

RESEARCH ARTICLE

 OPEN ACCESS

SEIRD Model for Qatar COVID-19 Outbreak: A Case Study

Ryad Ghanam^a, Edward L. Boone^b, Abdel-Salam G. Abdel-Salam^c^aDepartment of Liberal Arts and Sciences, Virginia Commonwealth University in Qatar, Education City, Doha, Qatar;^bDepartment of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, Virginia, 23284, USA; ^cDepartment of Mathematics, Statistics and Physics, College of Arts and Sciences, Qatar University, Doha, Qatar**ABSTRACT**

The COVID-19 outbreak of 2020 has required many governments to develop mathematical-statistical models of the outbreak for policy and planning purposes. This work provides a tutorial on building a compartmental model using Susceptibles, Exposed, Infected, Recovered and Deaths (SEIRD) status through time. A Bayesian Framework is utilized to perform both parameter estimation and predictions. This model uses interventions to quantify the impact of various government attempts to slow the spread of the virus. Predictions are also made to determine when the peak Active Infections will occur.

ARTICLE HISTORY

Received May 18, 2020

Accepted January 27, 2021

KEYWORDS

SEIRD, Compartmental Model, Bayesian Statistics, Intervention Analysis, Modeling Technique

1 Introduction

Coronavirus Disease 2019 (COVID-19) is a severe pandemic affecting the whole world with a fast spreading regime, requiring to perform strict precautions to keep it under control (Wu et al., 2020; Rezabakhsh et al., 2020). As there is no cure and target treatment yet, establishing those precautions become inevitable. These limitations (Giuliani et al., 2020) can be listed as social distancing, mask wearing and hand-washing, closure of businesses and schools and travel prohibitions (Chinazzi et al., 2020).

Corona Virus is a new human Betacoronavirus that uses densely glycosylated spike protein to penetrate host cells. The COVID-19 belongs to the same family classification with Nidovirales, viruses that use a nested set of mRNAs to replicate and it further falls under the subfamily of alpha, beta, gamma and delta Co-Vis. The virus that causes COVID-19 belongs to the Betacoronavirus 2B lineage and has a close relationship with SARS species. It is a novel virus since the monoclonal antibodies do not exhibit a high degree of binding to SARS-CoV-2. Replication of the viral RNA occurs when RNA polymerase binds and re-attaches to multiple locations (McIntosh, 2020; Fisher and Heyman, 2020).

Cases of COVID-19 started in December 2019 when a strange condition was reported in Wuhan, China. This virus has a global mortality rate of 3.4%, which makes it more severe in relation to flu. The elderly who have other pre-existing illnesses are succumbing more to the COVID-19. People with only mild symptoms recover within 3 to 7 days, while those with conditions such as pneumonia or severe diseases take weeks to recover. The recovery percentage of patients, for example, in China stands at 51%. The recovery percentage rate of COVID-19 is expected to hit 90% (WHO, 2020).

The virus has spread from China to 196 other countries and territories across the globe. From Wuhan, Hubei province, the virus spread to Mainland China, Thailand, Japan, South Korea, Vietnam, Singapore, Italy, Iran, and other countries. The State of Qatar was one of the countries that were affected by the COVID-19 spreading, and the first infected case was reported on 29th of February 2020 and it could be considered the 2nd highest in the Arab World with the number of confirmed cases 28,272 as of May 14, 2020.

For effectively specifying such security measures, it is essential to have a real-time monitoring system of the infection, recovery and death rates. We develop, implement and deploy a data-driven forecasting model for use by stakeholders in the State of Qatar to deal with the COVID-19 pandemic. The model will focus on infected, deaths and recovered as those are the only data available at this time.

This document is organized in the following manner. In Section 2 the Susceptible, Exposed, Infected, Recovered, Death (SEIRD) model that is employed is defined. Next, Section 3 introduces the data available and gives description. Section 4 shows how interventions are incorporated into the model. The Bayesian inference model specification is given in Section 5. Summaries

of the parameter estimation is shown in Section 6. Predictive performance is provided in Section 7. And finally, a discussion is found in Section 8.

2 The SEIRD Model

Let $S(t)$ be the number of people Susceptible at time t , $E(t)$ be the number of people Exposed at time t , $I(t)$ be the total number of Infected at time t , $R(t)$ be the cumulative number of recovered at time t and $D(t)$ be the cumulative number of Deaths at time t . This can be modeled with the following system of ordinary differential equations:

$$\begin{aligned}\frac{dS(t)}{dt} &= -\alpha S(t)E(t) \\ \frac{dE(t)}{dt} &= \alpha S(t)E(t) - \beta E(t) - \gamma E(t) \\ \frac{dI(t)}{dt} &= \beta E(t) - \gamma I(t) - \eta I(t) \\ \frac{dR(t)}{dt} &= \gamma I(t) \\ \frac{dD(t)}{dt} &= \eta I(t)\end{aligned}\tag{1}$$

where α is the transmission rate (per day \times individual²) from Susceptible to Exposed, β is the rate (per day) at which Exposed become Infected, γ is the rate (per day) at which Infected become recovered and η is the mortality rate (per day) for those Infected. Notice that, this model formulation makes several key assumptions:

1. Immigration, emigration, natural mortality and births are negligible over the time frame and hence are not in the model.
2. Once a person is in the Infected group, they are quarantined and hence they do not mix with the Susceptible population.
3. The Recovered and Deaths compartments are for those who first are infected.
4. Those who are exposed and never become infected(sick) recover at the same rate γ as those who become sick and recover.

Traditional analysis would include a steady state analysis, however, in this case the dynamics of the short term is of interest. Hence, this work does not address any steady state or equilibrium concerns. This work is concerned with fitting the model given in (1) to the COVID-19 data concerning the State of Qatar during the 2020 outbreak and using the model for forecasting several possible scenarios.

3 Data

The Johns Hopkins University (JHU) COVID-19 Github site (<https://github.com/CSSEGISandData/COVID-19>) includes for every country for each day the cumulative number of confirmed infections, cumulative number of recovered and the cumulative number of deaths for each day starting 22 January 2020. The data for Qatar was obtained. Notice that in model (1) the Recovered and Death states are cumulative as once one enters the compartment their is no exit. However, the Infected compartment has transitions from Exposed and to Recovered and Deaths. Hence the data provided for confirmed infections is cumulative and included both Recovered and Deaths and will need to be removed from this compartment's data. Let $CI(t)$ be the Confirmed Infections as reported by JHU at time t and let Infected $I(t)$ be defined as

$$I(t) = CI(t) - R(t) - D(t).$$

For clarity the term "Active Infections" will be used to denote the derived variable, $I(t)$, versus the Cumulative Infected, $CI(t)$, provided in the data.

Figure 1 shows the plots of the Active Infections, Recovered and Deaths data for Qatar for the days since 29 February 2020. Notice that the Active Infections are very low until around day 12 when there is large jump due to increased testing. The Active Infections then seems to plateau for until day 30, after which there is extreme growth in Active Infections. There seems to be a similar pattern for the Recovered with a delay showing the time of infection before recovery. The plot for Deaths shows no deaths until day 30 and then a steady increase in Deaths for the remaining days.

The State of Qatar, prepared an excellent flexible plan for risk management, grounded on national risk assessment, taking account of the global risk assessment done by WHO, focuses on reinforce capacities to reduce or eliminate the health risks from

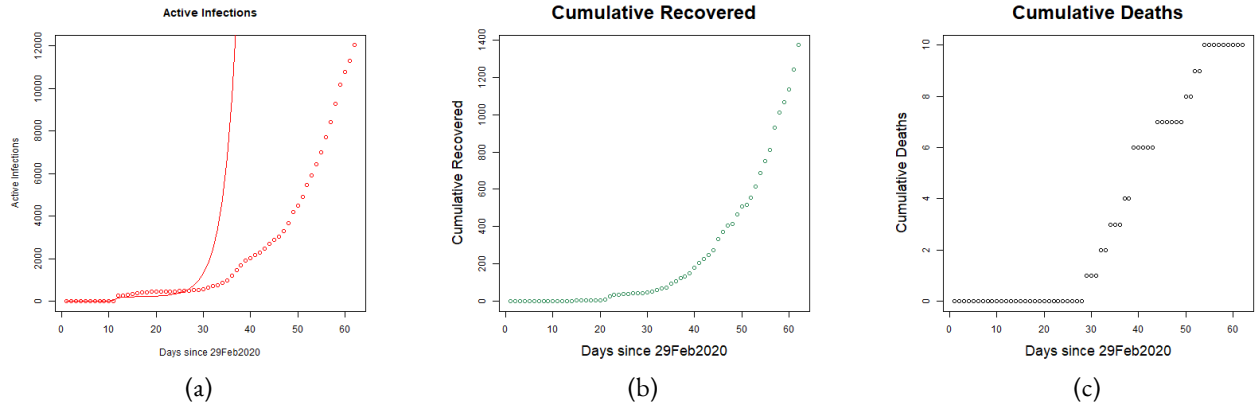


Figure 1: Plots of Number of Active Infections (a), Cumulative Number of Recovered (b) and Cumulative Number of Deaths (c) for the State of Qatar for the days since 29 February 2020 until 1 May 2020.

COVID-19. Embed complete emergency risk management strategy in the health sector. Furthermore, Enabling and promoting multi-sectoral linkage and integration across the whole-of-government and the whole-of-society (DOH-UK, 2020; DOH-Australia, 2020; WHO, 2020).

On March 9, 2020 (day 10), Qatar announced a closure of all universities and schools. It placed a travel ban on 15 countries: Bangladesh, China, Egypt, India, Iran, Iraq, Italy, Lebanon, Nepal, Pakistan, the Philippines, South Korea, Sri Lanka, Syria, and Thailand. On March 14, 2020 (day 15), Qatar expanded its travel ban to include three new countries: Germany, Spain and France (Hamad Medical Corporation, 2020; MPH-Qatar, 2019). The Ministry of Municipal and Environment on March 19, 2020 (day 19), closed all parks and public beaches and put public sector employees at 80% of work from home to curb the spread of coronavirus. The Council of Ministers on April 1, 2020 (day 31) further required that private sector employees must conduct 80% of their work from home. On April 8, 2020 (day 39), the Ministry of Commerce and Industry decided to temporarily close all restaurants, cafes, food outlets, and food trucks at the main public era. On April 23, 2020 (day 54) the holy month of Ramadan began where people were not allowed the typical social mingling the month typically provides. In fact the government highly encouraged extremely limited social mingling. A detailed list of government interventions are provided by Hamad Medical Corporation (2020) and MPH-Qatar (2019).

These interventions taken by the government change the dynamics of the system and hence need to be incorporated into the model. The next section details how we introduce interventions both from the government and interventions guided by the data. While the model interventions correspond to government policy changes, one cannot assume that any impact is solely due to those policies as other lurking variables may contribute to the impact.

4 Interventions

In Figure 1, one can see the jump at day 12 and a plateau until day 30. The model needs to be able to handle interventions made by the Government of the State of Qatar. The main parameter that policy can influence is α , the rate of transmission from Susceptible to Exposed. One way to implement this the use of indicator functions $W_k(t)$ defined as

$$W_k(t) = \begin{cases} 1 & \text{if } t > t_k \\ 0 & \text{otherwise} \end{cases}$$

where t_k is the time where the k^{th} intervention is taken and index $k = 1, 2, \dots, K$. For each intervention there needs to be a change to the value of α , denoted α_k , that captures the impact of the intervention. Let the vector $W(t) = (1, W_1(t), W_2(t), \dots, W_K(t))^T$ be the vector of the values of each $W_k(t)$ at time t . Let $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_K)^T$. This formulation gives the following transitions rates between $S(t)$ and $E(t)$:

$$\alpha(t) = \begin{cases} \alpha_0 & \text{if } 0 < t < t_1 \\ \alpha_0 + \alpha_1 & \text{if } t_1 < t < t_2 \\ \alpha_0 + \alpha_1 + \alpha_2 & \text{if } t_2 < t < t_3 \\ \vdots & \vdots \\ \alpha_0 + \alpha_1 + \dots + \alpha_K & \text{if } t_K < t \end{cases}$$

which will require the following constraints due to the fact that $\alpha(t) > 0$ for all t :

$$\begin{aligned}\alpha_0 &> 0 \\ \alpha_0 + \alpha_1 &> 0 \\ \alpha_0 + \alpha_1 + \alpha_2 &> 0 \\ &\vdots \\ \alpha_0 + \alpha_1 + \dots + \alpha_K &> 0.\end{aligned}$$

Let \mathcal{A} be the set defined by the constraints above.

In addition to changes in infection rates α , impulse functions can be used to model dramatic one time shifts in transitions between states. A Dirac delta function defined by

$$\delta(x) = 0, \text{ if } x \neq 0$$

which satisfies $\int_{-\infty}^{\infty} \delta(x) dx = 1$ (Dirac, 1958). This can be integrated in the model to capture spikes in the number of cases. In our case the State of Qatar data shows exhibits this type of behavior at day 12 where one can clearly see a large jump in the number of infections. This is incorporated into the model presented by a Dirac delta function, $\delta(t - \tau)$, in transition rate between Exposed and Infected, which is coupled with a coefficient to β_A to capture the impact of the jump.

5 Bayesian Analysis

Due to the complexity of the model the Bayesian inferential framework is chosen. Recall, Bayes formula is given by (Bayes and Price, 1763):

$$\pi(\theta | \mathbf{D}) = \frac{\pi(\theta)L(\mathbf{D} | \theta)}{\int_{\Theta} \pi(\theta)L(\mathbf{D} | \theta)d\theta}$$

where $\pi(\theta | \mathbf{D})$ is the posterior probability distribution for the parameters θ given the data \mathbf{D} , $\pi(\theta)$ is the prior distribution of θ and $L(\mathbf{D} | \theta)$ is the likelihood of the data given θ .

In order to specify the likelihood of the model in equation (1) the model modified to model the *mean* abundance in each compartment and is given by

$$\begin{aligned}\frac{d\lambda_S(t)}{dt} &= -W(t)^T \alpha \lambda_S(t) \lambda_E(t) \\ \frac{d\lambda_E(t)}{dt} &= W(t)^T \alpha \lambda_S(t) \lambda_E(t) - \beta \lambda_E(t) - \gamma \lambda_E(t) - \beta_A \lambda_E(t) \delta(t - \tau) \\ \frac{d\lambda_I(t)}{dt} &= \beta \lambda_E(t) + \beta_A \lambda_E(t) \delta(t - \tau) - \gamma \lambda_I(t) - \eta \lambda_I(t) \\ \frac{d\lambda_R(t)}{dt} &= \gamma \lambda_I(t) \\ \frac{d\lambda_D(t)}{dt} &= \eta \lambda_I(t)\end{aligned}\tag{2}$$

where $\lambda_S(t)$, $\lambda_E(t)$, $\lambda_I(t)$, $\lambda_R(t)$ and $\lambda_D(t)$ are the means of $S(t)$, $E(t)$, $I(t)$, $R(t)$ and $D(t)$, respectively and the parameters have the same definition as provided in the system given in equation (1). Since there is no data for $S(t)$ and $E(t)$, these compartments will be latent variables and will not directly factor into the likelihood. The likelihood for $I(t)$, $R(t)$ and $D(t)$ are given by

$$\begin{aligned}I(t) &\sim \text{Poisson}(\lambda_I(t)) \\ R(t) &\sim \text{Poisson}(\lambda_R(t)) \\ D(t) &\sim \text{Poisson}(\lambda_D(t)).\end{aligned}\tag{3}$$

To specify the prior distributions for α , β_A , β , γ and η one must incorporate the following constraints $\alpha > 0$, $\beta > 0$, $\gamma > 0$ and $\eta > 0$. Hence the following prior distributions are set:

$$\begin{aligned}\alpha &\sim \text{MVN}(a1, \sigma^2 \mathbf{I})C(\mathcal{A}) \\ \beta_A &\sim \text{Exp}(1) \\ \beta &\sim \text{Exp}(1) \\ \gamma &\sim \text{Exp}(1) \\ \eta &\sim \text{Exp}(1)\end{aligned}$$

where $C(\mathcal{A})$ is an indicator function takes the value 1 if $\alpha \in \mathcal{A}$. This serves to truncate the normal distribution in order to keep α in the feasible range of values.

The likelihood and prior distributions specifications lead to the following posterior distribution when $a = 1$ and $\sigma^2 = 1$:

$$\begin{aligned} \pi(\alpha, \beta_A, \beta, \gamma, \eta \mid \mathbf{D}) &\propto \pi(\alpha)\pi(\beta)\pi(\gamma)\pi(\eta)L(\mathbf{D} \mid \alpha, \beta_A, \beta, \gamma, \eta) \\ &= C(\mathcal{A})e^{-\frac{1}{2}(\alpha-1)^T(\alpha-1)-\beta_A-\beta-\gamma-\eta} \\ &\quad \times \prod_{t=1}^T \frac{\lambda_I(t)^{I(t)}\lambda_R(t)^{R(t)}\lambda_D(t)^{D(t)}e^{-\lambda_I(t)-\lambda_R(t)-\lambda_D(t)}}{I(t)!R(t)!D(t)!} \end{aligned} \quad (4)$$

The posterior distribution does not lend to any analytic solution, hence Markov chain Monte Carlo (MCMC) techniques will be used to sample from the posterior distribution (Gelman et al., 2013). Specifically Metropolis-Hastings sampler is used to obtain samples from the posterior distribution (Gilks et al., 1996; Albert, 2009). To tune the sampler, a series of short chains were generated and analyzed for convergence and adequate acceptance rates. These initial short chains were discarded as “burn-in” samples. The tuned sampler was used to generate 5,000 samples from $\pi(\alpha, \beta_A, \beta, \gamma, \eta \mid \mathbf{D})$ and trace plots were visually examined for convergence and deemed to be acceptable. All inferences will be made from these 5,000 samples. The model and sampling algorithm is custom programmed in the R statistical programming language version 3.6.3. The computation takes approximately 290 seconds using a AMD A10-9700 3.50GHz processor with 16GB of RAM to obtain 5,000 samples from the posterior distribution. For more on statistical inference see Wackerly et al. (2008); Casella and Berger (2002); Berger (1985).

6 Results

To apply the model the following initial conditions are specified: $S(0) = 2,782,000$, $E(0) = 3$, $I(0) = 1$, $R(0) = 0$ and $D(0) = 0$. Here $S(0)$ is the current population of the State of Qatar, $I(0)$, $R(0)$ and $D(0)$ are obtained directly from the data. The choice of $E(0)$ was used as it a minimal value that would allow the disease to spread but not so large as to make the spread rapid. Several values of $E(0)$ were explored and the value of 3 was found to have the best fit. Furthermore, model interventions were placed at days $t_1 = 12$, $t_2 = 19$, $t_3 = 36$, $t_4 = 40$ and $t_5 = 59$ with an Dirac delta impulse at time $\tau = 12$.

Table 1 shows the means, standard deviations and the 0.025%, 0.5% and 0.975% quantiles for the model parameters based on the 5,000 samples from the posterior distribution. Notice that, $\alpha_0 = 2.33 \times 10^{-7}$ and $\alpha_1 = -2.33 \times 10^{-7}$ (within rounding) are very close in magnitude with different signs indicating that the first intervention almost eliminated transmission. Similarly one can see that the second and third interventions $\alpha_2 = 1.70 \times 10^{-7}$ and $\alpha_3 = -1.70 \times 10^{-7}$ essentially are of the same magnitude with different signs which when added resulting in a very low transmission rate. However, $\alpha_4 = 6.39 \times 10^{-8}$ is a moderate increase with another moderate decrease in $\alpha_5 = -5.28 \times 10^{-8}$ which still leaves a final transmission rate of $\sum_{k=0}^K \alpha_k \approx 1.12 \times 10^{-8}$. Of particular note is the mean mortality rate $\eta = 0.00014 \approx 1/7142$ which means that about 1 in 7,142 people perish from the disease each day, which is quite low. Also note that the mean infection (confirmed) rate is $\beta = 0.07887 \approx 1/12.67$ which corresponds to about 1 in 12.67 exposed people become confirmed each day. The quantile intervals provide a 95% credible interval for the parameters and can be used to obtain a range of reasonable parameter values. For example for the parameter β the interval is (0.07319, 0.08381) meaning that the probability that β is between (0.07319, 0.08381) is 0.95. This can be used to create an interval for the risk interpretations as between $1/0.07319 \approx 13.66$ and $1/0.03298 \approx 11.93$ Exposed people are confirmed as infected each day. This also gives insight into how many people who may be in the population who are Exposed and may be infectious but do not yet exhibit symptoms. Recall, γ corresponds to the rate at which people recover from the disease which is 0.00977. This corresponds to $1/0.00977 \approx 1/102.35$ which corresponds to 1 in 102.35 people recover from the disease each day. This value may be artificially low due to delays in reporting.

While many of the parameters do not lend well to the traditional $H_0 : \theta = 0$ hypothesis testing as they must be positive. We can conduct simple hypothesis tests on the α parameters to look for significant changes due to interventions using contrasts. Specifically the sequential contrasts of $\alpha_1 - \alpha_0$, $\alpha_2 - \alpha_1$, $\alpha_3 - \alpha_2$, $\alpha_4 - \alpha_3$ and $\alpha_5 - \alpha_4$. These contrasts quantify the changes that in transmission rate from Susceptible to Exposed due to the interventions and are what policy makers want to see. Furthermore, they want a statistical test on whether or not the intervention performed in a statistically significant manner. This can be done by simply subtracting the MCMC samples to generate the contrast of interest. Using these subtracted samples one can look at the mean, standard deviation, quantiles and the proportion of samples above 0, $P(> 0)$. Table 2 shows these quantities for the contrasts listed above. Notice that the intervention at day 12 reduced the transmission rate by approximately 4.66×10^{-7} which is considerable and the proportion of samples above 0 was 0.000 indicating a statistically significant change due to the intervention. The intervention taken at day 19, $\alpha_2 - \alpha_1$, actually increased the transmission rate, where the intervention taken at day 36, $\alpha_3 - \alpha_2$, then reduced the transmission rate. Similarly the other two interventions increased and then decreased the transmission rates, respectively. Furthermore, all interventions be deemed statistically significant since $P(> 0)$ is either 0.000 or 1.000 indicating significance.

Table 1: Mean, Standard Deviation and $(Q_{0.025}, Q_{0.5}, Q_{0.975})$ for $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \beta_A, \beta, \gamma$ and η . Based on 5,000 samples from the posterior distribution

Parameter	Mean	Std Dev.	$(Q_{0.025}, Q_{0.5}, Q_{0.975})$
α_0	2.33×10^{-7}	9.30×10^{-10}	$(2.31 \times 10^{-7}, 2.33 \times 10^{-7}, 2.34 \times 10^{-7})$
α_1 (day 12)	-2.33×10^{-7}	1.03×10^{-9}	$(-2.34 \times 10^{-7}, -2.33 \times 10^{-7} - 2.31 \times 10^{-7})$
α_2 (day 19)	1.70×10^{-7}	1.01×10^{-9}	$(1.69 \times 10^{-7}, 1.70 \times 10^{-7}, 1.72 \times 10^{-7})$
α_3 (day 36)	-1.70×10^{-7}	1.11×10^{-9}	$(-1.72 \times 10^{-7}, -1.70 \times 10^{-7}, -1.69 \times 10^{-7})$
α_4 (day 40)	6.39×10^{-8}	1.68×10^{-9}	$(6.08 \times 10^{-8}, 6.42 \times 10^{-8}, 6.64 \times 10^{-8})$
α_5 (day 59)	-5.28×10^{-8}	8.24×10^{-9}	$(-6.47 \times 10^{-8}, -5.39 \times 10^{-8}, -3.32 \times 10^{-8})$
β_A (day 12)	1.45777	0.00841	(1.44649, 1.45598, 1.47108)
β	0.07887	0.00351	(0.07319, 0.07966, 0.08381)
γ	0.00977	8.71×10^{-5}	(0.00962, 0.00977, 0.00995)
η	0.00014	9.26×10^{-6}	(0.00012, 0.00014, 0.00016)

Table 2: Mean, Standard Deviation, $(Q_{0.025}, Q_{0.5}, Q_{0.975})$ and proportion of samples larger than zero $P(> 0)$ for sequential contrasts across α . Based on 5,000 samples from the posterior distribution

Contrast	Mean	Std Dev.	$(Q_{0.025}, Q_{0.5}, Q_{0.975})$	$P(> 0)$
$\alpha_1 - \alpha_0$	-4.66×10^{-7}	1.95×10^{-10}	$(-4.69 \times 10^{-7}, -4.66 \times 10^{-7}, -4.63 \times 10^{-7})$	0.000
$\alpha_2 - \alpha_1$	4.03×10^{-7}	4.25×10^{-10}	$(4.02 \times 10^{-7}, 4.03 \times 10^{-7}, 4.04 \times 10^{-7})$	1.000
$\alpha_3 - \alpha_2$	-3.41×10^{-7}	2.12×10^{-9}	$(-3.45 \times 10^{-7}, -3.40 \times 10^{-7}, -3.38 \times 10^{-7})$	0.000
$\alpha_4 - \alpha_3$	2.34×10^{-7}	7.62×10^{-10}	$(2.33 \times 10^{-7}, 2.34 \times 10^{-7}, 2.36 \times 10^{-7})$	1.000
$\alpha_5 - \alpha_4$	-1.16×10^{-8}	9.15×10^{-9}	$(-1.30 \times 10^{-7}, -1.18 \times 10^{-7}, -9.44 \times 10^{-8})$	0.000

The model formulation also allows for the individual transmission rates to be computed by simply summing up the α_k through to the desired time point. Table 3 gives the mean, standard deviation and $(Q_{0.025}, Q_{0.5}, Q_{0.975})$ for the transmission rate of Exposed to Infected across each time interval. This is done by simply add the corresponding MCMC samples. This is another perspective on how the transmission rate changes across the time frame. Notice that all of the transmission rates are positive which is required by the model specification. Also notice that the mean transmission rates vary in orders of magnitude from 2.40×10^{-10} to 2.33×10^{-7} . One interesting point that should be made is the highest transmission rate is at the beginning and the lowest transmission rate is at the end. This is evidence that the interventions taken by the Qatari government has ultimately reduced the transmission rate.

To assess the fit of the model the posterior predictive distribution was used and is given by

$$\pi(I_{new}(t), R_{new}(t), D_{new}(t) | \mathbf{D}) = \int L(I_{new}(t), R_{new}(t), D_{new}(t) | \alpha, \beta_A, \beta, \gamma, \eta) \times \pi(\alpha, \beta_A, \beta, \gamma, \eta | \mathbf{D}) d\alpha d\beta_A d\beta d\gamma d\eta.$$

Using the samples 5,000 samples from the posterior distribution 5,000 samples were generated from the posterior predictive distribution. At each time t the median, 0.025 and 0.975 quantiles were obtained to form a posterior predictive interval.

Figure 2 shows the model fits for Active Infections, Recovered and Deaths with posterior predictive bands. Notice that, the model does quite well at fitting the dynamics of the Active Infections including the jump at day 12 and captures the plateau and the exponential growth after the plateau as well. The Recovered model fits well as does the Deaths data. To assess the explained

Table 3: Mean, Standard Deviation and $(Q_{0.025}, Q_{0.5}, Q_{0.975})$ of transmission rate of Exposed to Infected for the specified time intervals.

Time	Sum	Mean	Std Dev.	$(Q_{0.025}, Q_{0.5}, Q_{0.975})$
$0 \leq t < 12$	α_0	2.33×10^{-7}	9.30×10^{-10}	$(2.31 \times 10^{-7}, 2.33 \times 10^{-7}, 2.34 \times 10^{-7})$
$12 \leq t < 24$	$\sum_{k=0}^1 \alpha_k$	2.40×10^{-10}	1.82×10^{-10}	$(8.16 \times 10^{-12}, 2.01 \times 10^{-10}, 6.61 \times 10^{-10})$
$24 \leq t < 28$	$\sum_{k=0}^2 \alpha_k$	1.71×10^{-7}	1.09×10^{-9}	$(1.69 \times 10^{-7}, 1.70 \times 10^{-7}, 1.72 \times 10^{-7})$
$28 \leq t < 40$	$\sum_{k=0}^3 \alpha_k$	1.43×10^{-10}	1.46×10^{-10}	$(3.07 \times 10^{-12}, 1.00 \times 10^{-10}, 5.70 \times 10^{-10})$
$40 \leq t < 59$	$\sum_{k=0}^4 \alpha_k$	6.41×10^{-8}	1.69×10^{-9}	$(6.10 \times 10^{-8}, 6.43 \times 10^{-8}, 6.66 \times 10^{-8})$
$59 \leq t$	$\sum_{k=0}^5 \alpha_k$	1.12×10^{-8}	7.58×10^{-9}	$(5.17 \times 10^{-10}, 1.04 \times 10^{-8}, 2.88 \times 10^{-8})$

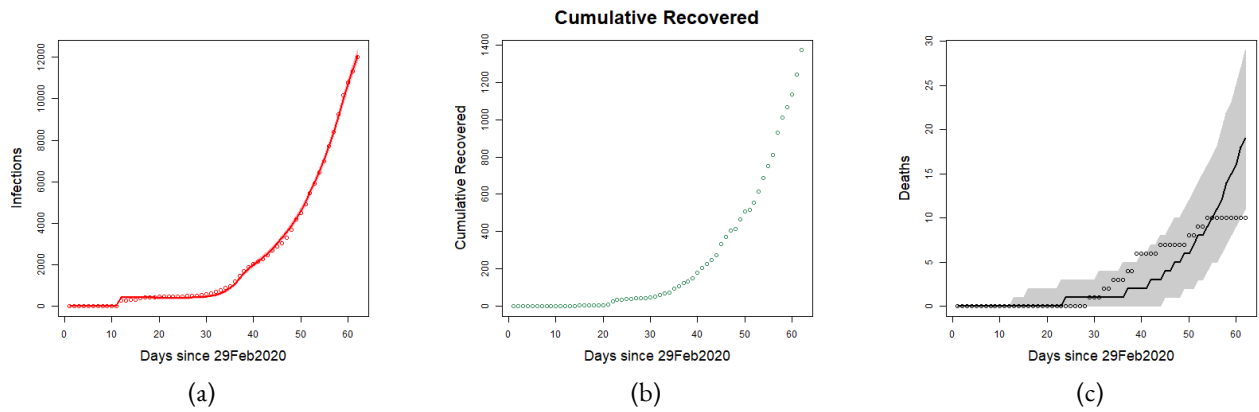


Figure 2: Plots of Total Active Infections (a), Cumulative Recovered (b) and Cumulative Deaths (c) for the State of Qatar for the days since 29 February 2020 until 1 May 2020 with posterior predictive bands. Posterior predictive bands are based on the 0.025 and 0.975 Quantiles from 5,000 samples from the posterior predictive distribution.

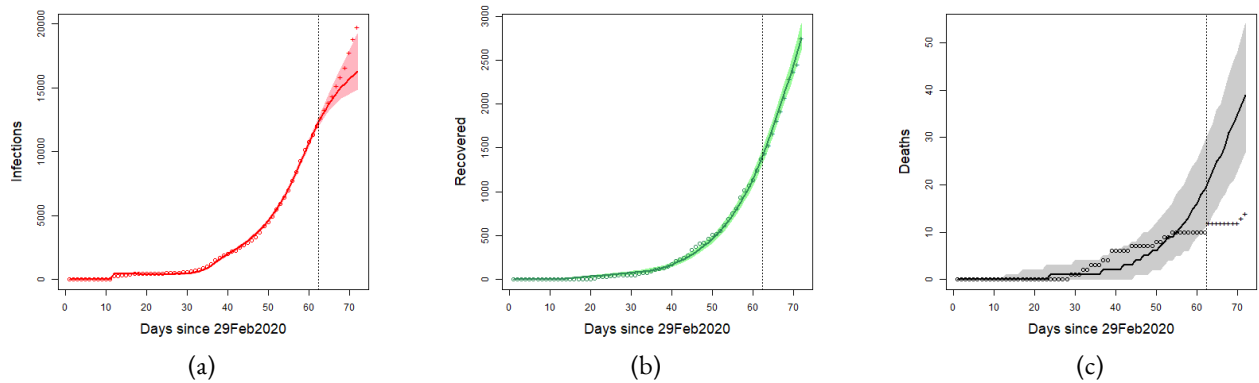


Figure 3: Plots of Total Active Infections (a), Cumulative Recovered (b) and Cumulative Deaths (c) for Qatar for the days since 29 February 2020 until 1 May 2020 with posterior predictive bands with actual values from 2 May 2020 until 10 May 2020 denoted with (+). Vertical line separates the training data from the predictions. Posterior predictive bands are based on the 0.025 and 0.975 Quantiles from 5,000 samples from the posterior predictive distribution.

variance a pseudo- R^2 was formed using the median from the posterior predictive distribution at each time as the point estimates. This, resulted in a pseudo- R^2 of 0.998 which indicates the fitted model explains approximately 99.8% of the variance in the data. Based on this the model is deemed to fit well. It should be noted that standard data splitting procedures for model validation are difficult in this scenario as removing values from the system may cause unstable behavior.

7 Predictive Performance

To assess the model performance predictive performance is utilized with days from 2 May 2020 and 10 May 2020 used as a test set. Using the samples from the posterior distribution the posterior predictive distribution was computed for each day of the test set and 95% posterior predictive intervals were created using the 2.5% and 97.5% quantiles. The test data is then compared to the posterior predictive intervals for each of the endpoints.

Figure 3 shows the data, predictive bands and the actual values for Active Infections (a), Recovered (b) and Deaths (c). Training data is presented by (o) and test set data is presented by (+) and the vertical line separates the training time frame from the prediction time frame. Notice that, the Active Infections (panel a) performs incredibly well with all test set points in the predictive interval. The Recovered (panel b) performs equally well. Deaths (panel c) performs the worst as the model consistently over predicts the true number of deaths. Overall the model does reasonably well at predicting the test set.

Another view of predictive performance is to examine pseudo-predictive- R^2 which compares the predicted values with the actual values for the test set. This calculation leads to a pseudo-predictive- $R^2 \approx 0.923$ which is lower than the pseudo- R^2

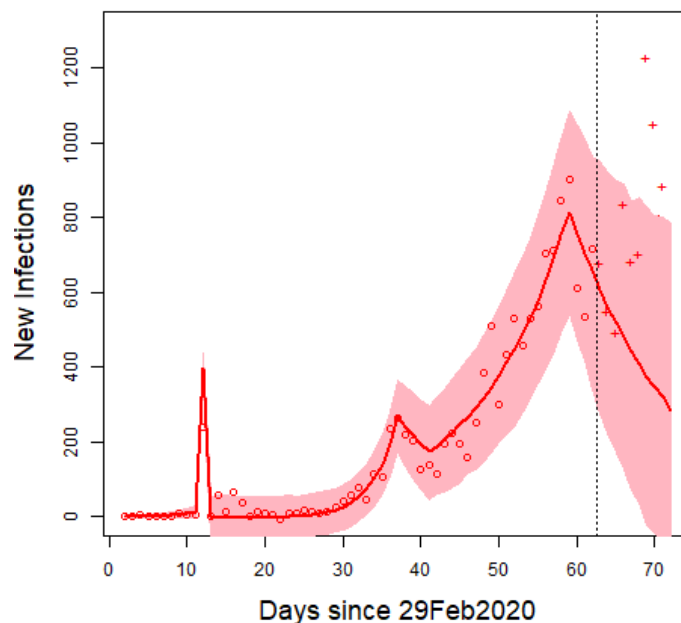


Figure 4: Plot of Total New Infections for Qatar for the days since 29 February 2020 until 1 May 2020 with posterior predictive bands with actual values from 2 May 2020 until 10 May 2020 denoted with (+). Posterior predictive point estimates are based on the median and bands are based on the 0.025 and 0.975 Quantiles from 5,000 samples from the posterior predictive distribution at each time point.

associated with the fit of the model to the data in the training set but is still very high. Of course several other measures of predictive performance exist however this is the easiest to understand as it measures the amount of variation explained by the predictions across the test set.

8 Discussion

This work has demonstrated how to build a SEIRD model for the COVID-19 outbreak in the State of Qatar, include interventions, estimate model parameters and generate posterior predictive intervals using a Bayesian framework. Furthermore, the model is able to treat the Susceptible and Exposed compartment as latent variables, as no data is observed about them other than approximate initial values. The model fits the data quite well with a pseudo- $R^2 \approx 0.998$ and predicts reasonably well with pseudo-predictive- $R^2 \approx 0.923$. One can also note that in the model definition, no immigration, emigration, natural births and natural mortality were not included and based on the high pseudo- R^2 would have a negligible effect on fit.

The modeling paradigm is quite flexible for modeling the COVID-19 data as it easily incorporates interventions into the system and can quantify the impact of the intervention. Furthermore, using simple differences the model can be used to predict new infections as well. Figure 4 shows the plots of the new infections with predictive bands based on 2.5% and 97.5% based on 5,000 samples from the posterior prediction distribution at each time point. Notice that the model does well at capturing the jump at day 12 and the bands capture most of the data. It is interesting how the dynamics change as the public sector is moved to working from home with an increase in infection rate compared to when

Another use of the model may be for long term predictions. While this is extrapolation, it does provide policy makers a tool for planning, provided nothing changes, i.e., no interventions are taken. It also allows policy makers to see the possible long term effects of their decisions. Figure 5 shows the long-term predictions of the model if no other interventions are made past 1 May 2020. Notice that the predictions do eventually decrease across the future time frame. Notice the width of the predictive bands for times farther in the future. This reflects the uncertainty associated with extrapolating into the future. However, one item that we can calculate from this is a 95% predictive interval for the peak infection time. By simply recording the maximum value for each of the predictive distribution trajectories from the MCMC samples one can obtain a distribution of the time for the maximum. In this case this gives the 95% predictive interval for the maximum to be (78,162). This means that the peak infection

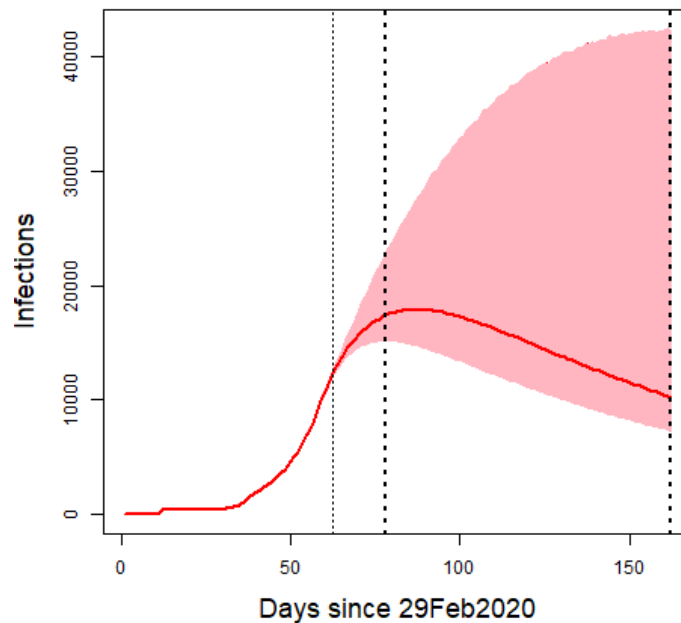


Figure 5: Plot of Long-term New Infections for Qatar for the days since 29 February 2020 until 28 July 2020 with posterior predictive bands. Training data from 29 February 2020 to 1 May 2020. Posterior predictive bands are based on the 0.025 and 0.975 Quantiles from 5,000 samples from the posterior predictive distribution. Dark dashed vertical lines give the 95% predictive interval for the maximum active infections.

time will be between day 78 (17 May 2020) and day 162 (9 August 2020) of the outbreak given that no other interventions or process changes occur. Fig. 5 also shows this interval given by the dark dashed vertical lines. The width of the interval quantifies the uncertainty about where the maximum active infections will occur. Since the width of the interval is 84 days, this indicates that there is a large amount of uncertainty on when the number of active infections will begin to decline.

To address possible sensitivity issues to the initial Exposed value $E(0) = 3$. By looking at the data values of $E(t)$ To study this the model was run again with values $E(0) = 5, 10, 15$ and 25 and the model pseudo- R^2 was calculated for each. All runs were under the exact same prior distribution specifications and using the same MCMC procedures. For $E(0) = 5$ the model produces a pseudo- $R^2 = 0.999$, with $E(0) = 10$ produces a pseudo- $R^2 = 0.990$, using $E(0) = 15$ a pseudo- $R^2 = 0.967$ is found and finally, with $E(0) = 25$ a pseudo- $R^2 = 0.865$ is given. This suggests that the model is not sensitive to low values of $E(0)$ but values that are too large produce a poorer and poorer performance especially at the beginning time frame.

Future work could be to add an overdispersion parameter into the model to allow for the more accurate capture of uncertainty. Furthermore, one can perform simulation studies to better understand how the model may perform under various scenarios. Feature selection methods could be employed to select where the interventions should be placed as well as other forms of interventions could be included in the model. Another possibility to address any deviations from the standard model a semi-parametric technique could be studied as well.

Acknowledgments

The authors would like to acknowledge the State of Qatar and the Ministry of Health for the daily updates and additional data. In addition the authors would like thank Virginia Commonwealth University in Qatar and Qatar University for supporting this effort.

References

Albert, J. (2009). *Bayesian Computation with R: Second Edition*, Springer, New York. 23

- Bayes, T., and R. Price (1763). An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price in a letter to John Canton. *A.M.F.R.S. Philosophical Transactions of the Royal Society of London* 53, 370–418. 22
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, Second Edition*, Springer, New York. 23
- Casella, G., and R. L. Berger (2002). *Statistical Inference, Second Edition*. Duxbury, Belmont, CA. 23
- Chinazzi, M., J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. Piontti, K. Mu, L. Rossi, K. Sun, and C. Viboud (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 368(6489), 395–400. 19
- Department of Health (2020). *Australian Health Sector Emergency Response Plan for Novel Coronavirus (COVID-19)*. Canberra, Australia. 21
- Department of Health (2020). *Coronavirus: action plan – A guide to what you can expect across the UK*. London, UK. 21
- Dirac, P. (1958). *The Principles of Quantum Mechanics, Fourth Edition*, Oxford at Clarendon Press. 22
- Fisher, D., and D. Heymann (2020). Q&A: The novel coronavirus outbreak causing COVID-19. *BMC Medicine* 18(57), <https://doi.org/10.1186/s12916-020-01533-w>. 19
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis, Third Edition*, CRC Press, Boca Raton, FL. 23
- Gilks, W. R., S. Richardson, D. J. Spiegelhalter (1996). *Markov chain Monte Carlo in Practice*, Chapman & Hall, CRC Press, Boca Raton, FL. 23
- Giuliani, D., M. M. Dickson, G. Espa, and F. Santi (2020). *Modelling and predicting the spread of Coronavirus (COVID-19) infection in NUTS-3 Italian regions*. arXiv preprint <https://arxiv.org/abs/2003.06664>. 19
- Hamad Medical Corporation. (2020). *Major Risks to Business Continuity*. Doha, Qatar. 21
- McIntosh, K. (2020). Coronavirus disease 2019 (COVID-19). Up-To-Date, <https://www.uptodate.com/contents/coronaviruses>. Accessed May 8, 2020. 19
- Ministry of Public Health (2019). *Qatar National Preparedness and Response Plan for Communicable Diseases*. Doha, Qatar. 21
- Rezabakhsh, A., A. Ala, and S. H. Khodaei (2020). Novel Coronavirus (COVID-19): A New Emerging Pandemic Threat. *Journal of Research in Clinical Medicine* 8(1), 5–6. 19
- Wackerly, D., W. Mendenhall, and R. L. Scheaffer (2008). *Mathematical Statistics with Applications, 7th Edition*. Thomson Brooks/Cole, Belmont, CA. 23
- World Health Organization (2020). *Coronavirus*, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed May 8, 2020. 19
- World Health Organization (2020). *COVID-19 Strategic Preparedness and Response Plan Operational Planning Guidelines to Support Country Preparedness and Response*. Geneva, Switzerland. 21
- Wu, F., S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, and M. L. Yuan (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. 19