Understanding the effect of sampling effort on covid-19 case numbers

Candidate Number:

Supervisor: Dr. Manolopoulou Loanna

Department of Statistical Science

University College London

Word count:

August 7, 2021

I, Candidate Number:, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

In the end of 2019, the new coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), outbroke in China and soon spread outside to the whole world. The general underestimation of coronavirus disease 2019 (COVID-19) in early periods made the UK suffer from the large epidemics. In addition, since no effective vaccine or other pharmaceutical approaches are proposed to contain the spread of the epidemic, the government had to implement non-pharmaceutical interventions in order to prevent it from further developing. In this paper, we study the effect of major interventions across the UK for the period from the start of the COVID-19 epidemics in February 2020 until when the lockdown restrictions are gradually eased in October 2020. Specifically, combining techniques of epidemic modelling, bayesian inferring and MCMC simulating, we establish the model to estimate the transmissions based on deaths that were observed in the following weeks. By matching the timeline of government policies with the estimated transmissions in time-series format, we are able to interpret the effect of the non-pharmaceutical interventions. Our results show that major non-pharmaceutical interventions, especially lockdowns, have had a large effect on reducing transmission. Moreover, the model further shows that the transmission is potential to increase again after easing the restrictions, indicating that a series of long-term non-pharmaceutical interventions are neccessary to keep transmission of SARS-CoV-2 under control.

Acknowledgements

Acknowledge all the things!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The global impact of COVID-19 is far-reaching, and it has been considered to be the most serious of respiratory viruses since the H1N1 influenza pandemic in 1918. Evidence shows that SARS-CoV-2 will continue to spread from person to person (Riou and Althaus, 2020). It can cause serious illness (Huang et al., 2020), and the elderly are prone to bear higher risk of severe consequences and even death (Surveillances, 2020). The first two COVID-19 cases in the UK were confirmed on January 31, 2020. Although the implementation of testing, quarantine and contact tracing may slow down early transmission (Hellewell et al., 2020) this is not enough to contain the outbreak in the UK (Davies et al., 2020). Aside from medical aids, *Non-Pharmaceutical Intervention* (NPI) has played a key role in reducing the basic reproduction number R_t and the impact of COVID-19 in the United Kingdom. It will continue to be the main public health tool against SARS-CoV-2 until all people at risk of COVID-19 can obtain an effective vaccine. However, NPI can have a severe negative impact on people's overall well-being, social operations, and the economy. Therefore, these interventions should be carefully used and guided by data so as to protect the most vulnerable individuals in society.

The NPI policy response to COVID-19 can be complex. These policies can be at a country level or a local level. Different countries resort to different methods to manage the pandemic, and it also evolves over time according to the developing situation of infection and economy. NPI usually contains closings of schools and universities, closings of workplaces, canceling public events,

limits on gatherings, closing of public transport, stay at home requirements, restrictions on internal movement between cities/regions, and restrictions on international travel. The timeline of NPI policies in the UK taken from the website `https://www.instituteforgovernment.org.uk` is shown in Figure 1.1.

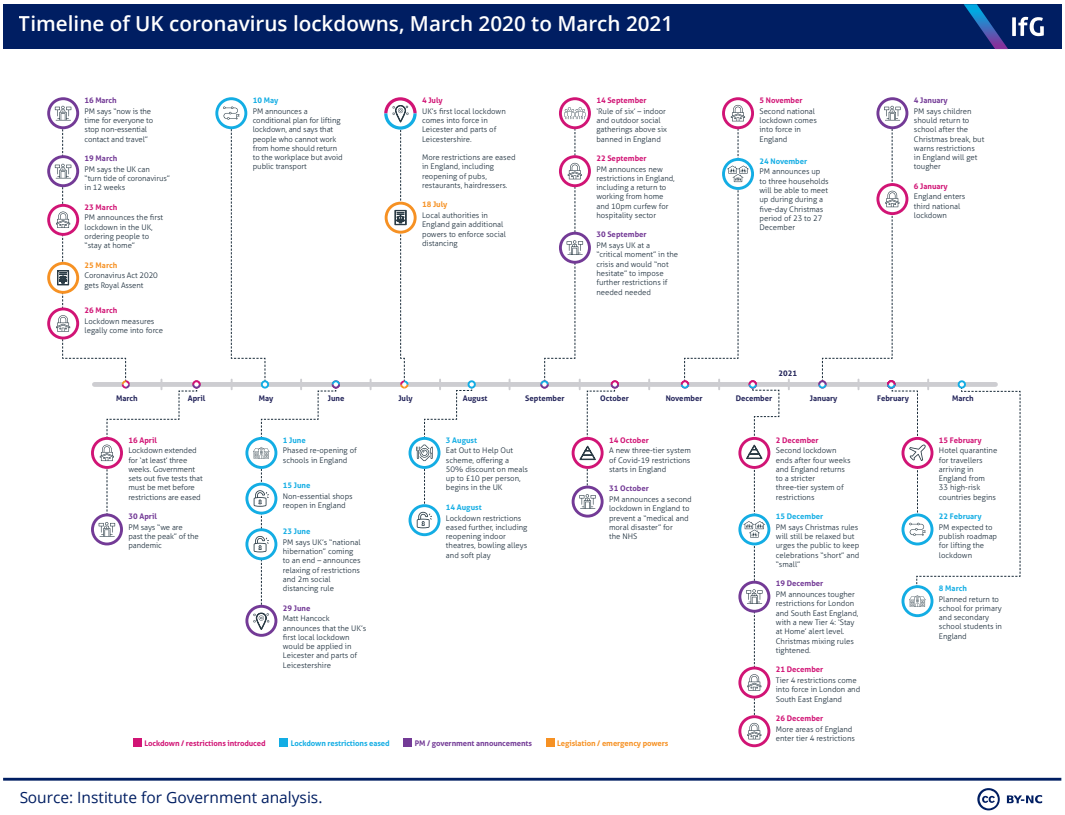

Figure 1.1: The timeline of NPI policies in the UK.

The interventions can be of different levels and extents, too. For example, according to the situation, the government can recommend closing or all schools open with alterations resulting in significant differences compared to non-Covid-19 operations, or require closing (only some levels or categories, eg. just high school, or just public schools), or even require closing all levels when the infection rate is high. For another example, in the international travel controls, the levels can be (from mild to severe) screening arrivals, quarantine arrivals from some or all regions, ban arrivals from some regions, and ban on all regions or border closer. Moreover,

The main purpose of these interventions is to reduce the reproducing number R_t, a basic epidemiological number, which represents the average number of infections produced by each infected case during the course of infection (Flaxman et al., 2020). This thesis will estimate effects of NPIs on Covid-19 in the UK. To be specific, the effect of major interventions will be studied across the UK from period from the start of the Covid-19 epidemics in Februrary 2020 until 1 October 2020, when the majority of universities reopened with an upturn of the epidemics. Before 1 October 2020, the delta strain has not yet become a pandemic in the UK, so it is reasonable to assume that SARS-CoV-2 during this period is caused by the same strain, leading to essentially invariant death rate, hospitalization rate, series interval, etc. The data used is collected from the four regions in the UK, i.e. England, Northern Ireland, Scotland, and Wales to study both the individual and shared effects on the time-varing reproduction number R_t of the effect of public isolation policies including lock down, self isolation, public events forbidden, social distance, and reopens. The goal is to assess whether there is evidence that interventions have so far been successful at reducing R_t to values below 1.

However, it is quite challenging to estimate the R_t of Covid-19 because of the high proportion of undetected infections in the health system (Jombart et al., 2020, Li et al., 2020, Zhao et al., 2020). Given the high incidence of non-specific and mild symptoms, the COVID-19 pandemic may be ignored in the new location until the first severe case or death is reported. Also, regular changes in testing policies that cause the proportion of detected infections to vary from time to time and from region to region. From the very beginning of the Covid-19 epidemics, most countries and regions can only detect a small number of suspected cases, and retain the ability to test critically ill patients or high-risk groups. Therefore, the reported data of detected infections if often biased downwards, and thus there is a need to resort to alternative ways to estimate R_t.

One of the alternative methods is to estimate the course of an epidemic is to calculate backwards from the number of deaths observed to the number of infections (Flaxman et al., 2020). I will use a Bayesian regression model to link the infection

cycle to the observed deaths and infer the total infected population (attack rate) and R_t. The key assumption here is that the distribution of the day from infection to death remains constant within a certain time range. Compared to the reported case data, reported deaths are likely to be far more reliable. Despite the limitations of the statistics of death, e.g. at the start of the Covid-19 epidemics, some cases may be attributed to other diseases, we assume that the death case confirmed by the government is the truth number in the UK. The reasons are that 1) reported deaths is deemed to be more accurate than other statistics of Covid-19 and 2) the "truth number" is hard to achieve without any reliable reported data.

The model relies on fixed estimates of some epidemiological parameters, such as the onset-to-death distribution, the infection fatality rate, and the generation distribution. In the model, I assume that only non-pharmaceutical interventions would impact the reproduction number (R_t), e.g. forbid public events, schools and universities lock-down self-isolating if ill, social distancing encouraged and city lockdown. Yet in practice, not all people will obey the lockdown rules, they want to protest on the street at their own risk, which increases the number of reproduction. Thus, a continuous variable describing the protest against the lockdown is added to the model. Furthermore, the continuous interval of covid-19 and the number of days of seed infection are assumed to be constant. This assumption originates from the built-in property of covid-19 and will not be affected by the NPIs. It is also implicitly assumed in the model that changes in R_t are an immediate response to interventions rather than gradual changes in behaviour and that individual interventions have a similar effect in different regions of the UK. The output of the model is the estimated value of R_t over time. Besides, the 95%, 60%, and 30% confidence intervals will also be reported.

The main experimental results are briefly shown as follows. Our results show that...

Chapter 2

Epidemic Model

Epidemic modelling, a task focus on learning the transmission pattern of epidemic disease, usually utilizes time-series data about populations, infected patients and deaths to describe how the disease transmits and how powerful the transmission can be. Particularly, the practical use of epidemic models must rely heavily on the experimental data. In other words, a reasonable model can only maintain major components that influence disease propagation shown in the data, and has no access to perfect interpretation to the underlying epidemic pattern.

As for the history of epidemic modelling, it is surprising that although the human kind have long been influenced by various infectious agents, the study of epidemic modelling did not start until the 19th century when the person-to-person contagion was beginning to be discussed. It is generally believed that the first series of mathematical epidemic models can date back to 1920s, when Kermack and McKendrick proposed their works in **?**. The articles provided epidemic modelling with a wide body of theory and applications for various infectious diseases **?** and they are still applicable in many modern epidemic situations. Basically, they assume people in consideration can only belong to 3 categories: susceptibles (S), infecteds (I) and removed(R), which leads to its alias "SIR models". Afterwards, various epidemic models have been proposed based on SIR models by changing the assumptions or the conditions, contributing to better interpretation to real-world diseases.

In our model, epidemic modelling provides the basic analyzing structure, indicating that how the data will be used to infer the parameters or predict the trans-

mission. In order to better understanding the model, we introduce some classic epidemic models in this section.

2.1 SIR models

SIR models are one of the earliest epidemic models. Its basic assumption lies on that individuals from an invariant community are initially equally susceptible, and one will never re-catch the disease after recovering from the infection. As we have mentioned, the population is divided into three distinct classes: Susceptible individuals (S); Infected individuals (I); and individuals who recover from the infections and been Removed from the infecting system(R). Schematically, the individual goes through consecutive states $S \to I \to R$.

Mathematically, let S_t, I_t, R_t be the number of susceptible, infected and removed individuals, respectively, at time t (Mentioned that R_t means removed individuals in this section). Assume that

- $S_t + I_t + R_t \equiv N$ (i.e. the population is closed);

- an individual comes into contact with any another individual at the rate α_1 per unit time;

- upon contact with an infected a susceptible individual contracts the disease with probability α_2, at which time he immediately becomes infected and infectious (no incubation period);

This defines a continuous time Markov Chain with the state $(S_t,\ I_t)$. Conditional on $S_t = S$ and $I_t = I$

$$\begin{aligned} \mathrm{P}_t(S_{t+h} = S-1,\ I_{t+h} = I+1) &= \alpha SIh + \mathrm{o}(h) \\ \mathrm{P}_t(S_{t+h} = S,\ I_{t+h} = I-1) &= \rho Ih + \mathrm{o}(h)\ , \end{aligned}$$

where $\alpha = \alpha_1 \times \alpha_2$, and

$$\begin{aligned} \mathrm{E}_t(S_{t+h} - S_t) &= -\alpha SIh + \mathrm{o}(h) \\ \mathrm{E}_t(I_{t+h} - I_t) &= \alpha SIh - \rho Ih + \mathrm{o}(h) \ . \end{aligned}$$

If we now formally take $h \to 0$ we arrive at the dynamical system

$$\begin{cases} \frac{dS_t}{dt} = -\alpha S_t I_t \\ \frac{dI_t}{dt} = \alpha S_t I_t - \rho I_t. \end{cases}$$

To in- vestigate the infection spread under this model, we only need to consider nonnegative solutions for S, I, and R. The epidemic stops when $I_t = 0$ for the first time. Suppose $I_0 > 0, S_0 > 0$, and $R_0 = 0$, the key question is, given parameters α, ρ and the initial number of infecteds and susceptibles, whether the infection spreads and how it develops with time. Notice that S_t decreases with t, and

$$\frac{dI_t}{dt} = I_t(\alpha S_t - \rho) \begin{cases} \leq I_t(S_0 - \rho) \leq 0 \text{ for all } t > 0, & \text{if } S_0 \leq \rho/\alpha \\ > 0 \text{ for some } t > 0, & \text{if } S_0 > \rho/\alpha. \end{cases}$$

In the case when $S_0 \leq \rho/\alpha$, the number of infecteds monotonically decreases with time, that is no epidemic can occurs. By an epidemic we mean the situation when $I_t > I_0$ for some $t > 0$. On the other hand, when $S_0 > \rho/\alpha, dI_t/dt > 0$ at least initially, and the number of infecteds increases in the beginning. We observe the *threshold phenomena* at $S_0 = \rho/\alpha$, or qualitatively different infection spread above and below this level. Meanwhile, the critical parameter $R_0 \equiv \alpha S_0/\rho$ is called the *basic reproduction number*, and is defined as the number of secondary infections introduced by one primary infection into a wholly susceptible population. As a result, SIR models suggest that in many epidemic models $R = 1$ is the critical value; $R < 1$ implies no epidemic and $R > 1$ that an epidemic is possible.

2.1.1 SIS models

SIS models are similar to SIR models, but assume that recovered individuals can get infected again. Mathematically, Let S be the number of susceptible individuals, and let I be the number of infected individuals. For an SIS model, infected individuals return to the susceptible class on recovery because the disease confers no immunity against reinfection. The simplest SIS model is given by

$$\frac{dS_t}{dt} = -\beta S_t I_t + \alpha I_t,$$
$$\frac{dI_t}{dt} = \beta S_t I_t - \alpha I_t.$$

where

- The $\beta S_t I_t$ term is understood as follows: An average infected individual makes contact sufficient to infect βN others per unit time. Also, the probability that a given individual that each infected individual comes in contact with is susceptible is S_t/N. Thus, each infected individual causes $(\beta N)(S_t/N) = \beta S_t$ infections per unit time. Therefore, I_t infected individuals cause a total number of infections per unit time of $\beta S_t I_t$.
- The αI_t term is even simpler to understand: α is the fraction of infected individuals who recover (and re-enter the susceptible class) per unit time.

We see that $\frac{d}{dt}(S_t + I_t) = 0$, indicating $S_t + I_t = N =$ constant., and we can get

$$\frac{dI_t}{dt} = (\beta N - \alpha)I - \beta I^2.$$

When the system becomes stable, we solve $dI_t/dt = 0$ and find 2 possible equilibria for this SIS model, one with $I_t = 0$ and the other with $I_t = N - \alpha/\beta$. As for the basic reproductive number, we have

$$R_0 \equiv \frac{\beta N}{\alpha},$$

and it can be shown that

- $R_0 < 1 \Rightarrow$ the equilibrium with $I = 0$ is stable,
- $R_0 > 1 \Rightarrow$ the equilibrium with $I = N - \alpha/\beta$ is stable.

2.1.2 SIR models with vital dynamics

In basic SIR models and SIS models, we assume the population is a constant. Nevertheless, real-world communities always have deaths and births. Consequently, in this section, we will introduce death rate μ and birth rate Γ into the model. The dynamic equations can be written as:

$$\begin{aligned}\frac{dS_t}{dt} &= \Gamma - \mu S_t - \frac{\beta I_t S_t}{N} \\ \frac{dI_t}{dt} &= \frac{\beta I_t S_t}{N} - \gamma I_t - \mu I_t \\ \frac{dR_t}{dt} &= \gamma I_t - \mu R_t\end{aligned}$$

where the equilibrium lies at

$$(S_t, I_t, R_t) = \left(\frac{\Gamma}{\mu}, 0, 0\right)$$

with basic reproduction number equals to $R_0 = \frac{\beta}{\mu+\gamma}$. and it can be shown that:

- $R_0 < 1 \Rightarrow (S_t, I_t, R_t) \to \left(\frac{\Gamma}{\mu}, 0, 0\right)$ is stable,
- $R_0 > 1 \Rightarrow (S_t, I_t, R_t) \to \left(\frac{\gamma+\mu}{\beta}, \frac{\mu}{\beta}(R_0 - 1), \frac{\gamma}{\beta}(R_0 - 1)\right)$, representing that the disease is not totally eradicated and remains in the population.

Chapter 3

Statistical Background

3.1 Bayesian Statistics

Bayesian statistics is named after Thomas Bayes, who formulated a specific case of Bayes' theorem in a paper published in 1763. During much of the 20th century, Bayesian methods were viewed unfavorably by many statisticians due to philosophical and practical considerations. Many Bayesian methods required much computation to complete, and most methods that were widely used during the century were based on the frequentist interpretation. However, with the advent of powerful computers and new algorithms like Markov chain Monte Carlo, Bayesian methods have seen increasing use within statistics in the 21st century (Gelman et al., 1995).

The core of Bayesian statistical methods is to use Bayes' theorem to compute and update probabilities after obtaining new data. Bayes' theorem introduced by Thomas Bayes is given by

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)}, \tag{3.1}$$

where θ is the vector of parameters, X is the data fitted to the model, and $p(X) \neq 0$. In this notation, we call $p(X|\theta)$ the likelihood, $p(\theta)$ the prior probability, $p(X)$ the evidence or else the normalizing constant, and $p(\theta|X)$ the posterior probability.

Bayes' theorem describes the conditional probability of an event based on data as well as prior information or beliefs about the event or conditions related to the event. In Bayesian inference, Bayes' theorem can be used to estimate the parameters

of a probability distribution or statistical model.

It can be observed that in Bayesian inference, two sources of information about unknown parameters of interest are being synthesized. First, given specific observation data, the likelihood function defines the possible values of the model parameter. The second is the distribution of prior beliefs, which is used to express the confidence of model parameters based on past experience. Then the product of the two is scaled and integrated into one within a reasonable range of parameter values. The result is the posterior confidence distribution of a given data parameter, which represents the understanding of the parameter based on the prior information and sample data.

Generally speaking, Bayesian statistics can solve more complex problems and provide more intuitive and meaningful inferences. In addition, through Bayesian methods, we can make direct probabilistic statements about the parameters of interest.

3.2 Bayesian Linear Regression

In statistics, Bayesian linear regression is a linear regression method in which statistical analysis is performed in the context of Bayesian inference. When the error of the regression model follows a normal distribution, and a specific form of the prior distribution is assumed, the posterior probability distribution of the model parameters can get an explicit result.

The Bayesian linear regression model starts with the same model as the frequentist linear regression, i.e. given a predictor vector x_i, the response variable y_i is

$$y_i = \alpha + \beta x_i + \varepsilon_i, \text{ for } i = 1, \ldots, n, \tag{3.2}$$

where errors ε_i are assumed to be independent and identically drawn from the a normal distribution with zero mean and constant variance σ^2, denoted as $\varepsilon_i \sim N(0, \sigma^2)$. Note that this assumption is the same as that in the frequentist linear regression for testing and constructing confidence intervals for the parameters α and β.

Under this assumption of ε_i, the random variable of each response Y_i conditioning on the observed data x_i and the parameters α, β, and σ^2, turns out to be normally distributed, i.e.

$$Y_i \,|\, x_i, \alpha, \beta, \sigma^2 \sim N(\alpha + \beta_i x_i, \sigma^2), \text{ for } i = 1, \ldots, n. \tag{3.3}$$

Thus, the likelihood of Y_i is given by

$$p(y_i \,|\, x_i, \alpha, \beta.\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(y_i - (\alpha + \beta x_i))^2}{2\sigma^2} \right). \tag{3.4}$$

By independence, the likelihood of $Y_1, \ldots, Y_n$ is given by the the product of each $p(y_i \,|\, x_i, \alpha, \beta.\sigma^2)$.

When considering the reference prior, the posterior distributions of α, β, and σ^2 is analogue to the frequentist results. Assume that the joint prior distribtuion of α, β, and σ^2 to be proportional to the inverse of σ^2, i.e.

$$p(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2}. \tag{3.5}$$

Using the hierachical model framework, this is equivalent to assuming

$$p(\alpha, \beta \,|\, \sigma^2) \propto 1 \text{ and } p(\sigma^2) \propto \frac{1}{\sigma^2}. \tag{3.6}$$

Then the marginal posterior distribution of β is the Student's t-distribution

$$\beta \,|\, y_1, \ldots, y_n \sim t(n-2, \hat{\beta}, \frac{\hat{\sigma}^2}{S_{xx}}) = t(n-2, \hat{\beta}, (se_\beta)^2), \tag{3.7}$$

with degrees of freedon $n-2$ centered at $\hat{\beta}$, with the scale parameter $\frac{\hat{\sigma}^2}{S_{xx}}$, which is the square of the standard error of $\hat{\beta}$ under the frequentist OLS model.

Similarly, α also follows the Student's t-distribution

$$\alpha \,|\, y_1, \ldots, y_n \sim t\left(n-2, \hat{\alpha}, \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) \right) = t(n-2, \hat{\alpha}, (se_\alpha)^2). \tag{3.8}$$

The following results will be used to calculate the confidence interval of the response variable Y.

The mean of the response variable Y, μ_Y, at a point x_i is

$$\mu_Y \,|\, x_i = E[Y|x_i] = \alpha + \beta x_i. \tag{3.9}$$

Under the reference prior, μ_Y has a posterior distribution

$$S^2_{Y|x_i} = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right). \tag{3.10}$$

Then any new prediction y_{n+1} at a point x_{n+1} also follows the Student's t-distribution

$$y_{n+1} \,|\, data, x_{n+1} \sim t(n-2, \hat{\alpha} + \hat{\beta} x_{n+1}, S^2_{Y|X_{n+1}}), \tag{3.11}$$

where

$$S^2_{Y|X_{n+1}} = \hat{\sigma}^2 + \hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}} \right) = \hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{S_{xx}} \right). \tag{3.12}$$

Chapter 4

Methodology

In this section, we demonstrate how we use the number of deaths to estimate the transmission of COVID-19. Under the distributional assumptions on variables, the model can be divided into three parts: (i) using epidemic model to interpret the relationship among *deaths, transmissions,* R_t; (ii) putting the training data into the epidemic model and employing bayesian regression to estimate the parameters; (3) making the predictions and utilizing MCMC method to generate the confidence intervals.

The Fig. 4.1 summarizes the overall procedures. In the beginning, the probability of a healthy person to be infected is determined by the reproduction number R_t and an empirical distribution of the infection time. Besides, the reproduction number R_t is considered to be determined ~~by some existing researches~~ and influenced by the non-pharmaceutical policies. Meanwhile, when the infection time interval keeps the same in a single day sampling from the empirical distribution, it is worthy highlighting that the number of infected individuals will also boost the transmission. After being infected, a person will show up in either the data of mechanical ventilation bed or the data of death, both of which follows certain constant distributions.

The structure of this section is as follows. We will firstly illustrate the notations in Table **??**. In Sec 4.1, we introduce the epidemic structure between the deaths (or beds) data and the infections. Then, we use bayesian regression to estimate R_t and other important posterior paramters in Sec 4.2. Last in Sec 4.3 we use MCMC

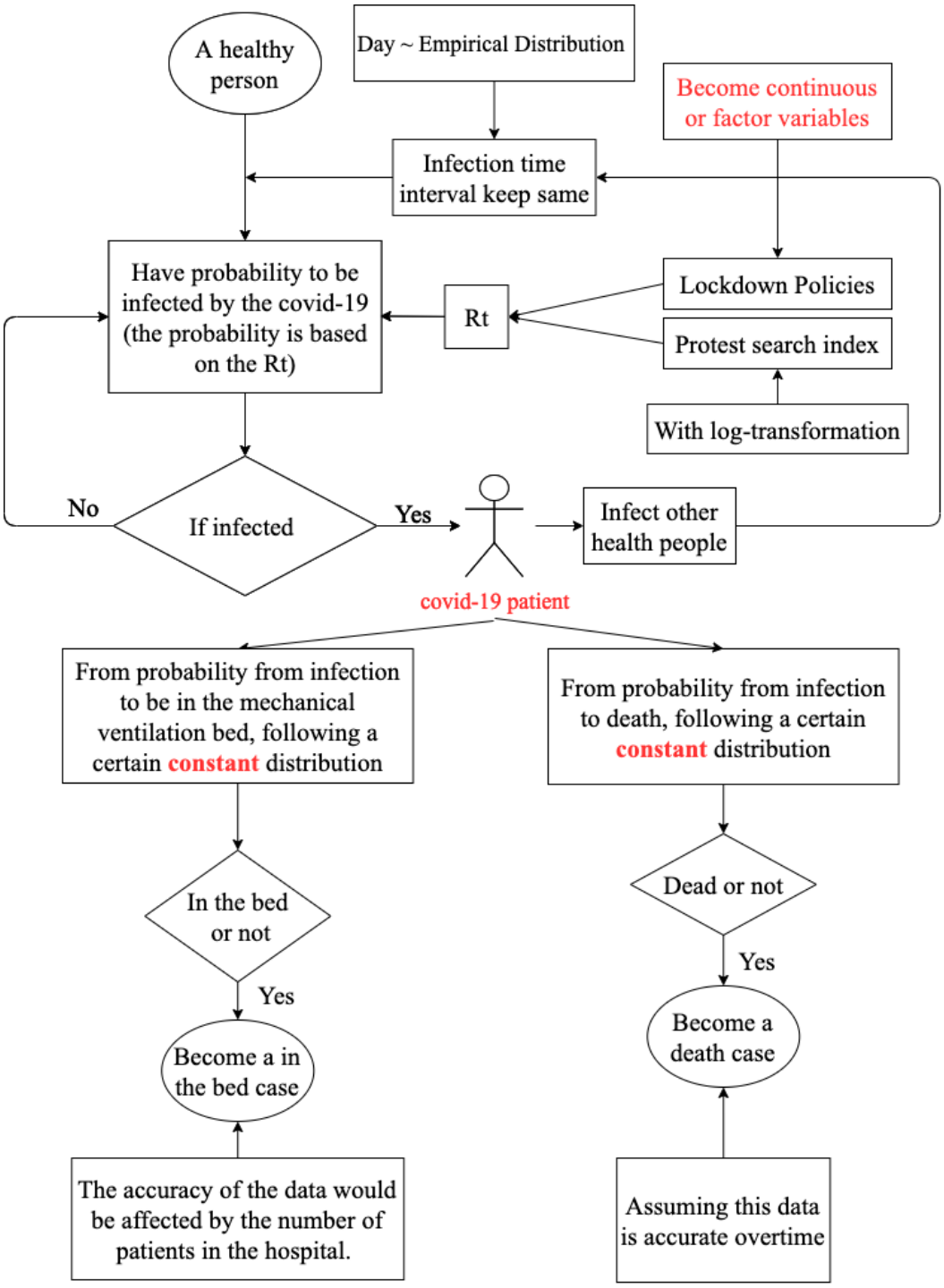

Figure 4.1: The model flowchart.

Notations	Meanings
t	Time stamp in days
y_t	expectation of response variable in day t
Y_t	response variables in day t (deaths or beds)
i_t	new infections in day t
R_t	reproduction number in day t
$p(\cdot)$	statistical distributions
π	distribution of time between infection to observation
ϕ, α, τ	parameters

method to sample from the given distributions to generate the confidence intervals for our predictions.

Before we mathematically illustrate our model, we first show the notations in this section. We further highlight that the capital letters without subscripts are vector form, i.e., $X = (X_1, \ldots, X_t)$, and an interval of time can be expressed with a colon ":", i.e. $0 : t = 0, 1, \ldots, t$ and $t : 0 = t, t-1 \ldots, 0$.

4.1 Basic epidemic model

4.1.1 Model deaths from infections

According to the variants of SIR model and the structure of bayesian analysis, we now explore the relationship between "death" and "infection". Let $Y = (Y_1, \ldots, Y_n)$ denote the observed non-negative vector of death data in n days. From a perspective of statistical modelling, we can use infections in the past few days $i_s, s < t$ to model Y_t. Nevertheless, since Y_t is regarded as an observation of a distribution, the relationship between infection and death must be bridged on the expectations. As a result, the model can be expressed as:

$$Y_t \sim p(y_t, \phi) \tag{4.1}$$

$$y_t = \alpha_t \sum_{s<t} i_s \pi_{t-s} \tag{4.2}$$

where $y_t = E(Y_t)$, $p(y_t, \phi)$ is the underlying distribution to generate Y_t, ϕ is the structural parameter of the distribution, α_t is the proportion of events at time

$t$that are recorded in the data, and π denotes the time distribution from infection to observation, indicating the hysteresis.

The equation can be explained as follows. The observation Y_t is a sample of the death distribution determined by the infections and some structural parameters; for each day, the infections have the day-like probabilities π_t to die and at rate α_t to be recorded.

4.1.2 Self-development of infections

According to the epidemic modelling, new infections i_t can be modeled through a renewal equation that controlled by both the reproduction number R_t and a degradation parameter g. Formally we have

$$i_t = R_t \sum_{s<t} i_s g_{t-s} \tag{4.3}$$

4.2 Parameter estimation

Recall that the observed data is $Y = (Y_1, \ldots, Y_n)$. In other words, we have no access to the previous information, including the initialization of the epidemic. In order to model the recursion of the infections, we set the unknown information as parameters and let all parameters are assigned priors. We have

$$i_{v:0}, R, \phi, \alpha \sim p(\cdot)$$

where $i_{v:0}$ can be used to initialize the renewal equation of the infections i_t, $R = (R_1, \ldots, R_n)$ denoting the reproduction numbers, $\alpha = (\alpha_1, \ldots, \alpha_n)$ denoting the vector of parameters.

Then, according to bayesian statistics, the posterior distribution of the parameters can be expressed as

$$p(i_{v:0}, R, \phi, \alpha | Y) \propto p(i_{v:0}) p(R) p(\phi) p(\alpha) \Pi p(Y_t | y_t, \phi)$$

4.3 Confidence interval

Chapter 5

Experiments

5.1 Prior Reproduction Number

The lock-down of city and schools would significantly reduce the reproduction number of the coronavirus, while the protest of the citizens would eliminate the efforts of lock-down. Hence, the shape of the prior reproduction number has been determined by the variation of these two conditions. To determine the number (location) of prior reproduction, the maximum reproduction number should be figured out at the early stage of epidemics(3.38, 95% confidence interval, 2.81 to 3.82) (Alimohamadi et al., 2020).

Recall that in figure 1.1, there are lots of lock-down and reopen policies for UK. An straightforward way to describe these policies is to use dummy variables. However, if all policies are add to the model as dummy variables, the Pareto k diagnostic value would come to the infinity, which means this model cannot fit well. Hence, I combine all policy variables to a single continuous variable, meaning the effort of lock-down in UK. The protest search index is taken with a logarithm transformation. By combining the lockdown policies and protest search index, I achieve a series of prior reproduction number R_t. Moreover, as the passage of time, the Real-time Assessment of Community Transmission (REACT) group has tested the monthly varied reproduction number. These results are also used to determine the parameters of prior reproduction number. To be specific, the round 1 REACT group result of it is $0.57\ (0.45, 0.72)$ between 1st May 2020 and 1 June 2020, which

is identical with the simulation result(Riley et al., 2020a). The round 2 result is 0.89 $(0.86, 0.93)$, between 1st May 2020 and early July 2020(Riley et al., 2020b). The round 3 result is 1.3 $(1.2, 1.4)$ between 24 July 2020 and 7 Sept 2020, and round 4 result is 1.7 $(1.4, 2.0)$ (Riley et al., 2021). The round 5 result is 1.06 $(0.74, 1.46)$ between 18 and 26 September 2020 (Riley et al., 2020c). Figure 5.1 shows the prior reproduction number (the blue area is the result of REACT group).

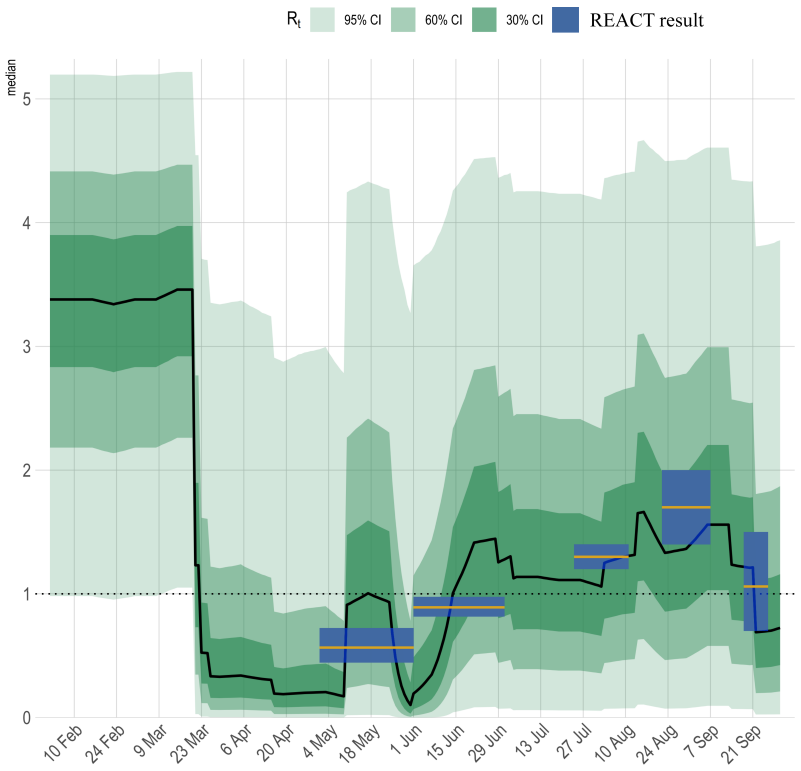

Figure 5.1: Prior reproduction number until Oct 1st, 2020

For the experimental implementations, I assume the following two assumptions hold.

1. The serial interval of covid-19 keeps constant with the mean of 5 days (Alimohamadi et al., 2020, Flaxman et al., 2020).

2. During the simulation period, delta strain did not occur.

The first assumption follows from previous works of estimating the serial interval for Covid-19.The second assumption ensures that the covid-19 is caused by the same strain, so the distributions of death rate, hospitalization rate, and other factors

remain unchanged overtime. This assumption guarantees the reproduction number R_t to be predictable.

Based on these assumptions, if the time series data of reproduction number R_t is known, the daily increasing cases could be inferred. In the following two sections, R_t is estimated through death data and through the number of people in. mechanical ventilation beds, respectively.

5.2 Simulation through Death

In this section, the R_t is estimated through death data. We show the three important assumptions as follows.

1. The daily death number is shown to be accurate. Figure 5.2 shows the fitness of observation from death data from February to October, 2020. The black line represents the death estimations and the blue areas are the 30%, 60%, and 95% intervals. The brown bars represent the observed death. From Figure 5.2, it could be found that this model fits the observed death data well, nearly all observed death data are in the 95% confident interval of the fitted model. Therefore, it is reasonable to consider the observed death data to be accurate overtime.

2. The death probability of an infected patient is assumed to be constant, and the mean value is 0.66%. Please refer to Mahase (2020) for more details.

3. The distribution of the duration from infection to death is assumed to be fixed and follows from a certain distribution. According to Flaxman et al. (2020), the modelled deaths are informed by the infection-to-death distribution.

Combining the assumptions 2 and 3 and prior reproduction number, the posterior reproduction number could be inferred as figure 5.3 and the variance of the reproduction number is significantly reduced.

Based on the simulation reproduction number, the daily increasing cases could be inferred. The result is shown in Figure 5.4.

From the figure 5.4, it could be found that the peak of simulated daily increasing cases is nearly 550K (450k ,980k) 95%, which is approximately 20 days ahead of the peak of the daily data (1250).

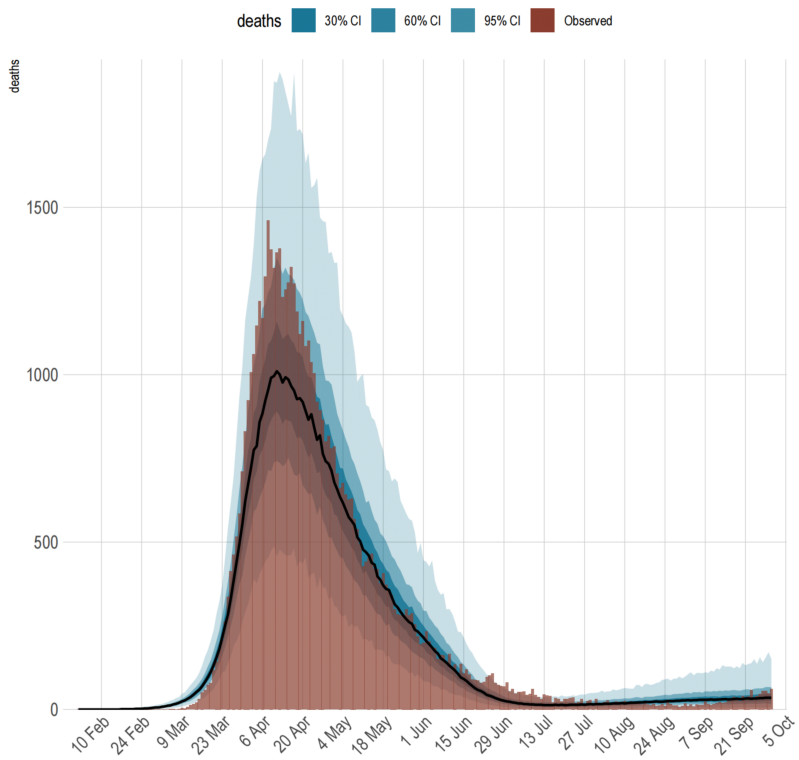

Figure 5.2: Fitness of observation from death data until Oct 1st, 2020

The cumulative Covid-19 patient tends to be $3.822\ (3.695, 3.886)$ million until 15 July 2020 in the result of REACT group (red)(Ward et al., 2021). And the similar experiment has conducted with the ONS, whose results are shown as the yellow ones. The Oxford University also has anther research in the Oxford area (purple), indicating $5.3\%\ (4\%, 6.9\%)$ people have been infected(Lumley et al., 2020). Meanwhile, in Greater Glasgow region (green), $8.57\%\ (6.095\%, 11.05\%)$ people have been infected(Thompson et al., 2020). The cumulative cases simulated from the death data could been seen from figure 5.5.

5.3 Simulation through People in Mechanical Ventilation Beds

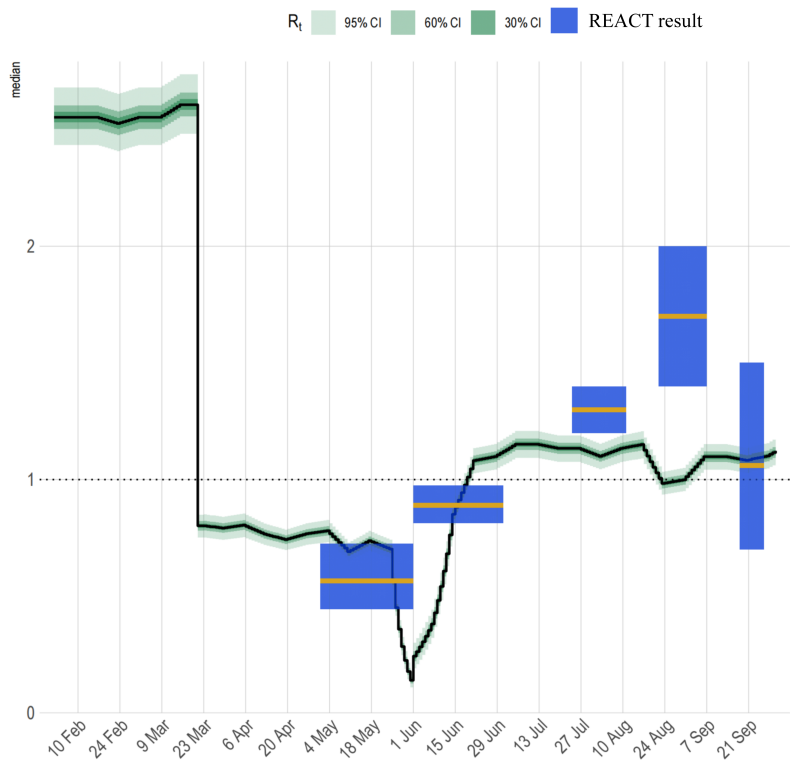

Figure 5.3: Simulation reproduction number from death data until Oct 1st, 2020

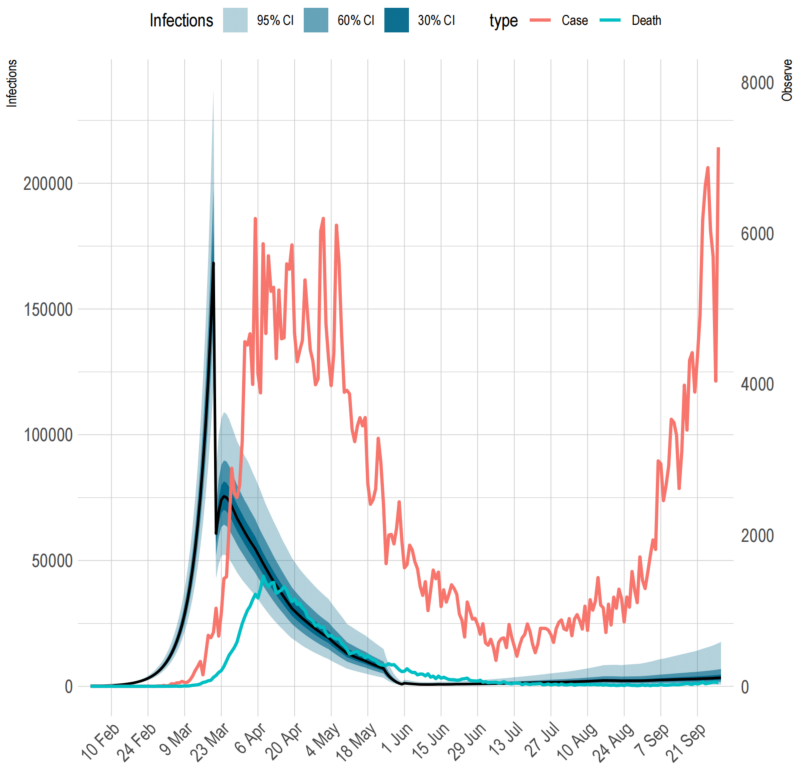

Figure 5.4: Simulation infection from death data until Oct 1st, 2020

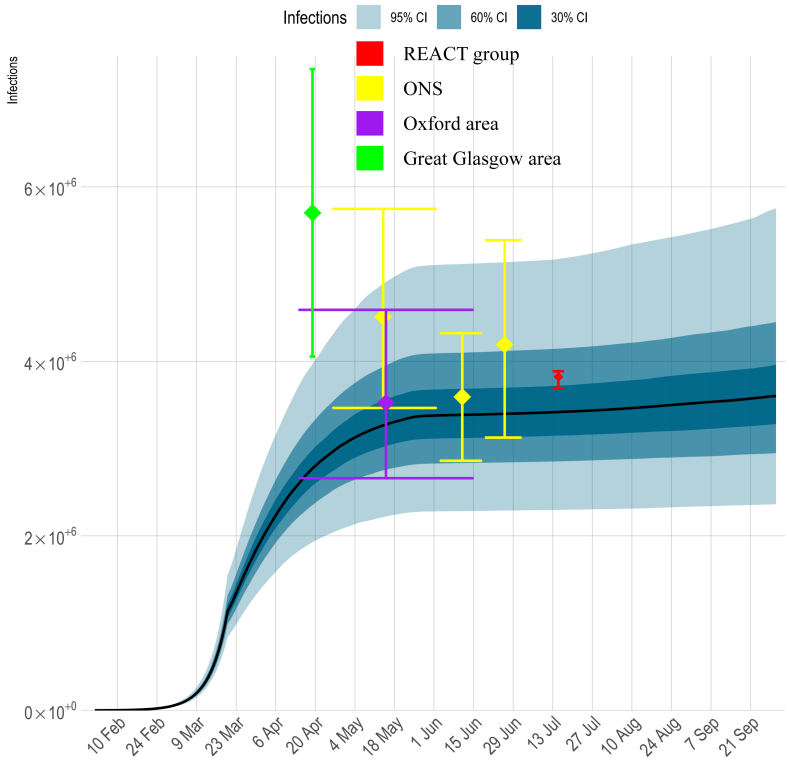

Figure 5.5: Cumulative simulated cases from death data until Oct 1st, 2020

Chapter 6

Discussions

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Appendix A

An Appendix About Stuff

(stuff)

Appendix B

Another Appendix About Things

(things)

Appendix C

Colophon

This is a description of the tools you used to make your thesis. It helps people make future documents, reminds you, and looks good.

(example) This document was set in the Times Roman typeface using LaTeX and BibTeX, composed with a text editor.

Bibliography

Yousef Alimohamadi, Maryam Taghdir, and Mojtaba Sepandi. Estimate of the basic reproduction number for covid-19: a systematic review and meta-analysis. *Journal of Preventive Medicine and Public Health*, 53(3):151, 2020.

Nicholas G Davies, Adam J Kucharski, Rosalind M Eggo, Amy Gimma, W John Edmunds, Thibaut Jombart, Kathleen O'Reilly, Akira Endo, Joel Hellewell, Emily S Nightingale, et al. Effects of non-pharmaceutical interventions on covid-19 cases, deaths, and demand for hospital services in the uk: a modelling study. *The Lancet Public Health*, 5(7):e375–e385, 2020.

Seth Flaxman, Swapnil Mishra, Axel Gandy, H Juliette T Unwin, Thomas A Mellan, Helen Coupland, Charles Whittaker, Harrison Zhu, Tresnia Berah, Jeffrey W Eaton, et al. Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature*, 584(7820):257–261, 2020.

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

Joel Hellewell, Sam Abbott, Amy Gimma, Nikos I Bosse, Christopher I Jarvis, Timothy W Russell, James D Munday, Adam J Kucharski, W John Edmunds, Fiona Sun, et al. Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 8(4):e488–e496, 2020.

Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, et al. Clinical features of

patients infected with 2019 novel coronavirus in wuhan, china. *The lancet*, 395 (10223):497–506, 2020.

Thibaut Jombart, Kevin Van Zandvoort, Timothy W Russell, Christopher I Jarvis, Amy Gimma, Sam Abbott, Sam Clifford, Sebastian Funk, Hamish Gibbs, Yang Liu, et al. Inferring the number of covid-19 cases from recently reported deaths. *Wellcome Open Research*, 5, 2020.

Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368(6490):489–493, 2020.

Sheila F Lumley, David W Eyre, Anna L McNaughton, Alison Howarth, Sarah Hoosdally, Stephanie B Hatch, James Kavanagh, Kevin K Chau, Louise O Downs, Stuart Cox, et al. Sars-cov-2 antibody prevalence, titres and neutralising activity in an antenatal cohort, united kingdom, 14 april to 15 june 2020. *Eurosurveillance*, 25(42):2001721, 2020.

Elisabeth Mahase. Covid-19: death rate is 0.66% and increases with age, study estimates. *BMJ: British Medical Journal (Online)*, 369, 2020.

Steven Riley, Kylie EC Ainslie, Oliver Eales, Benjamin Jeffrey, Caroline E Walters, Christina J Atchison, Peter J Diggle, Deborah Ashby, Christl A Donnelly, Graham Cooke, et al. Community prevalence of sars-cov-2 virus in england during may 2020: React study. *medRxiv*, 2020a.

Steven Riley, Kylie EC Ainslie, Oliver Eales, Caroline E Walters, Haowei Wang, Christina J Atchison, Peter Diggle, Deborah Ashby, Christl A Donnelly, Graham Cooke, et al. Transient dynamics of sars-cov-2 as england exited national lockdown. *medRxiv*, 2020b.

Steven Riley, Kylie EC Ainslie, Oliver Eales, Caroline E Walters, Haowei Wang, Christina J Atchison, Claudio Fronterre, Peter J Diggle, Deborah Ashby,

Christl A Donnelly, et al. High prevalence of sars-cov-2 swab positivity in england during september 2020: interim report of round 5 of react-1 study. *medRxiv*, 2020c.

Steven Riley, Kylie EC Ainslie, Oliver Eales, Caroline E Walters, Haowei Wang, Christina Atchison, Claudio Fronterre, Peter J Diggle, Deborah Ashby, Christl A Donnelly, et al. Resurgence of sars-cov-2: Detection by community viral surveillance. *Science*, 372(6545):990–995, 2021.

Julien Riou and Christian L Althaus. Pattern of early human-to-human transmission of wuhan 2019 novel coronavirus (2019-ncov), december 2019 to january 2020. *Eurosurveillance*, 25(4):2000058, 2020.

Vital Surveillances. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19)—china, 2020. *China CDC weekly*, 2(8): 113–122, 2020.

Craig P Thompson, Nicholas E Grayson, Robert S Paton, Jai S Bolton, José Lourenço, Bridget S Penman, Lian N Lee, Valerie Odon, Juthathip Mongkolsapaya, Senthil Chinnakannan, et al. Detection of neutralising antibodies to sars-cov-2 to determine population exposure in scottish blood donors between march and may 2020. *Eurosurveillance*, 25(42):2000685, 2020.

Helen Ward, Christina Atchison, Matthew Whitaker, Kylie EC Ainslie, Joshua Elliott, Lucy Okell, Rozlyn Redd, Deborah Ashby, Christl A Donnelly, Wendy Barclay, et al. Sars-cov-2 antibody prevalence in england following the first peak of the pandemic. *Nature communications*, 12(1):1–8, 2021.

Juanjuan Zhao, Quan Yuan, Haiyan Wang, Wei Liu, Xuejiao Liao, Yingying Su, Xin Wang, Jing Yuan, Tingdong Li, Jinxiu Li, et al. Antibody responses to sars-cov-2 in patients with novel coronavirus disease 2019. *Clinical infectious diseases*, 71 (16):2027–2034, 2020.