PLoS ONE*, 10(11):1–21, 2015.

[25] Gregory F. Cooper, Vijoy Abraham, Constantin F. Aliferis, John M. Aronis, Bruce G. Buchanan, Richard Caruana, Michael J. Fine, Janine E. Janosky, Gary Livingston, Tom Mitchell, Stefano Monti, and Peter Spirtes. Predicting dire outcomes of patients with community acquired pneumonia. *Journal of Biomedical Informatics*, 38(5):347–366, 2005.

[26] Robert C. Cope, Joshua V. Ross, Monique Chilver, Nigel P. Stocks, and Lewis Mitchell. Characterising seasonal influenza epidemiology using primary care surveillance data. *PLoS Computational Biology*, 14(8):1–21, 2018.

[27] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, 2013.

[28] Craig Dalton, David Durrheim, John Fejsa, Lynn Francis, Sandra Carlson, Edouard Tursan D'Espaignet, and Frank Tuyl. Flutracking: A weekly Australian community online survey of influenza-like illness in 2006, 2007 and 2008. *Communicable Diseases Intelligence Quarterly Report*, 33(IlI):316–322, 2009.

[29] Mark D.M. Davis, Niamh Stephenson, Davina Lohm, Emily Waller, and Paul Flowers. Beyond resistance: Social factors in the general public response to pandemic influenza. *BMC Public Health*, 15(1):1–9, 2015.

[30] Rachael Dexter and Paul Sakkal. 'Chaos': South Australia lockdown sparks panic buying with toilet paper stripped from shelves, nov 2020.

[31] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.

[32] Gytis Dudas, Luiz Max Carvalho, Trevor Bedford, Andrew J. Tatem, Guy Baele, Nuno R. Faria, Daniel J. Park, Jason T. Ladner, Armando Arias, Danny Asogun, Filip Bielejec, Sarah L. Caddy, Matthew Cotten, Jonathan D'Ambrozio, Simon Dellicour, Antonino Di Caro, Joseph W. Diclaro, Sophie Duraffour, Michael J. Elmore, Lawrence S. Fakoli, Ousmane Faye, Merle L. Gilbert, Sahr M. Gevao, Stephen Gire, Adrianne Gladden-Young, Andreas Gnirke, Augustine Goba, Donald S. Grant, Bart L. Haagmans, Julian A. Hiscox, Umaru Jah, Jeffrey R. Kugelman, Di Liu, Jia Lu, Christine M. Malboeuf, Suzanne Mate, David A. Matthews, Christian B. Matranga, Luke W. Meredith, James Qu, Joshua Quick, Suzan D. Pas, My V.T. Phan, Georgios Pollakis, Chantal B. Reusken, Mariano Sanchez-Lockhart, Stephen F. Schaffner, John S. Schieffelin, Rachel S. Sealfon, Etienne Simon-Loriere, Saskia L. Smits, Kilian Stoecker, Lucy Thorne, Ekaete Alice Tobin, Mohamed A. Vandi, Simon J. Watson, Kendra West, Shannon Whitmer, Michael R. Wiley, Sarah M. Winnicki, Shirlee Wohl, Roman Wölfel, Nathan L. Yozwiak, Kristian G. Andersen, Sylvia O. Blyden, Fatorma Bolay, Miles W. Carroll, Bernice Dahn, Boubacar Diallo, Pierre Formenty, Christophe Fraser, George F. Gao, Robert F. Garry, Ian Goodfellow, Stephan Günther, Christian T. Happi, Edward C. Holmes, Brima Kargbo, Sakoba Keïta, Paul Kellam, Marion P.G. Koopmans, Jens H. Kuhn, Nicholas J. Loman, N'Faly Magassouba, Dhamari Naidoo, Stuart T. Nichol, Tolbert Nyenswah, Gustavo Palacios, Oliver G. Pybus, Pardis C. Sabeti, Amadou Sall, Ute Ströher, Isatta Wurie, Marc A. Suchard, Philippe Lemey, and Andrew Rambaut. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*, 544(7650):309–315, 2017.

[33] Adam G. Dunn, Didi Surian, Julie Leask, Aditi Dey, Kenneth D. Mandl, and Enrico Coiera. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. *Vaccine*, 35(23):3033–3040, 2017.

[34] Ken T D Eames. Networks of influence and infection: parental choices and childhood disease. *Journal of The Royal Society Interface*, 6(38):811–814, 2009.

[35] Akira Endo, Sam Abbott, Adam J. Kucharski, and Sebastian Funk. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Research*, 5:67, 2020.

[36] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

[37] Neil M Ferguson, Daniel Laydon, Gemma Nedjati-gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-dannenburg, Amy Dighe, Han Fu, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Lucy C Okell, Sabine Van, Hayley Thompson, Robert Verity, Erik Volz, Haowei Wang, Yuanrong Wang, Patrick G T Walker, Caroline Walters, Peter Winskill, Charles Whittaker, Christl A Donnelly, Steven Riley, and Azra C Ghani. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID- 19 mortality and healthcare demand. (March), 2020.

[38] Luca Ferretti, Alice Ledda, Chris Wymant, Lele Zhao, Virginia Ledda, Lucie Abeler, Michelle Kendall, Anel Nurtay, Hao-yuan Cheng, Tachou Ng, Hsien-ho Lin, Rob Hinch, Joanna Masel, A Marm Kilpatrick, and Christophe Fraser. The timing of COVID-19 transmission. *medRxiv*, pages 1–16, 2020.

[39] Seth Flaxman, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W. Eaton, Mélodie Monod, Pablo N. Perez-Guzman, Nora Schmit, Lucia Cilloni, Kylie E.C. Ainslie, Marc Baguelin, Adhiratha Boonyasiri, Olivia Boyd, Lorenzo Cattarino, Laura V. Cooper, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djaafara, Ilaria Dorigatti, Sabine L. van Elsland, Richard G. FitzJohn, Katy A.M. Gaythorpe, Lily Geidelberg, Nicholas C. Grassly, William D. Green, Timothy Hallett, Arran Hamlet, Wes Hinsley, Ben Jeffrey, Edward Knock, Daniel J. Laydon, Gemma Nedjati-Gilani, Pierre Nouvellet, Kris V. Parag, Igor Siveroni, Hayley A. Thompson, Robert Verity, Erik Volz, Caroline E. Walters, Haowei Wang, Yuanrong Wang, Oliver J. Watson, Peter Winskill, Xiaoyue Xi, Patrick G.T. Walker, Azra C. Ghani, Christl A. Donnelly, Steven Riley, Michaela A.C. Vollmer, Neil M. Ferguson, Lucy C. Okell, and Samir Bhatt. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, 584(7820):257–261, 2020.

[40] Sebastian Funk, Shweta Bansal, Chris T. Bauch, Ken T.D. Eames, W. John Edmunds, Alison P. Galvani, and Petra Klepac. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics*, 10:21–25, 2015.

[41] Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent A.A. Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 106(16):6872–6877, 2009.

[42] Tapiwa Ganyani, Cécile Kremer, Dongxuan Chen, Andrea Torneri, Christel Faes, Jacco Wallinga, and Niel Hens. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance*, 25(17):1–8, apr 2020.

[43] Jemma L Geoghegan, Aldo F Saavedra, Sebastian Duchene, Sheena Sullivan, Ian Barr, and Edward C Holmes. Continental synchronicity of human influenza virus epidemics despite climactic variation. *PLoS Pathogens*, 14(1):1–16, 2018.

[44] Walter R. Gilks, Sylvia Richardson, and David J. Spiegelhalter. Markov Chain Monte Carlo in Practice. In *Markov Chain Monte Carlo in Practice*, page 512. Chapman and Hall, 1996.

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

[46] Julia R. Gog, Sébastien Ballesteros, Cécile Viboud, Lone Simonsen, Ottar N. Bjornstad, Jeffrey Shaman, Dennis L. Chao, Farid Khan, and Bryan T. Grenfell. Spatial Transmission of 2009 Pandemic Influenza in the US. *PLoS Computational Biology*, 10(6), 2014.

[47] Nick Golding, Freya M Shearer, Robert Moss, Peter Dawson, Dennis Liu, Joshua V Ross, Rob Hyndman, Cameron Zachreson, Nic Geard, Jodie Mcvernon, and David J Price. Estimating temporal variation in transmission of SARS-CoV-2 and physical distancing behaviour in Australia. *Doherty Institute*, (July):1–50, 2020.

[48] Marta C. González, César A. Hidalgo, and Albert László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[49] Google LLC. Google COVID-19 Community Mobility Reports, 2020.

[50] Katelyn M. Gostic, Lauren McGough, Edward B. Baskerville, Sam Abbott, Keya Joshi, Christine Tedijanto, Rebecca Kahn, Rene Niehus, James A. Hay, Pablo M. De Salazar, Joel Hellewell, Sophie Meakin,

James D. Munday, Nikos I. Bosse, Katharine Sherrat, Robin N. Thompson, Laura F. White, Jana S. Huisman, Jérémie Scire, Sebastian Bonhoeffer, Tanja Stadler, Jacco Wallinga, Sebastian Funk, Marc Lipsitch, and Sarah Cobey. Practical considerations for measuring the effective reproductive number, Rt. *PLOS Computational Biology*, 16(12):e1008409, dec 2020.

[51] Bryan T. Grenfell, Ottar N Bjørnstad, and Jens Kappey. Travelling waves and spartial hierarchies in measles epidemics. *Nature*, 414(December):716–723, 2001.

[52] Jamie T. Griffin, Tini Garske, Azra C. Ghani, and Paul S. Clarke. Joint estimation of the basic reproduction number and generation time parameters for infectious disease outbreaks. *Biostatistics*, 12(2):303–312, 2011.

[53] Caroline Guerrisi, Clément Turbelin, Thierry Blanchon, Thomas Hanslik, Isabelle Bonmarin, Daniel Levy-Bruhl, Daniela Perrotta, Daniela Paolotti, Ronald Smallenburg, Carl Koppeschaar, Ana O. Franco, Ricardo Mexia, W. John Edmunds, Bersabeh Sile, Richard Pebody, Edward Van Straten, Sandro Meloni, Yamir Moreno, Jim Duggan, Charlotte Kjelsø, and Vittoria Colizza. Participatory syndromic surveillance of influenza in Europe. *Journal of Infectious Diseases*, 214(May):S386–S392, 2016.

[54] Caroline Guerrisi, Clément Turbelin, Cécile Souty, Chiara Poletto, Thierry Blanchon, Thomas Hanslik, Isabelle Bonmarin, Daniel Levy-Bruhl, and Vittoria Colizza. The potential value of crowdsourced surveillance systems in supplementing sentinel influenza networks: The case of France. *Eurosurveillance*, 23(25), 2018.

[55] M. Elizabeth Halloran, Neil M. Ferguson, Stephen Eubank, Ira M. Longini, Derek A.T. Cummings, Bryan Lewis, Shufu Xu, Christophe Fraser, Anil Vullikanti, Timothy C. Germann, Diane Wagener, Richard Beckman, Kai Kadau, Chris Barrett, Catherine A. Macken, Donald S. Burke, and Philip Cooley. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 105(12):4639–4644, 2008.

[56] William S. Hart, Philip K. Maini, and Robin N. Thompson. High infectiousness immediately before COVID-19 symptom onset highlights the importance of contact tracing. *medRxiv*, 2020.

[57] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.*, volume 27. Second edition, 2009.

[58] Xi He, Eric H.Y. Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y. Wong, Yujuan Guan, Xinghua Tan, Xiaoneng Mo, Yanqing Chen, Baolin Liao, Weilie Chen, Fengyu Hu, Qing Zhang, Mingqiu Zhong, Yanrong Wu, Lingzhai Zhao, Fuchun Zhang, Benjamin J. Cowling, Fang Li, and Gabriel M. Leung. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26(5):672–675, 2020.

[59] Samuel Henly, Gaurav Tuli, Sheryl A. Kluberg, Jared B. Hawkins, Quynh C. Nguyen, Aranka Anema, Adyasha Maharana, John S. Brownstein, and Elaine O. Nsoesie. Disparities in digital reporting of illness: A demographic and socioeconomic assessment. *Preventive Medicine*, 101:18–22, 2017.

[60] H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5):559–570, 2000.

[61] Yoko Ibuka, Gretchen B. Chapman, Lauren A. Meyers, Meng Li, and Alison P. Galvani. The dynamics of risk perceptions and precautionary behavior in response to 2009 (H1N1) pandemic influenza. *BMC Infectious Diseases*, 10(1):296, 2010.

[62] Michael L. Jackson and Jennifer C. Nelson. The test-negative design for estimating influenza vaccine effectiveness. *Vaccine*, 31(17):2165–2168, 2013.

[63] Nicholas P. Jewell. *Statistics for Epidemiology.* CRC Press, 2003.

[64] Peng jun Lu, Alissa O'Halloran, Erin D. Kennedy, Walter W. Williams, David Kim, Amy Parker Fiebelkorn, Sara Donahue, and Carolyn B. Bridges. Awareness among adults of vaccine-preventable diseases and recommended vaccinations, United States, 2015. *Vaccine*, 35(23):3104–3115, 2017.

[65] Raja Jurdak, Kun Zhao, Jiajun Liu, Maurice AbouJaoude, Mark Cameron, and David Newth. Understanding human mobility from Twitter. *PLoS ONE*, 10(7):1–16, 2015.

[66] Kyriaki Kalimeri, Matteo Delfino, Ciro Cattuto, Daniela Perrotta, Vittoria Colizza, Caroline Guerrisi, Clement Turbelin, Jim Duggan, John Edmunds, Chinelo Obi, Richard Pebody, Ricardo Mexia, Ana Franco, Yamir Moreno, Sandro Meloni, Carl Koppeschaar, Charlotte Kjelsø, Daniela Paolotti, and Ricardo Jorge. Unsupervised Extraction of Epidemic Syndromes from Participatory Influenza Surveillance Self-reported Symptoms. pages 1–24, 2018.

[67] Matt J. Keeling and Pejman Rohani. *Modeling Infectious Diseases*. Princeton University Press, 2008.

[68] Petra Klepac, Stephen Kissler, and Julia Gog. Contagion! The BBC Four Pandemic - the model behind the documentary. *Epidemics*, (March):0–1, 2018.

[69] Carl E Koppeschaar, Vittoria Colizza, Caroline Guerrisi, Clément Turbelin, Jim Duggan, W John Edmunds, Charlotte Kjelsø, Ricardo Mexia, Yamir Moreno, Sandro Meloni, Daniela Paolotti, Daniela Perrotta, Edward van Straten, and Ana O Franco. Influenzanet: Citizens Among 10 Countries Collaborating to Monitor Influenza in Europe. *JMIR Public Health and Surveillance*, 3(3):e66, 2017.

[70] Mirjam E. Kretzschmar, Ganna Rozhnova, Martin C.J. Bootsma, Michiel van Boven, Janneke H.H.M. van de Wijgert, and Marc J.M. Bonten. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health*, 5(8):e452–e459, 2020.

[71] Heidi J Larson. The biggest pandemic risk? Viral misinformation. *Nature*, 562(7727):309–309, 2018.

[72] Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172(9):577–582, 2020.

[73] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The Parable of Google Flu: Traps in Big Data Analysis. 343(March):1203–1205, 2014.

[74] Elizabeth C. Lee, Jason M. Asher, Sandra Goldlust, John D. Kraemer, Andrew B. Lawson, and Shweta Bansal. Mind the scales: Harness-

ing spatial big data for infectious disease surveillance and inference. *Journal of Infectious Diseases*, 214(Suppl 4):S409–S413, 2016.

[75] Dennis Liu, Lewis Mitchell, Robert C. Cope, Sandra J. Carlson, and Joshua V. Ross. Elucidating user behaviours in a digital health surveillance system to correct prevalence estimates. *Epidemics*, 33:100404, 2020.

[76] James O. Lloyd-Smith, Sebastian J. Schreiber, Peter Ekkehard Kopp, and Wayne Marcus Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, 2005.

[77] Fred Sun Lu, Suqin Hou, Kristin Baltrusaitis, Manan Shah, Jure Leskovec, Rok Sosic, Jared Hawkins, John Brownstein, Giuseppe Conidi, Julia Gunn, Josh Gray, Anna Zink, and Mauricio Santillana. Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. *JMIR Public Health and Surveillance*, 4(1):e4, 2018.

[78] Richard McElreath. *Statistical Rethinking: a Bayesian course with examples in R and Stan*. CRC Press, 2016.

[79] Daniella Meeker, Jeffrey A. Linder, Craig R. Fox, Mark W. Friedberg, Stephen D. Persell, Noah J. Goldstein, Tara K. Knight, Joel W. Hay, and Jason N. Doctor. Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices a randomized clinical trial. *JAMA - Journal of the American Medical Association*, 315(6):562–570, 2016.

[80] Stefano Merler, Marco Ajelli, Laura Fumanelli, Marcelo F.C. Gomes, Ana Pastore y. Piontti, Luca Rossi, Dennis L. Chao, Ira M. Longini, M. Elizabeth Halloran, and Alessandro Vespignani. Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: A computational modelling analysis. *The Lancet Infectious Diseases*, 15(2):204–211, 2015.

[81] Andrew V. Metcalfe and Paul S.P. Cowpertwait. *Introductory Time Series with R*, volume 31. Springer, New York, 2009.

[82] Lewis Mitchell and Joshua V Ross. A data-driven model for influenza transmission incorporating media effects. *The Royal Society Open Science*, 3(10):1–10, 2016.

[83] Kelly R. Moran, Geoffrey Fairchild, Nicholas Generous, Kyle Hickmann, Dave Osthus, Reid Priedhorsky, James Hyman, and Sara Y. Del Valle. Epidemic forecasting is messier than weather forecasting: The role of human behavior and internet data streams in epidemic forecast. *Journal of Infectious Diseases*, 214(Suppl 4):S404–S408, 2016.

[84] Robert Moss, Alexander Zarebski, Sandra Carlson, and James McCaw. Accounting for Healthcare-Seeking Behaviours and Testing Practices in Real-Time Influenza Forecasts. *Tropical Medicine and Infectious Disease*, 4(1):12, 2019.

[85] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W. John Edmunds. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, 5(3):0381–0391, 2008.

[86] Allan H. Murphy. Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient, 1988.

[87] Ruchit Nagar, Qingyu Yuan, Clark C. Freifeld, Mauricio Santillana, Aaron Nojima, Rumi Chunara, and John S. Brownstein. A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *Journal of Medical Internet Research*, 16(10):e236, 2014.

[88] New South Wales Government Health. COVID-19 weekly surveillance reports, 2020.

[89] Hiroshi Nishiura, Tetsuro Kobayashi, Takeshi Miyama, Ayako Suzuki, Sung mok Jung, Katsuma Hayashi, Ryo Kinoshita, Yichi Yang, Baoyin Yuan, Andrei R. Akhmetzhanov, and Natalie M. Linton. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International Journal of Infectious Diseases*, 94(January):154–155, 2020.

[90] Elaine O. Nsoesie, Luisa Flor, Jared Hawkins, Adyasha Maharana, Tobi Skotnes, Fatima Marinho, and John S. Brownstein. Social Media as a Sentinel for Disease Surveillance: What Does Sociodemographic Status Have to Do with It? *PLoS currents*, 8:1–26, 2016.

[91] Daniela Paolotti, Anna Sara Carnahan, Vittoria Colizza, Ken T.D. Eames, John Edmunds, M. Gabriela M Gomes, C. Koppeschaar, Moa

Rehn, Ronald Smallenburg, Clément Turbelin, Sander P. van Noort, and Alessandro Vespignani. Web-based participatory surveillance of infectious diseases: The Influenzanet participatory surveillance experience. *Clinical Microbiology and Infection*, 20(1):17–21, 2014.

[92] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Oliver Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(1):2825–2830, 2011.

[93] Sen Pei, Sasikiran Kandula, Wan Yang, and Jeffrey Shaman. Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences*, 115(11):2752–2757, 2018.

[94] Maria Peppa, W. John Edmunds, and Sebastian Funk. Disease severity determines health-seeking behaviour amongst individuals with influenza-like illness in an internet-based cohort. *BMC Infectious Diseases*, 17(1):1–13, 2017.

[95] David J Price, Freya M Shearer, Michael Meehan, Emma Mcbryde, Nick Golding, Jodie Mcvernon, and James M Mccaw. Estimating the case detection rate and temporal variation in transmission of COVID-19 in Australia Technical Report 14 th April 2020. (April), 2020.

[96] Jacek Radzikowski, Anthony Stefanidis, Kathryn H Jacobsen, Arie Croitoru, Andrew Crooks, and Paul L Delamater. The Measles Vaccination Narrative in Twitter: A Quantitative Analysis. *JMIR Public Health and Surveillance*, 2(1):e1, 2016.

[97] Carrie Reed, Sandra S. Chaves, Pam Daily Kirley, Ruth Emerson, Deborah Aragon, Emily B. Hancock, Lisa Butler, Joan Baumbach, Gary Hollick, Nancy M. Bennett, Matthew R. Laidler, Ann Thomas, Martin I. Meltzer, and Lyn Finelli. Estimating Influenza Disease Burden from Population-Based Surveillance Data in the United States. *PLoS ONE*, 10(3):e0118369, 2015.

[98] Moa Rehn, Anna Sara Carnahan, Hanna Merk, Sharon Kühlmann-Berenzon, Ilias Galanis, Annika Linde, and Olof Nyrén. Evaluation of an internet-based monitoring system for influenza-like illness in Sweden. *PLoS ONE*, 9(5):1–10, 2014.

[99] Ralf Reintjes, Enny Das, Celine Klemm, Jan Hendrik Richardus, Verena Keßler, and Amena Ahmad. "Pandemic public health paradox": Time series analysis of the 2009/10 influenza A/H1N1 epidemiology, media attention, risk perception and public reactions in 5 European countries. *PLoS ONE*, 11(3):1–14, 2016.

[100] Timothy W. Russell, Nick Golding, Joel Hellewell, Sam Abbott, Lawrence Wright, Carl A.B. Pearson, Kevin van Zandvoort, Christopher I. Jarvis, Hamish Gibbs, Yang Liu, Rosalind M. Eggo, W. John Edmunds, Adam J. Kucharski, Arminder K. Deol, C. Julian Villabona-Arenas, Thibaut Jombart, Kathleen O'Reilly, James D. Munday, Sophie R. Meakin, Rachel Lowe, Amy Gimma, Akira Endo, Emily S. Nightingale, Graham Medley, Anna M. Foss, Gwenan M. Knight, Kiesha Prem, Stéphane Hué, Charlie Diamond, James W. Rudge, Katherine E. Atkins, Megan Auzenbergs, Stefan Flasche, Rein M.G.J. Houben, Billy J. Quilty, Petra Klepac, Matthew Quaife, Sebastian Funk, Quentin J. Leclerc, Jon C. Emery, Mark Jit, David Simons, Nikos I. Bosse, Simon R. Procter, Fiona Yueqian Sun, Samuel Clifford, Katharine Sherratt, Alicia Rosello, Nicholas G. Davies, Oliver Brady, Damien C. Tully, and Georgia R. Gore-Langton. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Medicine*, 18(1):1–9, 2020.

[101] Darpun D. Sachdev, Hannah K. Brosnan, Michael J. A. Reid, Michelle Kirian, Stephanie E. Cohen, Trang Q. Nguyen, and Susan Scheer. Outcomes of Contact Tracing in San Francisco, CaliforniaTest and Trace During Shelter-in-Place. *JAMA Internal Medicine*, 2020.

[102] Marcel Salathe and Sebastian Bonhoeffer. The effect of opinion clustering on disease outbreaks. *Journal of The Royal Society Interface*, 5(29):1505–1508, 2008.

[103] Marcel Salathé and Shashank Khandelwal. Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control. *PLoS Computational Biology*, 7(10), 2011.

[104] Lone Simonsen, Julia R. Gog, Don Olson, and Cécile Viboud. Infectious disease surveillance in the big data era: Towards faster and locally relevant systems. *Journal of Infectious Diseases*, 214(November):S380–S385, 2016.

[105] Paul Smith. Covid-19 in Australia: most infected health workers in

Victoria's second wave acquired virus at work. *BMJ (Clinical research ed.)*, 370(August):m3350, 2020.

[106] Mark S. Smolinski, Adam W. Crawley, Kristin Baltrusaitis, Rumi Chunara, Jennifer M. Olsen, Oktawia Wójcik, Mauricio Santillana, Andre Nguyen, and John S. Brownstein. Flu near you: Crowdsourced symptom reporting spanning 2 influenza seasons. *American Journal of Public Health*, 105(10):2124–2130, 2015.

[107] John Snow. *On the Mode of Communication of Cholera*. London, 1855.

[108] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. *AI 2006: Advances in Artificial Intelligence*, 4304(1):1015–1021, 2006.

[109] Stan Development Team. Stan Modeling Language: User's Guide and Reference Manual. *Version 2.17.0*, 2017.

[110] Sheena G. Sullivan, Jane Raupach, Lucinda J. Franklin, Kate Pennington, Christina Bareja, and Rachel de Kluyver. A brief overview of influenza surveillance systems in Australia, 2015. *Communicable diseases intelligence quarterly report*, 40(3):E351–E355, 2016.

[111] Kaiyuan Sun, Wei Wang, Lidong Gao, Yan Wang, Kaiwei Luo, Lingshuang Ren, Zhifei Zhan, Xinghui Chen, Shanlu Zhao, Yiwei Huang, Qianlai Sun, Ziyan Liu, Maria Litvinova, Alessandro Vespignani, Marco Ajelli, Cécile Viboud, and Hongjie Yu. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *medRxiv*, 2424(November):2020.08.09.20171132, 2020.

[112] Robin N. Thompson, Jake E. Stockwin, Rolina. D. van Gaalen, Jonathan A. Polonsky, Zhian N. Kamvar, P. Alex Demarsh, Elisabeth Dahlqwist, Sixuan Li, Eve Miguel, Thibaut Jombart, Justin Lessler, Simon Cauchemez, and Anne Cori. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29(July), 2019.

[113] Sander P. van Noort, Claudia T. Codeco, Carl E. Koppeschaar, Marc van Ranst, Daniela Paolotti, and M. Gabriela M Gomes. Ten-year performance of Influenzanet: ILI time series, risks, vaccine effects, and care-seeking behaviour. *Epidemics*, 13:28–36, 2015.

[114] Michaela Vollmer, Swapnil Mishra, H Juliette T Unwin, Axel Gandy, Thomas Mellan, Valerie Bradley, Harrison Zhu, Helen Coupland, Iwona Hawryluk, Michael Hutchinson, Oliver Ratmann, Melodie Monod, Patrick Walker, Charlie Whittaker, Lorenzo Cattarino, Constance Ciavarella, Lucia Cilloni, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Nicholas Brazeau, Giovanni Charles, Laura Cooper, Zulma Cucunuba, Gina Cuomo-Dannenburg, Amy Dighe, Bimandra Djaafara, Jeff Eaton, Sabine van Elsland, Richard FitzJohn, Keith Fraser, Katy Gaythorpe, Will Green, Sarah Hayes, Natsuko Imai, Ben Jeffrey, Edward Knock, Daniel Laydon, John Lees, Tara Mangal, Andria Mousa, Gemma Nedjati-Gilani, Pierre Nouvellet, Daniela Olivera, Kris Parag, Michael Pickles, Hayley Thompson, Robert Verity, Caroline Walters, Haowei Wang, Yuanrong Wang, Oliver Watson, Lilith Whittles, Xiaoyue Xi, Azra Ghani, Steven Riley, Lucy Okell, Christl Donnelly, Neil Ferguson, Ilaria Dorigatti, Seth Flaxman, and Samir Bhatt. A sub-national analysis of the rate of transmission of COVID-19 in Italy. *35*, (May), 2020.

[115] Jacco Wallinga and Marc Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, 2007.

[116] World Health Organization. Global Epidemiological Surveillance Standards for Influenza, 2014.

[117] World Health Organization. Influenza (Seasonal) fact sheet, 2018.

[118] Cameron Zachreson, Lewis Mitchell, Michael J. Lydeamore, Nicolas Rebuli, Martin Tomko, and Nicholas Geard. Risk mapping for COVID-19 outbreaks in Australia using mobility data. *Journal of The Royal Society Interface*, 18(174):20200657, 2021.