A Data Driven Change-point Epidemic Model for Assessing the Impact of Large Gathering and Subsequent Movement Control Order on COVID-19 Spread in Malaysia

Sarat C. Dass[1], Wai M. Kwok[1], Gavin J. Gibson[2], Balvinder S. Gill[3], Bala M. Sundram[3] and Sarbhan Singh[3]

[1]School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia
[2]School of Mathematical and Computer Sciences, Heriot-Watt University Edinburgh, UK
[3]Institute for Medical Research, Ministry of Health, Malaysia

Abstract

The second wave of COVID-19 in Malaysia is largely attributed to a mass gathering held in Sri Petaling between February 27, 2020 and March 1, 2020, which contributed to an exponential rise of COVID-19 cases in the country. Starting March 18, 2020, the Malaysian government introduced four consecutive phases of a Movement Control Order (MCO) to stem the spread of COVID-19. The MCO was implemented through various non-pharmaceutical interventions (NPIs). The reported number of cases reached its peak by the first week of April and then started to reduce, hence proving the effectiveness of the MCO. To gain a quantitative understanding of the effect of MCO on the dynamics of COVID-19, this paper develops a class of mathematical models to capture the disease spread before and after MCO implementation in Malaysia. A heterogeneous variant of the Susceptible-Exposed-Infected-Recovered (SEIR) model is developed with additional compartments for asymptomatic transmission. Further, a change-point is incorporated to model the before and after disease dynamics, and is inferred based on data. Related statistical analyses for inference are developed in a Bayesian framework and are able to provide quantitative assessments of (1) the impact of the Sri Petaling gathering, and (2) the extent of decreasing transmission during the MCO period. The analysis here also quantitatively demonstrates how quickly transmission rates fall under effective NPI implemention within a short time period.

Keywords: COVID-19, Epidemic Models, Bayesian Inference, Movement Control Order, Large Gathering, Lockdown

1 Introduction

Imported cases from China contributed to the first COVID-19 wave in Malaysia from January 25, 2020 to February 26, 2020 [1]. This first wave had a total of 22 cases out of which 20 were directly linked to foreign travel while the remaining two cases were local transmissions [2, 3]. The second wave of COVID-19 in Malaysia was largely attributed to a mass gathering held in Sri Petaling between February 27, 2020 and March 1, 2020, which contributed to an exponential rise of COVID-19 cases in the country. This gathering involved over 16,000 participants including a large number of foreigners from countries that later registered COVID-19 cases [4]. On March 18, 2020, the Malaysian government introduced a nationwide lockdown which was the Phase I Movement Control Order (MCO) throughout the country to stem the spread of COVID-19. Phase 1 MCO was enforced for a 2-week period starting from March 18, 2020 to March 31, 2020. During this phase, various non-pharmaceutical interventions (NPIs) were strictly enforced by means of movement restrictions, wearing of face masks, social distancing and hand hygiene practices to reduce disease transmission. The Phase I MCO was then evaluated after 2 weeks based on case trends and model forecasts by the Ministry of Health (MOH) Malaysia. As local transmission persisted, the MCO was extended a total of four times until May 12, 2020. Phase 2, 3 and 4 of the MCO covered periods starting from April 1, 2020 until May 12, 2020 (Phase 2 - April 1, 2020 to April 14, 2020, Phase 3 - April 15th to April 30, 2020, and Phase 4 - May 1, 2020 to May 12, 2020). Subsequently, from May 4, 2020, the MCO was eased into the Conditional MCO (CMCO) until June 9, 2020. However, during CMCO, there were still several identified hot-spots of COVID-19 which were placed under Enhanced MCO (EMCO) with the aforementioned strict movement control restrictions.

The implementation of MCO proved to be effective - the reported COVID-19 cases reached its peak around the first week of April and subsequently started to reduce. However, concerns remained whether a rebound in transmission would occur when the MCO was lifted and if compliance to NPIs were not followed strictly at that time. In order to gain a quantitative understanding of the effect of MCO, we developed a class of mathematical models to capture the dynamics of COVID-19 spread before and after the MCO implementation. A variant of the Susceptible-Exposed-Infected-Recovered (SEIR) model is proposed and developed in this paper which incorporates heterogeneity in the transmission dynamics, additional compartments for asymptomatic transmission and a change-point, chosen adaptively based on data, to reflect the shift in spread dynamics after the MCO implementation. The models developed in this paper are able provide a quantitative assessment of the extent of COVID-19 spread during the pre-MCO (large

gathering) and MCO periods by means of a measure of infectivity developed from them. This measure is similar to the basic reproduction number, commonly denoted by $\mathscr{R}_0$, but can be calculated for more complex epidemic models such as the ones proposed here.

Deterministic compartmental models, such as the Susceptible-Infected-Recovered (SIR) or the Susceptible-Exposed-Infectious-Recovered (SEIR) models, provide a good theoretical framework to study infectious disease spread, and have been widely used and reported in the literature. However, more complex versions of these models, and their stochastic counterparts require data-rich inputs to model all aspects of the disease dynamics. Data-rich inputs, if lacking, can be compensated using reliable and informative prior elicitation. Considering the acute nature and scale of the pandemic as well as the urgent need for a multisectorial response, comprehensive data availability of the pandemic was limited in Malaysia. For example, the open source website `outbreak.my` initially reported a transmission network for all cases; however, it was unable to cope with the scale of the pandemic when it intensified. Factoring in this data limitation, we choose to develop models that are deterministic, rather than stochastic, while ensuring that they are able to capture salient transmission dynamics satisfactorily. As mentioned earlier, we enhance the deterministic models by incorporating compartments for asymptomatic transmission and a change-point to reflect the shift in disease dynamics. We also take into account heterogeneity in the disease spread such as varying contacts among susceptibles within the closed population. The starting point of our proposed models are the class of epidemic models with power transmission dynamics which are shown to incorporate heterogeneity (see [5, 6]).

Several studies in the literature [7, 8, 9, 10, 11] have analyzed the effects of NPIs in reducing the number of COVID-19 cases. In [7], the effects of different types of NPIs on COVID-19 cases are modeled using a negative-binomial distribution whose underlying parameters incorporate country information, type of NPI implemented and change-point effects. The associated Bayesian analysis is carried out using Markov Chain Monte Carlo (MCMC) algorithms to arrive at posterior parameter estimates and credible intervals. No epidemic models are considered in this work. A generalization of the SEIR epidemic model is considered in [9] to understand the dynamics of transmission in New York, USA, under various NPI settings. However, the model is complex and requires data-rich inputs for the estimation of all unknown parameters. As a result, the authors derive baseline epidemiological parameters from published literature and not from actual observed cases in New York, and in the end conduct a simulation study based on the assumed parameter values. The study in [8] extends the work of [12] and computes a time-varying basic reproduction number as a way of gauging the effect of NPIs over time. Both these works

assume that serial intervals (i.e., the time between onset of symptoms for the infector and the infectee) can be computed for each case, which is another situation requiring data-rich input.

Studies that use compartmental epidemic models as a way of gauging the time-varying effects of NPIs have emerged over the course of the pandemic [13, 14, 15]. Compartmental epidemic models naturally model disease spread via contact rates which directly quantify the extent of NPIs (since, as mentioned earlier, NPIs are designed to reduce person-to-person transmission). Thus, epidemic models provide a natural approach for considering time-varying effects of the MCO period. Further, in this paper, the estimation of SEIR parameters is carried out based on local considerations and local data; they are not obtained from published literature based on studies conducted elsewhere where their local dynamics can be vastly different.

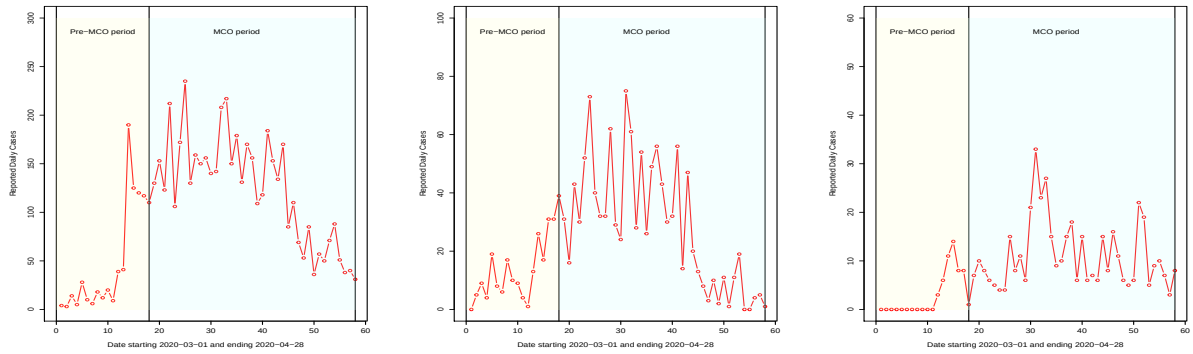
We seek to address one important aspect of Malaysia's multifaceted response to the COVID-19 pandemic, that is, to inform the health officials at MOH and aid them in their decision-making. Thus, our model was developed under local considerations using local data. Our model and related analyses are able to provide a quantitative assessment of (1) the impact of the Sri Petaling gathering, and (2) the extent of decreasing transmission during the MCO period by incorporating a time-varying contact rate parameter, which is estimated using locally available data. In essence, the proposed models here are being used as a lens to interpret the observed data in terms of when, and to what extent, a reduction in COVID-19 transmission occurred as result of the implementation of MCO.

The remainder of the paper is organized as follows: Section 2 presents the material and methods used in this paper: Section 2.1 gives the various data sources used in this study, Section 2.2 describes the epidemic models that we propose in this paper and Section 2.5 outlines the related Bayesian inference methodology developments. Section 3 gives the results of our analyses, and Section 4 provides general discussion and insights derived from these results. Finally, several conclusions and potential future work are outlined in Section 5.

2 Methods

2.1 Data Collection

Daily situation reports on COVID-19 cases in Malaysia are published by the National Crisis Preparedness and Response Centre (CPRC) of MOH, as well as other official websites (such as `outbreak.my`). The data on daily COVID-19 cases have been published



(a) (b) (c)

Figure 1: Reported daily cases for (a) Malaysia, (b) Selangor and (c) Sarawak. The time period considered is from March 1, 2020 to April 28, 2020.

since 21 January 2020 and is publicly available. The reports consist of confirmed daily and cumulative cases, recovered cases and deaths, as well as cases requiring ICU care and ventilator support. Cases by states are also available for the 13 states and 3 federal territories. In this study, we studied characteristics of the second COVID-19 wave in Malaysia starting from March 1, 2020 (corresponding to the final day of the Sri Petaling gathering). Data used for the current study are confirmed daily cases for Malaysia, and for two states: Selangor and Sarawak. These states were chosen to illustrate the aggressive transmission propagated by the Sri Petaling gathering. Selangor is the state where Sri Petaling is located and from where a majority of the participants originated, whereas Sarawak represents a state which was essentially not affected by this gathering. The time period of study is between March 1, 2020 (end of Sri Petaling gathering) and April 28, 2020, covering the period immediately after the Sri Petaling gathering and the first three phases of the MCO. Our study duration is further divided into two periods. The first period ranges from March 1, 2020 until March 18, 2020, which covers the subsequent 17 days after the gathering. The second period is taken from March 18, 2020 until April 28, 2020, covering the three successive Phases 1, 2 and 3 of the MCO. Figure 1 gives the trajectories of reported daily COVID-19 cases between March 1, 2020 and April 28, 2020 for Malaysia, and the states of Selangor and Sarawak.

All COVID-19 cases reported by MOH were confirmed by real-time reverse transcriptase-polymerase chain reaction (real-time RT-PCR) tests. A positive case was reported when the person in question was found to be positive for SARS-CoV-2 via a real-time RT-PCR test. Upon confirmation, the individual was isolated at COVID-19 designated hospitals and healthcare facilities. Active cases are defined as infected persons who were currently

undergoing treatment, and hence, isolated and removed; the individual is assumed to be unable to infect other susceptibles in the population, and hence "removed" from further modelling steps. This study did not consider transmission from positive isolated COVID-19 patients to health personnel as there was no evidence of this type of transmission occurring in the COVID-19 designated hospitals in Malaysia.

2.2 The SEIR model

The typical and well-known SEIR compartmental model consists of four compartments (Susceptible, Exposed, Infected and Recovered) representing different stages of evolution of an infectious disease, such as COVID-19, in a population. Susceptible individuals come in contact with one or more infected individuals in the population, and subsequently, become exposed to the virus. The virus then incubates within these individuals for some time. At the end of the incubation period, the exposed person becomes infectious and transmits the disease to other susceptibles in the population who come in close contact to him/her. The infected person is assumed to be infectious for a certain period (called the infectious period) after which the person recovers, dies or becomes immune. The deterministic SEIR model is given by a set of nonlinear ordinary differential equations (ODEs):

$$\dot{S}(t) = -h(S, I) \tag{1}$$

$$\dot{E}(t) = h(S, I) - \delta\, E(t) \tag{2}$$

$$\dot{I}(t) = \delta\, E(t) - \gamma\, I(t), \quad \text{and} \tag{3}$$

$$\dot{R}(t) = \gamma\, I(t) \tag{4}$$

where $S(t)$, $E(t)$, $I(t)$ and $R(t)$ represents, respectively, the susceptible, exposed, infected and recovered compartments representing the total number of individuals in each compartment at time t (here, $\dot{x}(t)$ denotes the derivative of $x(t)$ with respect to time t for $x \in \{S, E, I, R\}$), N is the total population size, and $h(S, I) = \beta \frac{S(t)}{N} I(t)$ is the rate of new infections (or, the number of new cases in the population). The parameters that govern the trajectory of the SEIR model are $\theta \equiv (\beta, \delta, \gamma, i_0, e_0)$ which are, respectively, the transmission rate (i.e., number of individuals in the population an infected person comes in contact with and successfully transmits the disease per unit time), the rate of incubation of the disease, the rate of infectiousness, the initial number of infectives and the initial number of exposed individuals. Since $\dot{S}(t) + \dot{E}(t) + \dot{I}(t) + \dot{R}(t) = 0$, it follows that $S(t) + E(t) + I(t) + R(t) = N$ for all t. Reparametrizing $S(t) = S(t)/N$, $E(t) = E(t)/N$, $I(t) = I(t)/N$ and $R(t) = R(t)/N$, the renormalized versions of S, E, I and R represent

the proportion, rather than the total number of individuals, in each compartment. In the renormalized SEIR model, $S(t) + E(t) + I(t) + R(t) = 1$ and the rate of new infections become

$$h(S, I) = \beta\, S\, I. \tag{5}$$

Based on initial values of $S(0) \equiv s_0 = 1 - \frac{i_0}{N} - \frac{e_0}{N}$, $E(0) = \frac{e_0}{N}$, $I(0) = \frac{i_0}{N}$, and $R(0) = 0$ at time $T_0 = 0$, the SEIR ODE system can be solved numerically to yield the values of $S(t), E(t), I(t)$ and $R(t)$ for every $t \in [T_0, T_1]$ where T_1 denotes the final time-point. In (1)-(4), the incubation period, $1/\delta$, is inversely proportional to the incubation rate δ, and similarly, the infectious period $1/\gamma$ is inversely proportional to the infectious rate γ. The correspondence between rates and exponential sojourn times is only approximately so since it is not so straightforward to establish this correspondence with an individual-based stochastic model given the non-linear nature of the process.

Modifications to typical SEIR formulation (1)-(4) are made to adapt it to the local Malaysian context. Here, the last compartment of the SEIR model is not "Recovered" but "Removed", representing all infectious individuals who are effectively isolated following a positive test result. In Malaysia, such patients are isolated in hospital wards to avoid further contacts with susceptibles. For COVID-19 in particular, the onset of symptoms does not necessarily indicate the start of infectiousness; in fact, the onset of infectiousness may be somewhat earlier. Thus, the infectious period $1/\gamma$ represents the period of effective infectiousness, that is, the period from the start of infectiousness (asymptomatic or symptomatic) until the individual is isolated and then can no longer infect others. Based on this understanding, $1/\delta$ represents the incubation period, which is the period starting from getting infected until the onset of infectiousness. The connection between the SEIR modelling formulation and actual evolution stages of COVID-19 in patients is shown in Figure 2. Recent studies on COVID-19 have clearly reported growing evidence of asymptomatic infections [16, 17, 18] which the current SEIR model does not incorporate. We address the issue of asymptomatic transmission in our subsequent model development in Section 2.3.

Observed data in Malaysia consists of the total number of confirmed daily cases as reported by `outbreak.my` and CPRC (i.e., the number of cases that were tested positive, and hence, effectively isolated). Hence, only the R compartment of the SEIR model can be related to observed data while other compartments of the SEIR model remain unobserved. In the subsequent model development in Section 2.3, the R compartment is further split into two: observed and cryptic sub-compartments corresponding to reported and asymptomatic infections; see Section 2.3 for more details. An additional assumption made is that individuals from the R compartment do not return to the S compartment -

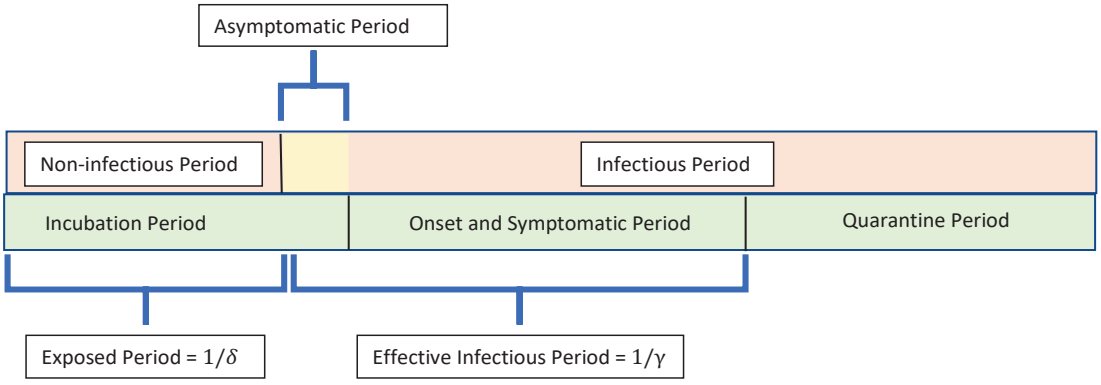


Figure 2: Salient periods in the evolution of COVID-19 and their connection to durations in the SEIR model.

at least on the timescales over which the epidemic is observed; see, for example, [19, 20]

In order to provide a quantitative assessment of the impact of the MCO, we modify (1)-(4) to incorporate an instantaneous time-varying transmission rate, $\beta \equiv \beta(t)$ (see [21], for example) which is able to quantify the extent of disease transmission at time t. The MCO can be deemed effective if the function $\beta(t)$ shows a decay reflecting an increasing effectiveness in reducing transmission among individuals over time due to implementation of the NPIs.

2.3 The modified SEIR model

The aforementioned SEIR model does not account for heterogeneity, asymptomatic transmission and change points. To this end, we propose models with power transmission dynamics that incorporate heterogeneity in the disease parameters; see, for example, [5, 6]. Further, the infectious compartment in (3) of the SEIR model is now split into two, I_o and I_c, for symptomatic (or, observed) and asymptomatic (or, cryptic) individuals, who, respectively, exhibit and do not (or, mildly) exhibit symptoms but are nevertheless infectious. Correspondingly, the R compartment is also split into two, as mentioned earlier, to accomodate quarantined and un-quarantined cases. It is assumed that the proportion of individuals transiting from E to I_o is p. The remaining exposed individuals (a proportion of $1-p$) transition into the I_c compartment and remain undetected throughout their disease experience.

From now on, we consider only renormalized state values which represent proportions, and not actual numbers, of the population. The modified SEIR model with cryptic and

observed infectiousness is given by the following system of ODEs:

$$\begin{aligned}
\dot{S}(t) &= -h(S, I_o, I_c) && (6)\\
\dot{E}(t) &= h(S, I_o, I_c) - \delta\, E(t) && (7)\\
\dot{I}_o(t) &= p\,\delta\, E(t) - \gamma_o\, I_o(t) && (8)\\
\dot{R}_o(t) &= \gamma_o\, I_o(t) && (9)\\
\dot{I}_c(t) &= (1-p)\,\delta\, E(t) - \gamma_c\, I_c(t) && (10)\\
\dot{R}_c(t) &= \gamma_c\, I_c(t) && (11)
\end{aligned}$$

where the rate of new infections is now given by

$$h(S, I_o, I_c) = \left(\alpha + \beta_o(t)\,[I_o(t)]^{w_{i,o}} + \beta_c(t)\,[I_c(t)]^{w_{i,c}}\right) \cdot [S(t)]^{w_s} \qquad (12)$$

as opposed to $\beta\, S(t)\, I(t)$ in (5) for the SEIR model. A key difference of the model formulation in (6) - (12) is the power transmission dynamics used to model heterogeneity in the population; see [5, 6]. In [6], a gamma distribution is elicited on the varying disease transmission parameters within the population which gives rise to the powers w_s, $w_{i,o}$ and $w_{i,c}$ on $S(t)$, $I_o(t)$ and $I_c(t)$, respectively, with $w_s \geq 1$, $w_{i,o} \geq 1$ and $w_{i,c} \geq 1$. The lower bounds on w_s, $w_{i,o}$ and $w_{i,c}$ recover the original SEIR model dynamics with no heterogeneity. The remaining parameters have the following interpretation: (1) α represents a small yet significant force of infection that starts the local infection process but is eventually overwhelmed by it. In our context, α represents the initial force of infection arising from, say, foreign infectious individuals attending the large gathering at Sri Petaling starting February 27th. (2) The parameters γ_o and γ_c have the same interpretation as γ in (3), that is, they are rates of infectiousness but for the observed and cryptic compartments, respectively. (3) The parameter δ is the same as before, namely, it is the rate of incubation associated with the exposed compartment. (4) The parameters β_o and β_c are transmission rates for the observed and cryptic compartments, respectively, in the modified SEIR model with $\beta_c = \mu\,\beta_0$ and $\mu \in [0, 1]$. In other words, we assume that the transmission rate for asymptomatic individuals is smaller than that of symptomatic individuals; this is a plausible assumption to make as asymptomatic individuals generally possess a lower viral load which leads to lower chances of a successful transmission. On the other hand, a longer asymptomatic infectious period may compensate for the lower transmission rates for an asymptomatic individual, and this possibility is captured by the model via the parameter γ_c.

A change-point is incorporated into the modified SEIR model to capture the shift in disease dynamics before and after the start of the MCO. For this, an unknown threshold,

T^*, is chosen so that observed daily cases fall either to the left or right of T^*. We denote the observation window to the left of T^* by $\mathcal{W}_L$ which consists of dates from March 1st up to and including T^*. The window to the right of T^* is denoted by $\mathcal{W}_R$ which consists of dates after T^* up to and including April 28th, 2020. The change-point date T^* is inferred from data; it is not taken as March 18th, 2020, the date of the start of MCO. Choosing T^* in this data-driven way is justified as the impact of MCO on observed cases may not be realized immediately. The modified SEIR model with change-point T^* is governed by the ODE system (6)-(12) for all time points $t \in \mathcal{W}_L$. For $t \in \mathcal{W}_R$, the same ODE system (6)-(12) is considered but now with a different set of parameter values in $\mathcal{W}_R$ compared to $\mathcal{W}_L$, and with a new β_o expressed as a function of t,

$$\beta_o(t) = \gamma_o \left[e^{a_0 + a_1 (t - T^*)} + c \left[1 - e^{a_1 (t - T^*)} \right] \right], \tag{13}$$

to model possible changes in disease transmission over time. The functional form of $\beta_o(t)$ in (13) has an initial value of $\gamma_o e^{a_0}$ at $t = T^*$ after which it decreases (provided $a_1 < 0$) to the asymptotic value of $c\gamma_o$ as $t \to \infty$. Thus, $c\gamma_o$ represents the residual disease transmission that may be present even during the MCO period, for example, due to close contact between family members in the same household. The general functional form of $\beta_o(t)$ subsumes the constant disease transmission rate model as a special case by taking $a_1 = c = 0$ in (13). The constant rate submodel has the advantage of not explicitly assuming any functional form for the change in disease transmission over time. On the other hand, it can only ascertain if there is an overall change (increase or decrease) in transmission after the change point T^*. Similar to the relationship $\beta_c = \mu\beta_o$ in $\mathcal{W}_L$, we assume $\beta_c(t) = \mu\,\beta_o(t)$ for the window $\mathcal{W}_R$ based on a different μ value. In Section 3, the submodel is used first followed by the full model to fit to observed data.

In the model formulation of (6)-(11), only the R_o compartment is modelled directly using a likelihood function based on daily observed cases. The other compartments of the modified SEIR model remain latent and do not have any direct observation processes for modelling based on likelihoods; see Section 2.5 for further details.

2.4 Quantitative assessment of disease spread

The basic reproduction number, $\mathscr{R}_0$, is defined as the number of secondary infections caused by one primary individual during his/her infectious period. It is the most important quantitative indicator reported to assess whether the disease is in control or not. It is well-known that the threshold value of 1 for $\mathscr{R}_0$ distinguishes between the situations

where a major epidemic occurs versus the disease dying out eventually in the population. In fact, several studies in the literature report a gradual decrease in $\mathscr{R}_0$ after lockdown [22, 23, 24]. For the SEIR model in (1) - (4), $\mathscr{R}_0$ is given by the well-known formula $\mathscr{R}_0 = \beta/\gamma$. Time-varying measures of secondary infections per primary infected, $\mathscr{R}_t$, are also available in the literature. However, for the modified SEIR model presented in Section 2.3, $\mathscr{R}_0$ and $\mathscr{R}_t$ cannot be computed. Therefore, we resort to an alternative quantitative assessement of disease spread - the total number of infections (i.e., generational) caused by the introduction of one additional infectious individual into the infection process at time point t. This procedure is illustrated using the I_o compartment. Based on (6)-(12), new values for the ODE system are calculated from time point t onwards with current state values serving as initializations of the ODE system for all except one compartment: For the I_o-compartment, the current value $I_o(t)$ is replaced by $I_o(t) + 1/N$ as the initial value. The new rate of incidence is given by $h(S^*, I_o^*, I_c^*)$ over the infectious period of the individual when the ODE system is propagated using (6)-(12) in $[t, t+1/\gamma_o]$. The increase in the rate of incidence by the introduction of this individual in the I_o compartment is given by

$$\Delta_t(I_o) = \int_{\left[t, t+\frac{1}{\gamma_o}\right]} \left[h\Big(S^*(u), I_o^*(u), I_c^*(u)\Big) - h\Big(S(u), I_o(u), I_c(u)\Big) \right] du. \tag{14}$$

Similarly, the increase in the rate of incidence by the introduction of one infectious individual into the I_c compartment at time point t can be calculated. This is denoted by $\Delta_t(I_c)$. The final increase in the incidence is the mixture

$$\Delta_t = p\,\Delta_t(I_o) + (1-p)\Delta_t(I_c) \tag{15}$$

where p is the proportion of exposed individuals who enter the I_o compartment in (8).

2.5 Bayesian Inference

Model fitting and inference is carried out in a Bayesian framework.

2.5.1 The Likelihood

The likelihood relating components of the R_o-compartment to the total number of daily cases, D_t, $t \in \mathcal{W}_L \cup \mathcal{W}_R$, is taken to be the negative binomial probability function, that is,

$$D_t \overset{\text{iid}}{\sim} NB(\cdot\,;\, \gamma_o\, I_o(t),\, \tau) \tag{16}$$

where

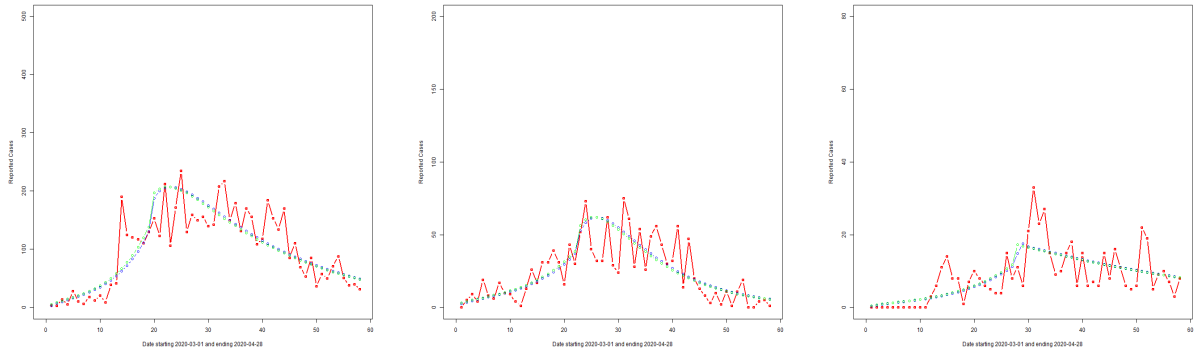
$$NB(x\,;\,\nu,\,\tau) = \frac{\Gamma(x+r)}{\Gamma(r)\,x!}\,q^r\,(1-q)^x, \qquad x = 0, 1, 2, \cdots \tag{17}$$

has mean $\nu \equiv q\,r/(1-q)$ and variance $qr/(1-q)^2 \equiv \nu + \tau\,\nu^2$; in (17), $\Gamma(u)$ is the Gamma function $\int_{s=0}^{\infty} s^{u-1}\,e^{-s}\,ds$ evaluated for $u > 0$. Thus, r need not be integer-valued in (17). The parameter τ measures overdispersion with respect to the Poisson likelihood recovered when $\tau = 1$; see, for example, [25, 26]. The observed data on daily cases in Malaysia exhibited significant overdispersion in the order of the mean. As a result, the Poisson likelihood did not fit well and we resorted to the negative binomial likelihood instead. More discussion on this aspect is presented later.

2.5.2 Assignment of Priors

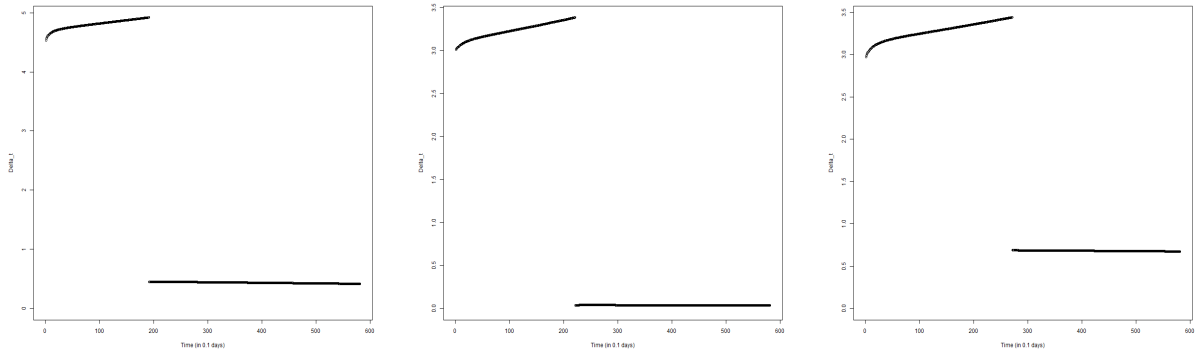
Let $\theta^{(j)} \equiv (\beta_o^{(j)}, \mu^{(j)}, \gamma_o^{(j)}, \gamma_c^{(j)}, \delta^{(j)}, p^{(j)}, w_{i,o}^{(j)}, w_{i,c}^{(j)}, w_s^{(j)}, \alpha^{(j)})$ for $j = \{L, R\}$ be the parameters of the ODE system (6)-(12) for $t \in \mathcal{W}_L$ and $t \in \mathcal{W}_R$, respectively. The collection of all parameters is denoted by $\Theta = \theta^{(L)} \cup \theta^{(R)} \cup (\tau, T^*)$ where τ is the overdispersion parameter of the negative binomial likelihood and T^* is the change-point. The uncertainty in θ is elicited via prior distributions in the Bayesian inferential framework. In what follows, we describe the priors used on the generic parameter θ after which the discussion on prior elicitation can be extended to $\theta^{(L)}$ and $\theta^{(R)}$ in a straightforward manner. It is important to note that the priors on θ depend on hyperparameters. Here, we only elicit the forms of the priors; the discussion on the exact choice of the corresponding hyperparameters is relegated to the next few paragraphs. Independent uniform priors $U(a_{\xi_1}, b_{\xi_1})$ are chosen for each $\xi_1 \in \{\mu, p, w_{i,o}, w_{i,c}, w_s, \alpha\}$ with corresponding hyperparameters a_{ξ_1} and b_{ξ_1}. The prior on (γ_o, δ) is chosen independently in the following way: For $\xi_2 \in \{\gamma_o, \delta\}$, $1/\xi_2 \sim U(a_{\xi_2}, b_{\xi_2})$. The reciprocal transformation is used for prior elicitation since these parameters have a rate interpretation, and hence, their reciprocals represent the corresponding duration (either incubation or infectious periods) with benchmark values reported in the literature [16, 17, 18, 27]. The cryptic infectious period $1/\gamma_c$ is generally longer than the observed infectious period $1/\gamma_o$ (since the former remains undetected). Hence, it follows that $1/\gamma_o < 1/\gamma_c$. This restriction can be incorporated into the prior elicitation for $1/\gamma_o$ and $1/\gamma_c$ by first generating $1/\gamma_o \sim U(a_{\gamma_o}, b_{\gamma_o})$ as above and then setting $1/\gamma_c = 1/\gamma_o + \xi_3$ where $\xi_3 \sim U(a_{\xi_3} \equiv 0, b_{\xi_3} > 0)$. The prior on the change point T^* is taken to be uniform on dates from March 18th 2020 to March 31st 2020 both inclusive. Since we define T^* as the number of days after March 1st 2020, $T^* \sim U(17, 30)$. The prior on τ is taken to be $U(a_\tau, b_\tau)$ for the entire observation window $\mathcal{W}_L \cup \mathcal{W}_R$.

The prior on β_o is elicited indirectly via a reparametrization: For $\mathcal{W}_L$, we take $\beta_o =$



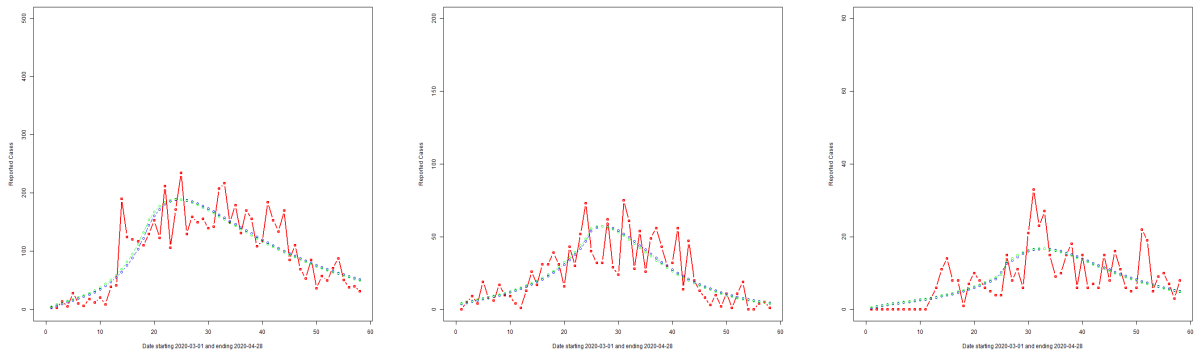
(a) (b) (c)

Figure 3: Reported daily cases (red line), and overlay plots of $\Delta R_o(t)$ (blue line) and $\gamma_o I_o(t)$ (green line) based on $\hat{\Theta}_{MAP}$ for the constant rate submodel for (a) Malaysia, (b) Selangor and (c) Sarawak.



(a) (b) (c)

Figure 4: Plots of Δ_t versus t for the constant rate submodel for (a) Malaysia, (b) Selangor and (c) Sarawak.



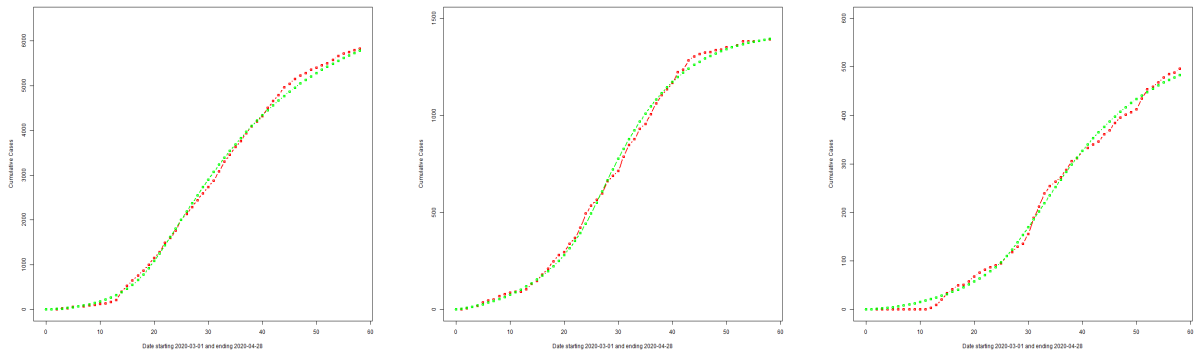
(a) (b) (c)

Figure 5: Reported daily cases (red line), and overlay plots of $\Delta R_o(t)$ (blue line) and $\gamma_o\, I_o(t)$ (green line) based on $\hat{\Theta}_{MAP}$ for (a) Malaysia, (b) Selangor and (c) Sarawak.

$e^{a_0}\,\gamma_o$ and consider a uniform prior on a_0, independent of γ_o. Our prior elicitation adopts this reparametrization for the following reason: For $\beta_o = e^{a_0}\,\gamma_o$, the basic reproduction number $\mathscr{R}_0 \equiv e^{a_0}$ is the more fundamental quantity for simpler models, such as the SIR and SEIR models, compared to β_o. Thus, $\mathscr{R}_0$ for these models can be benchmarked based on similar flu-like epidemics in the past which, in turn, provide a suggested range of values for the prior elicitation of a_0 in W_L; see [28]. Note that β_o which represents the contact rate of the target population cannot be determined by direct observation. Hence, putting a prior directly on β_o is difficult. Similarly, since $1/\gamma_o$ is the observed infectious period, benchmark values can be obtained from the literature for its prior elicitation. To summarize, we put direct priors on parameters that have an epidemiological interpretation and which can be benchmarked from reported literature for eliciting the appropriate prior.

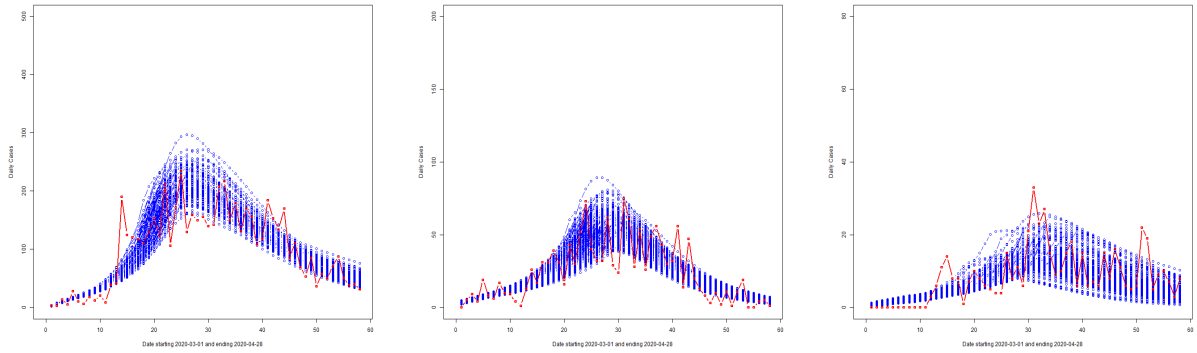
For $\mathcal{W}_R$, $\beta_o(t)$ has the form in (13) for coefficients a_0, a_1 and c. For a_0, we consider two cases: When using the submodel to determine whether an overall reduction in transmission occurs or not, we take $a_0 \sim U(L_{a_0}, U_{a_0})$ with $L_{a_0} < 0$ and $U_{a_0} > 0$. If the full model of (13) is considered, a_0 is chosen deterministically to ensure continuity of the infection process before and after the change-point. More specifically, a_0 in $\mathcal{W}_R$ is chosen so that the rate of incidence $h(S(t), I_o(t), I_c(t))$ in $\mathcal{W}_L$ and $\mathcal{W}_R$ coincide at $t = T^*$. For the full model, a_1 is given a uniform prior with support on $[L_{a_1}, U_{a_1}]$. We take $L_{a_1} < 0$ and $U_{a_1} = 0$ if such a reduction in disease transmission is established by the submodel.

Prior elicitation on parameters in W_R is based on convenience and ease of interpretation. It is easier to elicit priors on the components of $\beta_o(t)$ in (13) compared to the entire function itself. Further, the component parameters of $\beta_o(t)$ satisfy restrictions that are easy to understand; for example, $c \leq e^{a_0}$ since c and e^{a_0} are the lower and upper bounds of



(a) (b) (c)

Figure 6: Reported cumulative cases (red line), and overlay plots of $R_o(t)$ (green line) based on the MAP estimate for (a) Malaysia, (b) Selangor and (c) Sarawak.



(a) (b) (c)

Figure 7: Illustration of the variability of $\gamma_o\, I_o(t)$ from the posterior: Reported cumulative cases (red line), and overlay plots of $\gamma_o\, I_o(t)$ for (a) Malaysia, (b) Selangor and (c) Sarawak.

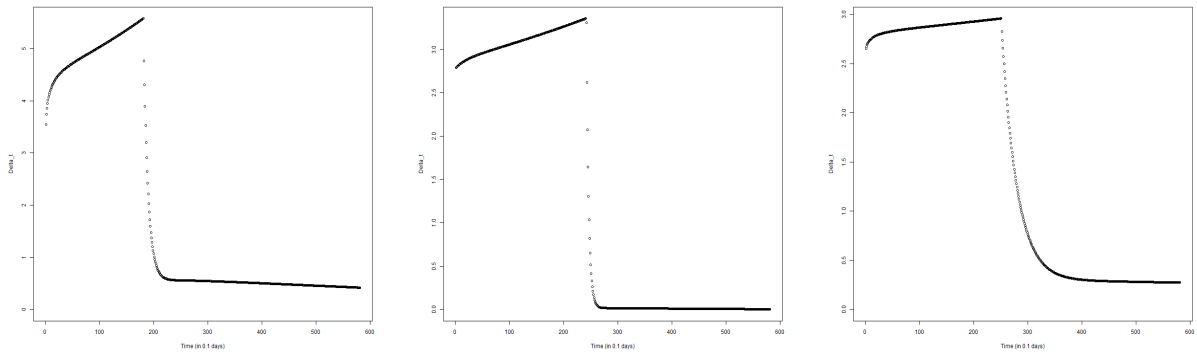
the decay curve, respectively. Hence, a reasonable prior to put on c is $U(0, e^{a_0})$. The prior elicitation on remaining unknown parameters are explained in detail in the Appendix.

2.5.3 Computational algorithm

Based on the negative binomial likelihood and prior elicitation in Sections 2.5.1 and 2.5.2, respectively, the posterior of Θ can be derived using Bayes theorem as

$$\pi(\Theta \,|\, \mathbf{D}) \;\propto\; \prod_{t \in \mathcal{W}_L} NB(D_t\,;\, \gamma_o^{(L)} I_o^{(L)}(t),\, \tau) \cdot \prod_{t \in \mathcal{W}_R} NB(D_t\,;\, \gamma_o^{(R)} I_o^{(R)}(t),\, \tau) \cdot \pi(\Theta) \quad (18)$$

$$= \;\mathbf{L}(\Theta) \cdot \pi(\Theta) \quad (19)$$



(a) (b) (c)

Figure 8: Plots of Δ_t versus t: Malaysia (a), Selangor (b) and Sarawak (c).

where $\mathbf{D} = \{D_t : t \in \mathcal{W}_L \cup \mathcal{W}_R\}$ is the collection of observed daily cases from March 1st until April 28, 2020, $\mathbf{L}(\Theta)$ is the entire likelihood for $\mathbf{D}$ and $\pi(\Theta)$ is the prior on Θ as described in Section 2.5.2. Bayesian inference is carried out using Monte Carlo importance sampling. A total of M samples, Θ_i for $i = 1, 2, \cdots, M$, are generated from the prior specification $\pi(\Theta)$. The likelihood, $\mathbf{L}(\Theta_i)$, is computed for each Θ_i and normalized to obtain weights

$$w_i = \frac{\mathbf{L}(\Theta_i)}{\sum_{i=1}^{M} \mathbf{L}(\Theta_i)} \tag{20}$$

for $i = 1, 2, \cdots, M$. To compute $\mathbf{L}(\Theta_i)$, one requires to numerically evaluate the solution of the ODE system (6)- (12). This is achieved using the `deSolve` package in R. The Bayesian computational algorithm described here is developed using R and the RStudio® user interface. Through this importance sampling step, an approximation to the Maximum-a-Posteriori (MAP) estimator of Θ is given by

$$\hat{\Theta}_{MAP} = \arg\max_{\Theta} \pi(\Theta \,|\, \mathbf{D}) \approx \arg\max_{1 \le i \le M} w_i \, \pi(\Theta_i) \tag{21}$$

with the approximation becoming more accurate as $M \to \infty$. Resampling Θ_is with weights w_i in (20) gives a ensemble of size M, $\{\Theta_i^*\}_{i=1}^{M}$, from the target posterior $\pi(\Theta \,|\, \mathbf{D})$ which can be used to provide a quantification of uncertainty around $\hat{\Theta}_{MAP}$.

3 Results

The study period is from March 1, 2020 until April 28, 2020 with $T_0 = 0$ and $T_1 = 59$. The impacts of the Sri Petaling gathering and MCO implementation are analyzed here using the proposed model described in Section 2.3. Reported daily cases at the national level

Country/State	Window	Parameter	MAP	95% Credible Interval
Malaysia	$\mathcal{W}_L$	$1/\gamma_o$	7.562	(6.629, 8.203)
		$1/\gamma_c$	8.153	(6.961, 8.992)
		$1/\delta$	7.594	(6.357, 7.935)
		a_0	2.861	(1.713, 3.395)
		μ	0.640	(0.250, 0.760)
		p	0.690	(0.538, 0.897)
	$\mathcal{W}_R$	$1/\gamma_o$	7.289	(6.094, 7.430)
		$1/\gamma_c$	7.676	(6.766, 9.230)
		$1/\delta$	8.011	(6.280, 8.224)
		a_1	-1.180	(-2.414, -0.373)
		c	6.522	(1.309, 27.797)
		μ	0.686	(0.067, 0.882)
		p	0.605	(0.479, 0.982)
	–	T^*	18	(17, 22)

Table 1: MAP estimates and associated credible intervals for Malaysia.

Country/State	Window	Parameter	MAP	95% Credible Interval
Selangor	$\mathcal{W}_L$	$1/\gamma_o$	7.952	(6.434, 8.320)
		$1/\gamma_c$	8.611	(6.826, 9.120)
		$1/\delta$	6.747	(5.615, 7.905)
		a_0	2.041	(1.296, 2.956)
		μ	0.468	(0.211, 0.774)
		p	0.785	(0.519, 0.885)
	$\mathcal{W}_R$	$1/\gamma_o$	7.342	(6.100, 7.482)
		$1/\gamma_c$	9.066	(6.839, 9.099)
		$1/\delta$	7.099	(5.518, 8.033)
		a_1	-2.384	(-2.953, -0.303)
		c	4.037	(0.144, 13.069)
		μ	0.374	(0.177, 0.881)
		p	0.708	(0.486, 0.948)
	–	T^*	22	(17.5, 25)

Table 2: MAP estimates and associated credible intervals for Selangor.

Country/State	Window	Parameter	MAP	95% Credible Interval
Sarawak	$\mathcal{W}_L$	$1/\gamma_o$	8.947	(6.647, 9.567)
		$1/\gamma_c$	11.399	(8.936, 14.138)
		$1/\delta$	7.251	(7.056, 7.972)
		a_0	1.627	(1.139, 2.618)
		μ	0.250	(0.027, 0.879)
		p	0.735	(0.502, 0.979)
	$\mathcal{W}_R$	$1/\gamma_o$	7.158	(6.195, 7.954)
		$1/\gamma_c$	11.364	(8.993, 14.278)
		$1/\delta$	7.648	(6.910, 8.287)
		a_1	-0.354	(-1.896, -0.099)
		c	0.540	(0.168, 2.705)
		μ	0.433	(0.013, 0.988)
		p	0.880	(0.521, 0.987)
	–	T^*	25	(18, 32.5)

Table 3: MAP estimates and associated credible intervals for Sarawak.

as well as for Selangor and Sarawak are used for model fitting and parameter estimation. Note that all states in Malaysia implemented MCO Phases 1-3 using the same guidelines and protocols. Thus, one can gauge the impact of the Sri Petaling gathering on COVID-19 spread in Malaysia based on a comparison between states and the national experience. Here, Selangor and Sarawak are chosen as two such representative states with high and low population densities, respectively.

First, we investigate if the MCO implementation had an overall effect of reducing COVID-19 transmission rates. For this purpose, the constant rate submodel of (13) is used and the prior on a_0 in $\mathcal{W}_R$ is chosen to be uniform with support on both positive and negative values. The Bayesian inference methodology of Section 2.5 is carried out with $M^* = 50,000$ to obtain $\hat{\Theta}_{MAP}$ and samples Θ_i^* from the posterior of Θ in (19). The curves of $\gamma_o\, I_o(t)$ and $\Delta R_o(t) \equiv R_o(t) - R_o(t-1)$ for each day t are obtained based on $\hat{\Theta}_{MAP}$ and are displayed in Figure 3. This submodel captures broad features (increasing and decreasing trends) of the reported cases trajectories in all three panels for Malaysia, Selangor and Sarawak. To quantify the overall change in transmission before and after MCO implmentation, Δ_t (see (15)) is obtained for t in $\mathcal{W}_L$ and $\mathcal{W}_R$. The plots of Δ_t versus t are shown in Figure 4. A consistent feature of the plots in all three panels is

that they first increase for time points $t \leq T^*$ followed by a significant drop for $t > T^*$. Hence, we conclude that an exponential rise in cases occurred right after the completion of the Sri Petaling gathering on March 1st 2020, and the implementation of the MCO successfully stemmed the exponential rise at the national level, Selangor and Sarawak.

With reduced disease transmission established in $\mathcal{W}_R$, we next proceed to utilize the functional form (13) of $\beta_o(t)$ as a quantitative model for transmission decay in $\mathcal{W}_R$. The Bayesian inference methodology of Section 2.5 is applied to the full model with $M^* = 50,000$ to obtain $\hat{\Theta}_{MAP}$ and samples Θ_i^* from the posterior of Θ in (19). The curves of $\gamma_o I_o(t)$ and $\Delta R_o(t) \equiv R_o(t) - R_o(t-1)$ for each day t are obtained based on $\hat{\Theta}_{MAP}$ and are displayed in Figure 5. Daily cumulative cases and the curve of $R_o(t)$ are displayed in Figure 6. We note from these figures that the proposed model captures broad features of the observed data and is an improvement over the constant rate sub-model. Uncertainty estimates are obtained for all unknown parameters in Θ based on the ensemble $\{\Theta_i^*\}_{i=1}^{M^*}$. Variabilility estimates can be obtained for all parameters and their functions. As an illustration, we demonstrate the extent of variability inherent in the posterior visually for the expected $R_o(t)$ curve given by $\gamma_o I_o(t)$ (see (16)) for $t \in [T_0, T_1]$. This is displayed in Figure 7 which shows that most of the reported case numbers are well within the limits of variability of the posterior. Hence, the proposed model together with the negative binomial likelihood are able to explain the variability in the reported case numbers. However, there are a few exceptions, the most notable being the reported case number on Day 14 for Malaysia in Figure 7(a). We present an explanation for this outlying case later in the Discussion section.

Further results from the Bayesian analyses are summarized in Tables 1, 2 and 3. These tables give the MAP estimates of parameters and their corresponding 95% credible intervals for Malaysia, Selangor and Sarawak, respectively. We provide a summary of the salient findings here. The symptomatic and asymptomatic infectious periods as well as the incubation periods are found to be around 6-8 days for Malaysia, Selangor and Sarawak. These findings are similar to values reported in the literature for other countries; see, for example, [16, 17, 18, 27]. Change-points T^* are estimated not too far away from the date of MCO implementation, March 18th 2020. For Malaysia, $T^* = 18$ is the MAP estimate which corresponds to March 19th, 2020, and the associated 95% credible interval is $(17, 21)$. For Selangor, the MAP estimate of T^* is $T^* = 22$ with $(17.5, 25)$ being the 95% credible interval. For Sarawak, the transition date is less precise. The MAP estimate is $T^* = 25$ (March 26th 2020) but the 95% credible interval $(18, 32.5)$ is much larger indicating higher uncertainty in T^*. This can be attributed to the fact that the trajectory of reported case numbers for Sarawak shows a slower and more gradual increase, then

decrease, compared to Malaysia and Selangor (see Figure 5).

The plots of Δ_t versus t for Malaysia, Selangor and Sarawak are provided in Figure 8. We note that all three panels in Figure 8 indicate a decay in the transmission rates after T^*. The measure Δ_t is calculated to be approximately 3.55 , 2.79 and 2.65 for Malaysia, Selangor and Sarawak, respectively, at the start of $\mathcal{W}_L$, that is, when $t = T_0$. Hence, Selangor achieved a rate of increase that is closer to the national level compared to Sarawak. Sarawak's cases increased but at a much slower rate compared to Selangor and Malaysia. Starting from $t = T_0$, Δ_t showed an increase in $\mathcal{W}_L$, reaching values of 5.58, 3.35 and 2.96 at $t = T^*$, respectively, for Malaysia, Selangor and Sarawak. After T^*, Δ_t for Malaysia, Selangor and Sarawak registered a decay demonstrating the effectiveness of the MCO. Δ_t declined sharply to a value around 0.55, 0.01 and 0.37, respectively, for Malaysia, Selangor and Sarawak at $t = T^* + 10$, and after that, it declined more gradually to its corresponding asymptote. Based on Δ_t, it is seen that the initial transmission rates tend to be higher for areas with a higher population density (comparing Selangor and Sarawak). On the other hand, based on the MAP estimates of a_1 in $\mathcal{W}_R$ of -2.38 and -0.35 for Selangor and Sarawak, respectively (see Tables 2 and 3), higher population density areas also experience a faster decline in the transmission rates under an effective implementation of the MCO. Although the MAP estimate of a_1 for Malaysia ($a_1 = -1.18$ from Table 1) is not as negative as it should be, we will show in the next section that a redistribution of cases further improves this estimate of a_1 and brings it closer to that of Selangor (see Section 4 for the details).

4 Discussion on Reporting Delays, Case Redistribution and Overdispersed Likelihoods

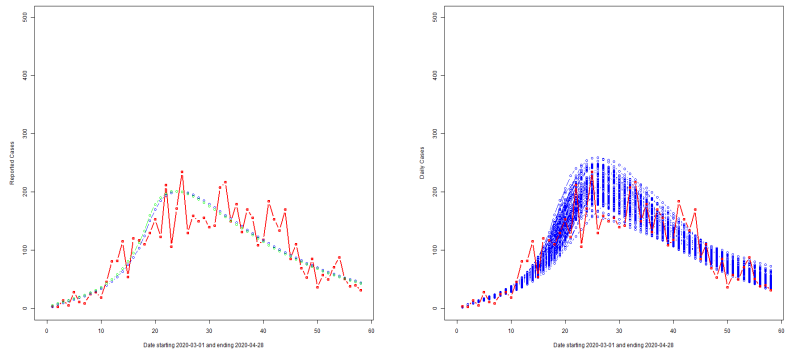
Figure 7 indicates the presence of outliers that fall outside the limits of variability of the posterior. The most notable outlier is the total number of new cases reported on Day 14 for Malaysia. Generally speaking, such outliers highlight a mismatch between the proposed model and the observed data, and point towards model inadequacy. However, we wish to emphasize that this is not the case here. One key consideration is the effect of delay, that is, whether or not the reported case numbers coincide with the day of testing. It is highly likely that a lag occurred in the reporting of cases since the COVID-19 experience was new to Malaysia. Based on the report [29], it is reasonable to assume that delays in testing and reporting were expected during the initial days of the COVID-19 outbreak in Malaysia. The peak on Day 14 seem to suggest a significant backlog of reporting of cases.

The effects of reporting delays on observed case trajectories and parameter inference are illustrated here based on a simulation study. A delay-in-reporting model based on the multinomial distribution is assumed: Let $X \sim Mult(D_t; p_1, p_2, \cdots, p_K)$, where $X = (X_1, X_2, \cdots, X_K)$ with $K = 5$ and X_k is the number of cases (out of the total reported cases on day t, D_t) that is to be redistributed to day $t-k+1$ for $k = 1, 2, \cdots, K$. The probabilities p_k, $k = 1, 2, \cdots, K$ are chosen according to a truncated geometric distribution taking values in $k-1$ for $k = 1, 2, \cdots, K$ with success probability 0.4. The cases redistribution model is applied to new cases reported from Day 10 until Day 15. The redistributed reported case trajectory, the best fit curves and associated variabilities are shown in Figure 9. Comparing Figures 7(a) and 9(b), one can immediately notice that the reported case numbers in Figure 9(b) are better explained by the variabilities of the underlying model and the negative binomial likelihood. Parameter estimates and credible intervals for the redistributed case numbers are given in Table 4. We present the salient findings here. Comparing Tables 1 and 4, we find that estimates of the infectious (both symptomatic and asymptomatic) periods have now become shorter. This is expected and reasonable since the model and likelihood do not have to account for the sudden steep rise in cases on Day 14 by preferring a larger infectious period. Nevertheless, the new infectious periods are still within the 6-8 day range and are consistent with previously reported literature. The redistribution of case numbers have also reduced the uncertainty around the MAP value of $T^* = 18$: The credible interval for T^* in Table 4 is narrower compared to that in Table 1. The MAP estimate of a_1 is now -2.075, which is closer to that of Selangor compared to Sarawak.

A final point to be discussed is our preference for the negative binomial likelihood compared to the more traditional Poisson likelihood for modelling COVID-19 case numbers. Our initial investigation used the Poisson likelihood for reported case numbers but we found that the underlying model together with the Poisson likelihood was not able to capture inherent variabilities in the observed data. Hence, we opted for the overdispersed negative binomial likelihood which was able to satisfactorily represent the observed data via its overdispersion parameter τ. This is evidenced by the variability bands presented in Figures 7 and 9(b) which successfully enclose most of the reported case numbers. This coverage is further improved in Figure 9(b) by a redistribution of delayed cases. We also provide the loglikelihood values corresponding to the Poisson and negative binomial observation models in Table 5 for Malaysia (with original case numbers), Malaysia (with redistributed case numbers), Selangor and Sarawak. Note that the negative binomial loglikelihood values are consistently larger than the Poisson counterparts indicating a better model fit to observed data.

Country/State	Window	Parameter	MAP	95% Credible Interval
Malaysia	$\mathcal{W}_L$	$1/\gamma_o$	7.121	(6.330, 8.223)
		$1/\gamma_c$	7.694	(6.829, 8.993)
		$1/\delta$	7.379	(6.211, 7.954)
		a_0	3.018	(1.643, 3.451)
		μ	0.709	(0.218, 0.778)
		p	0.810	(0.545, 0.892)
	$\mathcal{W}_R$	$1/\gamma_o$	6.821	(6.058, 7.452)
		$1/\gamma_c$	7.322	(6.652, 8.921)
		$1/\delta$	7.074	(5.959, 8.195)
		a_1	-2.075	(-2.888, -0.476)
		c	8.438	(0.675, 29.336)
		μ	0.621	(0.105, 0.807)
		p	0.756	(0.540, 0.957)
	–	T^*	18	(17, 21)

Table 4: Summary results for Malaysia (with redistributed cases).



(a) (b)

Figure 9: Effect of redistribution: Panel (a) shows the redistributed daily cases and the corresponding best fit curves of $\gamma_o\, I_o(t)$ (blue line) and $\Delta R_o(t)$ (green line) based on $\hat{\Theta}_{MAP}$. Panel (b) shows the variability of the fit based on the ensemble set $\{\Theta_i^*\}_{i=1}^M$.

Country/State	Distribution	Log-likelihood values
Malaysia	Negative Binomial	-254.68
	Poisson	-490.42
Malaysia (redistributed)	Negative Binomial	-245.66
	Poisson	-417.66
Selangor	Negative Binomial	-191.41
	Poisson	-276.06
Sarawak	Negative Binomial	-136.30
	Poisson	-162.12

Table 5: Loglikelihood values of the NB and Poisson likelihoods.

5 Conclusion

Quantitative models and assessment of the impacts of the Sri Petaling gathering and implementation of MCO on COVID-19 spread in Malaysia are developed in this paper. The MCO implementation is found to be highly effective in containing (an exponential rise of) the COVID-19 outbreak in Malaysia. The analysis here quantitatively demonstrates how quickly transmission rates fall under effective NPI implemention within a short time period. Higher disease transmission is found in Selangor (a state with higher population density) compared to Sarawak. We also found that under MCO, the decline in transmission was faster in Selangor compared to Sarawak. The rise and fall of disease transmission in Selangor mirrored the national level whereas Sarawak showed a more gradual increase and decrease in COVID-19 transmission. The change points were mostly found to be close to the date of MCO implementation (18th March 2020) although Sarawak exhibited a larger uncertainty around that date due to its gradual and slower increasing and decreasing trends of reported case numbers. Our study developed a new model to represent COVID-19 spread in Malaysia that accounts for heterogeneity and asymptomatic transmissions. We found that reported case numbers in Malaysia exhibited large variabilities which can possibly be attributed to a delay in reporting, particularly during the early stages of the pandemic as the experience with handling COVID-19 was new to the country. Nevertheless, the model developed here together with the overdispersed negative binomial likelihood are able to capture salient features of COVID-19 spread in Malaysia and provide reliable quantitative assessments even under the challenges of limited and delayed data.

Acknowledgment

The authors would like to acknowledge the Global Challenges Research Fund (GCRF) "Post MCO Strategies for Controlling Covid-19 in Malaysia" by the Scottish Funding Council and Heriot-Watt University. The authors declare no conflicts of interest. The authors would also like to thank the Director General of Health Malaysia for his permission to publish this article.

Appendix: Prior Elicitation on Remaining Parameters

We describe the prior elicitation for the initial number of infectious and exposed individuals, i_0 and e_0, respectively, relevant only for $\mathcal{W}_L$. Conditional on δ and γ_o, i_0 (corresponding to observed, not cryptic) is given the prior elicitation $\pi(i_0 \,|\, \gamma_o, \delta) = U(m_{i_0} - \Delta_{i_0}, m_{i_0} + \Delta_{i_0})$ for some m_{i_0} and Δ_{i_0}. This prior, $\pi(i_0 \,|\, \gamma_o, \delta)$, on i_0 is motivated from (9) based on the differential equation for the R_o compartment. Note that $\dot{R}_o(t) = \gamma_o I_o(t)$ from (9), and hence, $i_0 = \dot{R}_o(0)/\gamma_o$. To obtain an estimate of $\dot{R}_o(0)$, a second order polynomial is fitted using least squares to the trajectory of cumulative cases in a window of $m \geq 3$ days starting from $T_0 \equiv 0$. $\dot{R}_o(0)$ is then estimated by $\dot{P}(0)$ where $\dot{P}(t)$ is the first derivative of the fitted polynomial, $P(t)$. The mean of the uniform distribution on i_0 is taken to be $m_{i_0} = \dot{P}(0)/\gamma_o$. The initial number of exposed individuals, e_0, is given the prior elicitation $\pi(e_0 \,|\, \gamma_o, \delta, p) = U(m_{e_0} - \Delta_{e_0}, m_{e_0} + \Delta_{e_0})$ for some m_{e_0} and Δ_{e_0}. To find an expression for m_{e_0}, we rewrite (8) and note that $e_0 = (\dot{I}_o(0) + \gamma_o\, I_o(0))/(p\,\delta)$. Next, substituting $\dot{P}(0)/\gamma_o$ for $I_o(0)$, the mean of the uniform distribution on e_0 is taken as $m_{e_0} = (\ddot{P}(0) + \gamma_o\, \dot{P}(0))/(\gamma_o\, p\, \delta)$ where $\ddot{P}(t)$ is the second derivative of $P(t)$ with respect to t. The half-widths for both priors on i_0 and e_0 are taken as $\Delta_{i_0} = \Delta_{e_0} = 5$. Thus, the initial prior distributions on i_0 and e_0 are based on the number of infectives and exposed in the original population; so they are un-normalized. This is because their estimates are calculated from reported case data. But these estimates are later normalized by the population size for input into the SIER and modified SEIR models.

The hyperparameters a_ξ and b_ξ for $\xi \in \{\mu, \gamma_o, \gamma_c, \delta, p, w_{i,o}, w_{i,c}, w_s, \alpha\}$ are chosen based on values reported in previous studies where available. For example, the incubation period, defined as the period from being infected by COVID-19 to the onset of symptoms, is typically reported to be between 6 and 8 days on average [30]. Hence, we take $a_\delta = 6$ and $b_\delta = 8$ for the prior elicitation of $1/\delta$. For the infectious period, we consider $a_{\gamma_o} = 6$ and $b_{\gamma_o} = 8$ to encompass corresponding values available from the literature; see, for example, [16, 17, 18, 27]. The values of parameters reported in the literature are only

taken as starting points.

References

[1] Kementerian Kesihatan Malaysia, KPK Press Statement 25 January 2020 - Detection of A New Case Infected by The 2019 Novel Coronavirus (2019-nCoV) in Malaysia, Press Release (2020).
URL https://kpkesihatan.com/2020/01/25/kenyataan-akhbar-kpk-25-januari-2020-pengesanan-kes-baharu-yang-disahkan-dijangkiti-2019-novel-coronavirus-2019-ncov-di-malaysia/

[2] Kementerian Kesihatan Malaysia, KPK Press Statement 25 February 2020 - The Latest Situation of the Coronavirus Disease 2019 (COVID-19) Infection in Malaysia, Accessed: 2020-05-21 (2020).
URL https://kpkesihatan.com/2020/02/25/kenyataan-akhbar-kpk-25-februari-2020-situasi-terkini-jangkitan-coronavirus-disease-2019-covid-19-di-malaysia/

[3] Kementerian Kesihatan Malaysia, KPK Press Statement 27 February 2020 - the Latest Situation of the Coronavirus Disease 2019 (COVID-19) Infection in Malaysia, Accessed: 2020-02-27 (February 2020).
URL https://kpkesihatan.com/2020/02/27/kenyataan-akhbar-kpk-27-februari-2020-situasi-terkini-jangkitan-coronavirus-disease-covid-19-di-malaysia/

[4] V. Babulal, N. Z. Othman, Sri Petaling Tabligh gathering remains Msia's largest Covid-19 cluster, New Straits Times (2020).

[5] A. S. Novozhilov, On the spread of epidemics in a closed heterogeneous population, Math. Biosci. 215 (2) (2008) 177–185.
URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2580825/

[6] A. S. Novozhilov, Epidemiological Models with Parametric Heterogeneity : Deterministic Theory for Closed Populations, Mathematical Modelling of Natural Phenomena 7 (3) (2012) 147–167.
URL https://www.cambridge.org/core/journals/mathematical-modelling-of-natural-phenomena/article/epidemiological-models-with-parametric-

heterogeneity - deterministic - theory - for - closed - populations / 87C0DA28F1DC79BF69F0384C0BF5AE8B

[7] N. Banholzera, E. van Weenena, B. Kratzwalda, A. Seeligera, D. Tschernuttera, P. Bottrighia, A. Cenedesea, J. P. Sallesa, W. Vachy, S. Feuerriegely, The estimated impact of non-pharmaceutical interventions on documented cases ofCOVID-19: A cross-country analysis, medRxiv (2020).
URL https://www.medrxiv.org/content/10.1101/2020.04.16.20062141v3

[8] B. J. Cowling, S. T. Ali, T. W. Y. Ng, T. K. Tsang, J. C. M. Li, M. W. Fong, Q. Liao, M. Y. Kwan, S. L. Lee, S. S. Chiu, J. T. Wu, P. Wu, G. M. Leung, Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study, Lancet Public Health 5 (2020) 279–88.
URL https://doi.org/10.1016/S2468-2667(20)30090-6

[9] C. N. Ngonghala, E. Iboi, S. Eikenberry, M. Scotch, C. R. MacIntyre, M. H. Bonds, A. B. Gumel, Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel Coronavirus, Mathematical Biosciences 325 (2020) 1–15.
URL https://www.sciencedirect.com/science/article/pii/S0025556420300560?via%3Dihub

[10] N. Imai, K. A. Gaythorpe, S. Abbott, S. Bhatia, S. van Elslandand Kiesha Prem, Y. Liu, N. M. Ferguson, Adoption and impact of non-pharmaceutical interventions for COVID-19, Wellcome Open Research (2020).
URL https://doi.org/10.12688/wellcomeopenres.15808.1

[11] M. de Figueiredo, D. Codina, M. M. Figueiredo, S. M, C. León, Impact of lockdown on COVID-19 incidence and mortality in China: an interrupted time series study, Bull World Health Organ [Submitted] (2020).
URL http://dx.doi.org/10.2471/BLT.20.256701

[12] R. Thompsona, J. Stockwind, R. van Gaalene, J. Polonskyf, Z. Kamvarg, P. Demarshh, E. Dahlqwist, S. Lij, E. Miguelk, T. Jombartg, J. Lesslerm, S. Cauchemezn, A. Corig, Improved inference of time-varying reproduction numbers during infectious disease outbreaks, Epidemics 29 (2019).
URL https://www.sciencedirect.com/science/article/pii/S1755436519300350?via%3Dihub

[13] A. Al Wahaibi, A. Al Manji, A. Al Maani, B. Al Rawahi, K. Al Harthy, F. Alyaquobi, A. Al-Jardani, E. Petersen, S. Al Abri, Covid-19 epidemic monitoring after non-pharmaceutical interventions: The use of time-varying reproduction number in a country with a large migrant population, International Journal of Infectious Diseases 99 (2020) 466–472.
URL https://www.sciencedirect.com/science/article/pii/S1201971220306688

[14] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, Nature Medicine (2020) 1–6.
URL https://www.nature.com/articles/s41591-020-0883-7?fbclid=IwAR156AOapdnJ9R8QG4s5odlVrQgA9nhbQV7OU1KDttNw4Pq7Y860hRd54B4

[15] F.-C. Hu, The Estimated Time-Varying Reproduction Numbers during the Ongoing Pandemic of the Coronavirus Disease 2019 (COVID-19) in 12 Selected Countries outside China, medRxiv (2020).
URL https://www.medrxiv.org/content/10.1101/2020.05.10.20097154v1.full.pdf

[16] P. Yu, J. Zhu, Z. Zhang, Y. Han, A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period, The Journal of Infectious Diseases 221 (11) (2020) 1757–1761.
URL https://academic.oup.com/jid/article/221/11/1757/5739751

[17] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, V. J. Lee, Presymptomatic Transmission of SARS-CoV-2 - Singapore, January 23 - March 16, 2020, MMWR Morb Mortal Wkly Rep 2020 69 (2020) 411–415.
URL https://www.cdc.gov/mmwr/volumes/69/wr/mm6914e1.htm?s_cid=mm6914e1_w

[18] A. Kimball, K. Hatfield, M. Arons, A. James, J. Taylor, K. Spicer, A. C. Bardossy, L. P. Oakley, S. Tanwar, Z. Chisty, J. M. Bell, M. Methner, J. Harney, J. R. Jacobs, C. M. Carlson, H. P. McLaughlin, N. Stone, S. Clark, C. Brostrom-Smtih, L. C. Page, M. Kay, J. Lewis, D. Russell, B. Hiatt, J. Gant, J. S. Duchin, T. A. Clark, M. A. Honein, S. C. Reddy, J. A. Jernigan, Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility — King County, Washington, March 2020, MMWR Morb Mortal Wkly Rep 2020 69 (2020)

377–381.
URL https://www.cdc.gov/mmwr/volumes/69/wr/mm6913e1.htm

[19] J. C. Blackwood, L. M. Childs, An introduction to compartmental modeling for the budding infectious disease modeler, Letters in Biomathematics 5 (1) (2018) 195–221.
URL https://lettersinbiomath.journals.publicknowledgeproject.org/index.php/lib/article/view/81/59

[20] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London Series A containing papers of a mathematical and physical character 115 (772) (1927) 700–721.
URL https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1927.0118

[21] M. T. Anguloa, J. X. Velasco-Hernandez, Robust qualitative estimation of time-varying contact rates in uncertain epidemics, Epidemics 24 (2018) 98–104.
URL https://doi.org/10.1016/j.epidem.2018.03.001

[22] N. G. Becker, K. Glass, B. Barnes, P. Caley, D. Philp, J. McCaw, J. McVernon, J. Wood, Using Mathematical Models to Assess Responses to an Outbreak of an Emerged Viral Respiratory Disease, Final Report to the Australian Government Department of Health and Ageing. National Centre for Epidemiology and Population Health, Australian National University (April 2006).
URL https://www1.health.gov.au/internet/publications/publishing.nsf/Content/mathematical-models

[23] J. Hwang, H. Park, J. Jung, S.-H. Kim, N. Kim, Basic and effective reproduction numbers of COVID-19 cases in South Korea excluding Sincheonji cases, medRxiv (January 2020).
URL http://medrxiv.org/content/early/2020/03/31/2020.03.19.20039347.abstract

[24] W. H. Organization, et al., Report of the WHO–China Joint Mission on Coronavirus Disease 2019 (COVID-19), 16–24 February 2020 (February 2020).
URL https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

[25] G. Frasso, P. Lambert, Bayesian inference in an extended SEIR model with nonparametric disease transmission rate: an application to the ebola epidemic in sierra leone,

612 Biostatistics 17 (4) (2019) 779–792.
613 URL https://doi.org/10.1093/biostatistics/kxw027

614 [26] X. Xu, T. Kypraios, P. O'Neill, Bayesian non-parametric inference for stochastic
615 epidemic models using Gaussian Processes, Biostatistics 17 (4) (2016) 619–633.
616 URL https://doi.org/10.1093/biostatistics/kxw011

617 [27] World Health Organization, Coronavirus disease 2019 (COVID-19) Situation Report
618 73 (April 2020).
619 URL https://www.who.int/docs/default-source/coronaviruse/situation-
620 reports/20200402-sitrep-73-covid-19.pdf

621 [28] E. Petersen, M. Koopmans, U. Go, D. H. Hamer, N. Petrosillo, F. Castelli,
622 M. Storgaard, S. Al Khalili, L. Simonsen, Comparing SARS-CoV-2 with SARS-CoV
623 and influenza pandemics, The Lancet infectious diseases (2020).
624 URL https://www.thelancet.com/pdfs/journals/laninf/PIIS1473-
625 3099(20)30484-9.pdf

626 [29] Emil Zainul, Malaysia to boost virus testing with S Korean test kits, Accessed: 2020-
627 07-21 (2020).
628 URL https://www.theedgemarkets.com/article/malaysia-boost-virus-
629 testing-s-korean-test-kits

630 [30] J. A. Backer, D. Klinkenberg, J. Wallinga, Incubation period of 2019 novel coron-
631 avirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January
632 2020, Eurosurveillance 25 (5) (2020) 2000062.