COVID-19 Incubation and Quarantine Times: Is a 14-Day Quarantine Period the Right Quarantine Time?

Yasin Khadem Charvadeh[1], Yuan Bian[1], Grace Y. Yi[1,2], and Wenqing He *[1]

[1]*Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada*

[2]*Department of Computer Science, University of Western Ontario, London, Ontario, Canada*

Abstract

To confine the spread of an infectious disease, setting a sensible quarantine time is crucial. To this end, it is imperative to well understand the distribution of incubation times of the disease. Regarding the ongoing COVID-19 pandemic, the 14-days is currently taken as a quarantine time almost by all the places. However, it is unclear how likely it is for an exposed individual to show signs of illness after being quarantined for 14-days. To explore the plausibility of setting 14-days as a quarantine time of COVID-19, we analyze a dataset of 178 COVID-19 cases dated from January 20, 2020 to February 29, 2020, with the information of exposure periods and dates of symptom onset collected. To gain a good understanding of possible scenarios, we employ different models to analyze incubation times of COVID-19 from different angles. Our findings suggest that the current recommended 14-day quarantine time is not long enough to control the probability of an early release of infected individuals to be small.

Keywords: COVID-19; incubation times; profile likelihood; quantile estimation; quarantine time

1 Introduction

As of January 2021, more than 90 million confirmed COVID-19 cases and 1.9 million deaths have been reported in more than 200 countries and regions [WHO]. To control the virus spread, WHO suggested a 14-day quarantine time

*whe@stats.uwo.ca

for potential COVID-19 infected cases. The 14-day quarantine guidance is primarily recommended to accommodate the incubation period of the virus, which is the length of time for a person exposed to the virus to become infectious.

Since the COVID-19 pandemic starts, many authors studied the incubation information. For example, Li et al. [2020] studied the incubation time using the data on the first 425 confirmed infections in Wuhan city, China. They assumed a generalized gamma distribution for the incubation times, and estimated that the average incubation time was 5.2 days (95% confidence interval: 4.1-7.0 days) and the 95th percentile of incubation times was 12.5 days. Nevertheless, in their dataset, only 10 cases were able to identify the exact date of exposure to the virus. Backer et al. [2020] used travel histories and symptom onsets of 88 confirmed cases to characterize the distribution of incubation times. They estimated the mean incubation time to be 6.4 days with a 95% confidence interval of 5.6-7.7 days. Charvadeh and Yi [2020] estimated the mean and median of the incubation period to be 5.8 and 5 days, respectively, by examining a cohort of 3397 infected cases dated from January 22, 2020 to March 29, 2020. Examining 1084 confirmed COVID-19 cases who initially had shown no signs of illness at their time of departure from Wuhan city, China, Qin et al. [2020] conducted a forward follow-up study. They estimated that the median incubation time was 7.76 days (95% confidence interval: 7.02-8.53 days), the mean incubation time was 8.29 days (95% confidence interval: 7.67-8.9 days), the 90th percentile of incubation times was 14.28 days (95% confidence interval: 13.64-14.90 days), and the 99th percentile of incubation times was 20.31 days (95% confidence interval: 19.15-21.47 days). He et al. [2020] employed a meta-analysis method to combine five studies up to February 2020 and estimated the average incubation time to be 5.08 days with a 95% confidence interval of 4.8-5.4 days. Banka and Comiskey [2020] conducted a modeling study to provide an updated estimate and meta-analysis of the incubation period distribution for COVID-19. Their findings highlighted the need for a longer quarantine time than the initially suggested 14-days by the WHO. Quilty et al. [2020] used an agent-based model to simulate incubation periods. Their findings indicated that self-isolation can prevent 39% of onward transmission from secondary cases, and that 14-day post-exposure quarantine for all contacts reduces transmission by 70%. In the study by Jiang et al. [2020], they manually collated the clinical data of 2015 COVID-19 patients from official websites of local Chinese health agencies reported between January 1, 2020 and February 25, 2020. The cohort in their study was believed to represent a wide spectrum of COVID-19 cases with different age groups as well as hospitalized and non-hospitalized cases included. Their findings showed that the incubation time of COVID-19 ranges from 0 to 33 days, and the median incubation periods for adults and children are 7 and 9 days, respectively. Furthermore, they recommended extending the quarantine period for adults from 14 days to 18, or even 21 days for a more effective quarantine.

While the principle of setting a quarantine time is clear, it is challenging to determine a plausible quarantine time because of the uncertainty and different revealings from different studies. Public health officials often use

the tail end of the incubation range to determine a quarantine time [Public Health Ontario]. A good understanding of the distribution of incubation times thereby becomes critical. However, studying incubation times is not trivial and a definite answer can not be found, because incubation periods of infected cases are heterogeneous and vary considerably. The incubation time of a case is associated with many factors such as gender, age, socio-economic status, underlying health conditions, and the disease transmission method (i.e., direct contact or indirect contact). It is therefore imperative to broadly examine the possible distribution of COVID-19 incubation times from different angles in order to uncover the underlying truth.

In this study, we aim to investigate whether the currently used 14-day quarantine period is long enough to effectively control the virus transmission for COVID-19 by examining different ways of modelling the distribution of incubation times of COVID-19. We study how the choice of a model may affect the decision of setting plausible values for the quarantine time. Our explorations are carried out on a dataset of 178 COVID-19 cases with clear-defined exposure periods and dates of symptom onset dated from January 20, 2020 to February 29, 2020. Our findings suggest that the current recommended 14-day quarantine time is not long enough to control the probability of an early release of infected individuals to be small.

The remainder of the article is organized as follows. In Section 2 we describe the data and outline several distributions that are useful for describing incubation times of an infectious disease. We present the methodology employed for inferential purposes and discuss modelling of incubation times of COVID-19 with the use of a real COVID-19 dataset when incubation times or interval-censored incubation times are available, respectively in Sections 3 and 4. We investigate different analyses in the hope of revealing broad scenarios for setting a sensible quarantine time for COVID-19. In Section 5 we explore the incubation times to determine the quarantine time based on percentiles. We conclude the manuscript with discussion in the last section.

2 COVID-19 Data and Model Framework

2.1 COVID-19 Data

We consider the data of 178 COVID-19 cases in Shiyan city, Hubei province, China, which were reported between January 20, 2020 and February 29, 2020 [Yang et al., 2020]. For each case in the dataset, the information, including gender, age, the source of infection, the date of exposure, the date of symptom onset, and the date of diagnosis, was collected through epidemiological survey. The date of exposure refers to either (a) entry and exit dates into Wuhan or (b) the earliest and latest dates of close contact with a Wuhan-imported/locally infected case. Let S_i be the

symptom onset time for case i, and let E_{Li} and E_{Ui} be the lower and upper time points for the exposure period for case i, respectively. Then, shown in Figure 1, the incubation time for case i is bounded by the interval $[t_{Li}, t_{Ui}]$, where $t_{Li} = S_i - E_{Ui}$ and $t_{Ui} = S_i - E_{Li}$. Table 1 reports t_{Li}, and t_{Ui} for each of the 178 cases. Figure 2 illustrates the interval incubation times by gender and three different age groups.

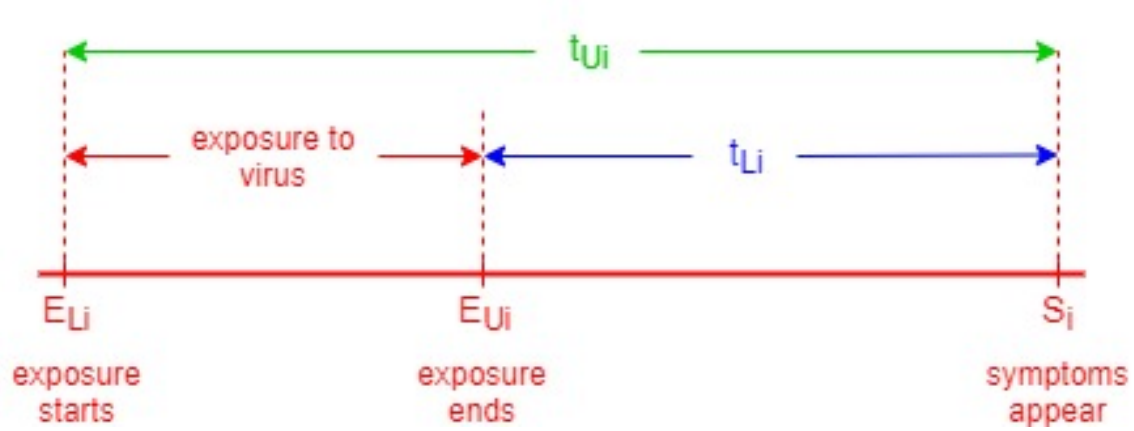


Figure 1: Timeline of the incubation period

2.2 Useful Distributions

Parametric modeling is often employed to describe the distribution of incubation times of an infectious disease. Typically, distributions such as gamma, Weibull, log-normal, and generalized gamma are commonly used to study the incubation time of SARS [Cowling et al., 2007]. Since SARS-CoV-2 virus is genetically closely related to severe acute respiratory syndrome coronavirus (SARS-CoV), those distributions are also considered in studies of the COVID-19 incubation time (e.g., McAloon et al. [2020]).

Let T denote a continuous non-negative random variable representing the incubation time with an observed value, denoted t, and let $f(t)$ denote its probability density function (pdf). It is often convenient to work with the log-location-scale distribution of T. For this, we start from a random variable, say Z, with the distribution, say $g(z)$, of the support on $(-\infty, \infty)$, and consider a family of distributions of the location-scale form:

$$\log T = \beta_0 + \sigma Z, \tag{1}$$

where β_0 and $\sigma > 0$ are parameters.

Due to its versatile nature, the Weibull distribution is one of the most widely used distributions for describing the incubation time of a disease. Letting Z in (1) have the standard extreme value distribution with the density

Table 1: Intervals bounded incubation times of the study subjects

ID	t_{Li}	t_{Ui}	ID	t_{Li}	t_{Ui}	ID	t_{Li}	t_{Ui}	ID	t_{Li}	t_{Ui}	ID	t_{Li}	t_{Ui}	ID	t_{Li}	t_{Ui}
1	1	3	2	0	4	3	4	7	4	0	4	5	1	2	6	0	3
7	0	3	8	8	10	9	4	9	10	3	6	11	0	6	12	0	3
13	1	5	14	4	7	15	2	3	16	1	5	17	8	12	18	5	6
19	2	7	20	0	4	21	6	8	22	4	10	23	0	5	24	1	5
25	3	10	26	1	7	27	4	7	28	1	3	29	1	6	30	3	6
31	3	6	32	0	5	33	3	7	34	2	6	35	6	9	36	4	7
37	0	7	38	8	9	39	0	7	40	3	7	41	1	3	42	2	9
43	4	5	44	1	2	45	0	3	46	5	6	47	0	7	48	6	7
49	2	6	50	3	8	51	1	6	52	4	7	53	4	8	54	4	5
55	4	9	56	1	3	57	8	13	58	6	8	59	9	14	60	4	7
61	4	7	62	4	9	63	9	12	64	2	6	65	4	6	66	8	9
67	4	10	68	9	13	69	1	6	70	6	9	71	7	8	72	5	7
73	3	4	74	5	9	75	4	6	76	5	8	77	10	12	78	2	8
79	4	6	80	2	4	81	5	9	82	8	15	83	11	12	84	7	9
85	2	8	86	2	4	87	1	4	88	9	11	89	2	3	90	1	3
91	5	7	92	4	8	93	4	5	94	6	13	95	1	8	96	5	8
97	4	6	98	4	8	99	4	8	100	2	4	101	2	7	102	0	4
103	9	12	104	9	10	105	6	9	106	9	12	107	3	7	108	14	17
109	2	6	110	6	9	111	9	13	112	8	13	113	2	8	114	14	15
115	7	11	116	8	13	117	10	13	118	4	5	119	5	9	120	4	7
121	10	13	122	7	9	123	0	6	124	6	11	125	7	9	126	6	12
127	8	9	128	7	11	129	9	10	130	14	17	131	3	7	132	5	6
133	12	19	134	6	11	135	5	7	136	11	13	137	7	9	138	6	9
139	2	8	140	8	12	141	9	14	142	7	9	143	13	15	144	4	7
145	9	10	146	1	5	147	11	14	148	9	11	149	13	14	150	7	9
151	9	11	152	9	13	153	7	12	154	8	10	155	3	6	156	4	7
157	11	14	158	6	9	159	16	19	160	12	15	161	11	15	162	4	9
163	9	11	164	11	13	165	6	12	166	11	14	167	2	5	168	9	15
169	13	14	170	0	4	171	18	24	172	6	9	173	3	6	174	1	4
175	17	20	176	8	13	177	7	9	178	7	9						

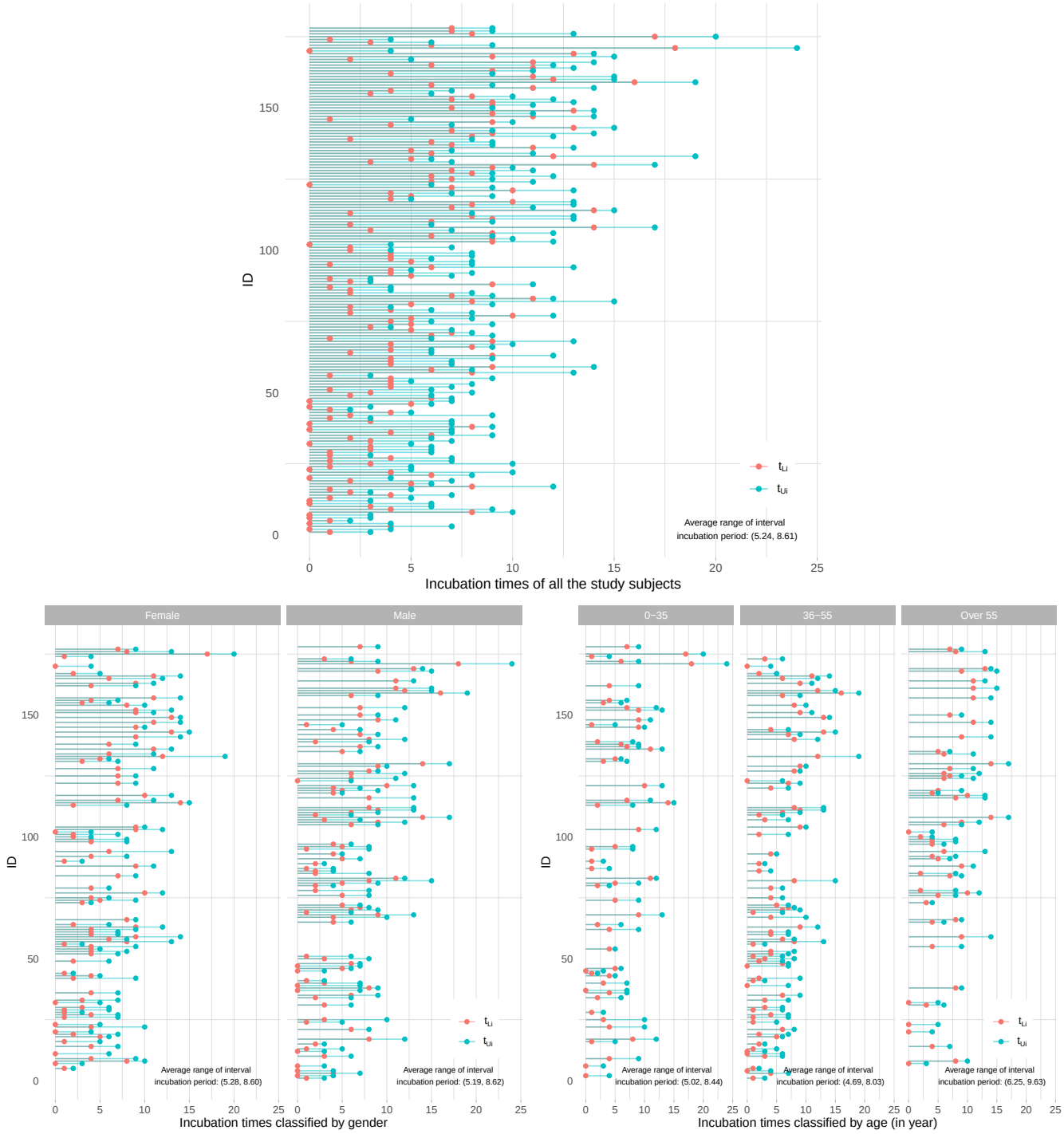


Figure 2: Lollipop charts for the intervals $[t_{Li}, t_{Ui}]$ in which the incubation times are bounded

$g(z) = \exp\{z - \exp(z)\}$, one obtain that the incubation time T has a Weibull distribution with the density function

$$f(t) = \lambda^p p t^{p-1} \exp\{-(\lambda t)^p\},$$

where $\lambda = \exp(-\beta_0)$ and $p = 1/\sigma$.

Another commonly used distribution for characterizing incubation times is the gamma distribution. If setting $\sigma = 1$ and letting Z in (1) have a generalized extreme-value distribution with the density $g(z) = \frac{\exp\{kz - \exp(z)\}}{\Gamma(k)}$, we obtain that the incubation time T follows a gamma distribution with the density function

$$f(t) = \frac{\lambda(\lambda t)^{k-1}\exp(-\lambda t)}{\Gamma(k)},$$

where $\lambda = \exp(-\beta_0)$, $k > 0$ is the shape parameter, and $\Gamma(\cdot)$ represents the gamma function.

If Z in (1) has a standard normal distribution with the density $g(z) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{z^2}{2}\}$, then the incubation time T has a log-normal distribution with the density function

$$f(t) = \frac{1}{(2\pi)^{1/2}\sigma t}\exp\left\{-\frac{1}{2}\left(\frac{\log\, t - \beta_0}{\sigma}\right)^2\right\}.$$

The aforementioned distributions can be cast under a more general distribution form. The incubation time T has a generalized gamma distribution with the density function

$$f(t) = \begin{cases} \frac{|k|}{\sigma t\Gamma(k^{-2})}\left[k^{-2}\{\exp(-\beta_0)t\}^{k/\sigma}\right]^{k^{-2}}\exp\left[-k^{-2}\{\exp(-\beta_0)t\}^{k/\sigma}\right], & \text{if } k \neq 0 \\ \frac{1}{(2\pi)^{1/2}\sigma t}\exp\left\{-\frac{1}{2}\left(\frac{\log\, t-\beta_0}{\sigma}\right)^2\right\}, & \text{if } k = 0 \end{cases},$$

if a standard generalized gamma distribution is specified for $\exp(Z)$ in (1), where β_0 is the location parameter, and σ and k are the scale and shape parameters, respectively. Generalized gamma distributions include Weibull ($k = 1$), gamma ($k = \sigma$), exponential ($k = \sigma = 1$) and log-normal ($k = 0$) distributions as special cases [Cox et al., 2007].

The log-logistic distribution is another choice for modeling incubation times [Cooley et al., 1996, Lover et al., 2014]. The incubation time T has a log-logistic distribution with the density function

$$f(t) = \frac{\lambda p(\lambda t)^{p-1}}{\{1 + (\lambda t)^p\}^2},$$

if Z in (1) follows a standard logistic distribution with the density $g(z) = \frac{\exp(z)}{\{1+\exp(z)\}^2}$, where $\lambda = \exp(-\beta_0)$ and $p = 1/\sigma$.

3 Analysis Methods When Incubation Times Are Available

3.1 Estimation Procedures

With the preparations in Section 2.2, we now turn to the problem of our interest. We are interested in delineating the distribution of incubation times associated with COVID-19. Before we present detailed analysis in the following section, here we introduce the basic notation and discuss the estimation strategy.

Let T denote the incubation time of COVID-19 for *any* individual, and assume that the distribution of T is characterized by the pdf $f(t;\theta)$, where $f(\cdot)$ takes a form from a class of distributions useful for describing incubation times of infectious diseases, such as those outlined in Section 2.2; θ is the parameter whose value is unknown. Let $F(\cdot;\theta)$ denote the corresponding cumulative distribution function of T. Let M_0 denote the *maximum* incubation time of *all* the individuals in the target population, thus, implicitly, we have $T \leq M_0$. Suppose that we have a sample of incubation times, $\{t_1,\cdots,t_n\}$, of n randomly selected COVID-19 patients.

One may be interested in estimating M_0 using the sample information. To this end, let M denote a time that is regarded to be sufficiently large to exceed any observed incubation time; in other words, the event of incubation times longer than M virtually has the probability of zero. For the clarity of exposition, we consider those values of M beyond a sufficiently large positive constant, say, M_1. That is, we take $[M_1,\infty)$ to be the parameter space of M in the following development.

We now consider the density function of T truncated by M, given by

$$\begin{aligned} f_{trunc}(t;\theta,M) &= \frac{f(t;\theta)}{P(T \leq M)} \\ &= \frac{f(t;\theta)}{F(M;\theta)}, \end{aligned}$$

where $F(M;\theta) = \int_0^M f(u;\theta)du$. Then the likelihood function of θ and M is obtained as

$$\begin{aligned} L(\theta,M|t_1,\cdots,t_n) &= \prod_{i=1}^{n} f_{trunc}(t_i;\theta,M) \\ &= \frac{\prod_{i=1}^{n} f(t_i;\theta)}{\{F(M;\theta)\}^n} \qquad \text{if } t_{(n)} \leq M, \end{aligned} \tag{2}$$

where $t_{(n)} = \max(t_1,\cdots,t_n)$.

Estimation of θ and M may be carried out using the likelihood method by maximizing $L(\theta,M|t_1,\cdots,t_n)$ with respect to θ and M simultaneously. However, as θ and M govern the likelihood differently, we employ an alternate

by breaking the one-step maximization into a two-step procedure. That is, we take the profile likelihood approach by maximizing the profile likelihood for one parameter with the other fixed at a given value.

Specifically, with a given value M, analogous to Farewell et al. [2005], we maximize $L(\theta, M|t_1, \cdots, t_n)$ with respect to θ and let $\hat{\theta}(M)$ denote the resulting value of θ. Then the profile likelihood for the parameter M is given by

$$L_p(M|t_1, \cdots, t_n) = L(\hat{\theta}(M), M|t_1, \cdots, t_n). \tag{3}$$

By the form of (2), for any given θ, $L(\theta, M|t_1, \cdots, t_n)$ is a non-increasing function in M, which suggests that M reaching its lower bound maximizes $L(\theta, M|t_1, \cdots, t_n)$ for any given θ, and thus, maximizes (3) as well. Thus, $\hat{M} = t_{(n)}$ is the maximum likelihood estimate (MLE) of M.

3.2 Relative Profile Likelihood

In reality, using an estimate of the maximum incubation time M_0 to be a quarantine time for the whole population is not completely sensible, because incubation times for a very small portion of patients can be extremely large while the majority of the infected cases may have incubation times under a small value. Consequently, we focus on identifying a practically feasible value M above which it is unlikely for the exposed individuals to develop symptoms of COVID-19. This problem can be approached by adopting the idea of the relative likelihood, which provides a convenient metric ranging between 0 and 1 to rank all parameter values according to their plausibilities in light of the data [Kalbfleisch, 1979].

Given the data $\{t_1, ..., t_n\}$ and the resultant MLEs of θ and M, denoted, respectively, $\hat{\theta}$ and $\hat{M}$, the relative plausibilities of values of θ and M may be assessed by comparing the likelihood of those values to the likelihood of the MLEs. In notation, the relative likelihood function of θ and M is defined as

$$\frac{L(\theta, M|t_1, \cdots, t_n)}{L(\hat{\theta}, \hat{M}|t_1, \cdots, t_n)}.$$

Since our focus is on parameter M, we propose to use the profile likelihood (3) and consider the relative profile likelihood (RPL), defined as

$$\text{RPL}(M) = \frac{L_p(M|t_1, \cdots, t_n)}{L_p(\hat{M}|t_1, \cdots, t_n)}.$$

For a pre-specified constant c with $0 \leq c < 1$, define

$$\mathscr{S}(c|t_1,\cdots,t_n) = \{M \geq M_1 : \mathrm{RPL}(M) \leq c\}. \tag{4}$$

The set $\mathscr{S}(c|t_1,\cdots,t_n)$ collects all the values of M such that the resulting profile likelihood is no bigger than c times the profile likelihood evaluated at the MLE $\hat{M}$. As $L_p(M|t_1,\cdots,t_n)$ is a non-increasing function of M, it is immediate that if $M^* \in \mathscr{S}(c|t_1,\cdots,t_n)$, then all values larger than M^* belong to $\mathscr{S}(c|t_1,\cdots,t_n)$. It is thereby of interest to identify the smallest value in $\mathscr{S}(c|t_1,\cdots,t_n)$ for a given c. Let $M_c = \min \mathscr{S}(c|t_1,\cdots,t_n)$. In the instance where the smallest value does not exist, we set M_c to be M_1. Apparently, $\mathscr{S}(c|t_1,\cdots,t_n)$ includes more values as c becomes bigger. By controlling the value of c, we are able to group those values of M so that the corresponding profile likelihood is upper bounded by a fraction of the maximum profile likelihood, which offers us a way to determine a suitable quarantine time. We hope to set a possibly smallest quarantine time so that the chance of an incubation time exceeding it is slim. To be conservative, we consider a small value of c, say $c = 0.1$, in the following analysis.

3.3 Determine Quarantine Time using Surrogate Incubation Times

A quick approach to analyzing of interval-censored incubation times described in Section 2.1 is to take the mid-point of the incubation interval for each case as the surrogate of the incubation time and then estimate the distribution of the incubation times accordingly. In this section, we take this approach and consider the five truncated parametric models described in Section 2.

Specifically, for $i = 1,\cdots,n$, let t_i^* be the midpoint of the interval $[t_{Li}, t_{Ui}]$, where $t_i^* = \frac{t_{Li}+t_{Ui}}{2}$. For the COVID-19 data we consider here, the midpoints of the incubation intervals $[t_{Li}, t_{Ui}]$ range from 1.5 to 21 days, with the mean 6.92 days and the standard deviation 3.75 days. We fit the surrogate data $\{t_i^* : i = 1,\cdots,n\}$ with each of the five distribution described in Section 2.2, and then estimate the model parameters using the method outlined in Section 3.1. To assess the performance of the model fit, we calculate the Akaike information criterion (AIC) and Schwarz's Bayesian information criterion (BIC) for each model, respectively, given by

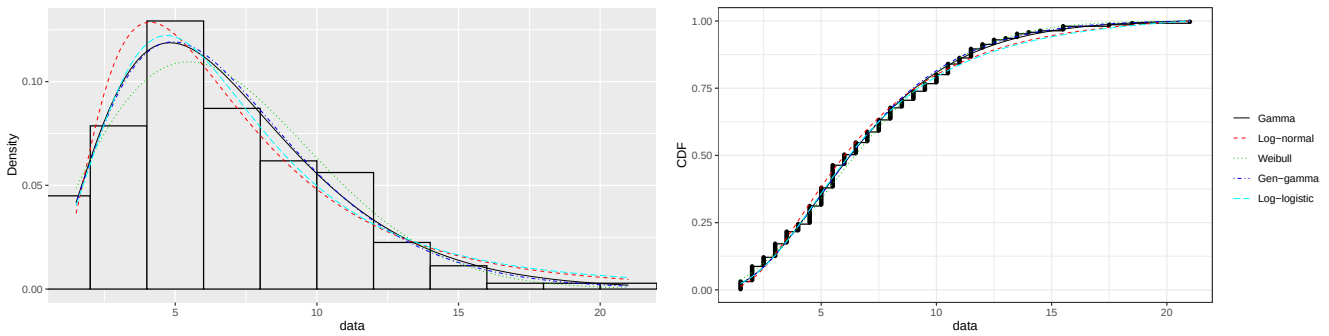
$$\begin{aligned} AIC &= -2\log L(\hat{\theta},\hat{M}|t_1^*,\cdots,t_n^*) + 2p; \\ BIC &= -2\log L(\hat{\theta},\hat{M}|t_1^*,\cdots,t_n^*) + p\log\, n, \end{aligned} \tag{5}$$

where p is the number of the model parameters. The results are reported in Table 2. The gamma model has the lowest AIC and BIC, while the Weibull and generalized gamma models have the highest AIC and BIC, respectively, though the differences are not large.

Table 2: Summary of the model fit for the surrogate data

Distribution	gamma	generalized gamma	log-normal	log-logistic	Weibull
AIC	945.30	948.88	947.88	949.78	950.86
BIC	951.66	958.42	954.24	956.15	957.22

In Figure 3a, we plot the pdf for five distributions discussed in Section 2.2 with the parameters replaced by their estimates, in contrast to the histogram for the data. Figure 3b shows the corresponding cumulative distribution function (CDF) of the five truncated models along with the empirical CDF of the data. While there are differences in these distributions, they fit the data generally well.



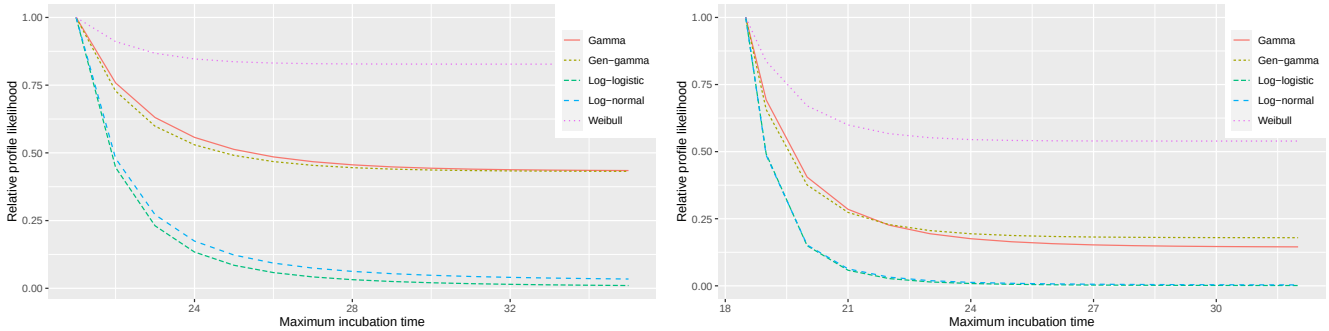
(a) Histogram and theoretical densities

(b) Empirical and theoretical CDFs

Figure 3: Two goodness-of-fit plots for the truncated distributions fitted to the surrogate data

Despite the fact that $\hat{M}$ is not affected by the distributional assumption of the incubation times, the shape of the profile likelihood of M, and therefore, a plausible quarantine time, depends on the distribution forms, especially the tails of the distributions [Farewell et al., 2005]. Figure 4a presents the RPL for M for the five considered models. The RPL for Weibull is nearly flat for M values greater than 24 and stay around 0.85, i.e., $RPL(M) \approx 0.85$ for any $M \geq 24$. This implies that any values of $M \geq 24$ do not decrease the corresponding profile likelihood to be below 0.85 of the profile likelihood at the MLE. Therefore, the truncated Weibull distribution does not provide us much information about the choice of a sensible quarantine time. For the truncated generalized gamma model, the RPL for M has a steeper change than that of the truncated Weibull model, but remains stable around 0.4 for M greater than 28. The RPL for truncated gamma model shows a similar trend to that of the truncated generalized gamma model. The RPLs

of the truncated log-normal and log-logistic models drops sharply towards 0, suggesting any values of $M \geq 26$ makes profile likelihood $L_p(M|t_1^*, \cdots, t_n^*)$ less than 10% of that evaluated at $\hat{M}$. Therefore, fitting the truncated log-logistic or log-normal model to the data suggests that a quarantine time of $M_c = 26$ days may be plausible if we regard $c = 0.1$ as a reasonable threshold.



(a) with the possible influential point

(b) without the possible influential point

Figure 4: RPLs for the surrogate data

Diving deep into the data, we find that the maximum observed surrogate incubation time, 21 days, comes from a 17 years old male patient. Speculating the possibility that this observation is influential outlier, we repeat the same analysis by removing it to see how the results may change; the new $\hat{M}$ becomes 18.5. The results are displayed in Figure 4b, showing that all the truncated distributions become more informative. In particular, the truncated log-logistic and log-normal models suggest 21 days as a reasonable quarantine time. These findings also demonstrate the considerable influence by potential outliers.

4 Analysis Methods with Interval-Censored Incubation Times

4.1 Estimation with Interval-Censored Data

In Section 3.3, using the midpoint of $[t_{Li}, t_{Ui}]$ as a surrogate of the true incubation time for subject i gives us quick analysis results. However, such a method may incur some bias because that the true incubation times may not be identical to the surrogate values. Furthermore, the method in Section 3.3 ignores the uncertainty induced from interval-censored data. In this section, we develop an inference method for interval-censored data.

Using the notation in Section 3, suppose the incubation times $\{t_1, \cdots, t_n\}$ for n study subjects are not directly observed, but instead, we observe a sequence of exposure windows $\{[t_{Li}, t_{Ui}], (i = 1, \cdots, n)\}$, where $t_i \in [t_{Li}, t_{Ui}]$ for $i = 1, \cdots, n$. Again, similar to the idea in Section 3.1, we consider the distribution of incubation times truncated by

M, a time point varying in $[M_1,\infty)$ where M_1 is a sufficiently large value so that all the exposure windows fall below M_1. Let I be the index set for the individuals with interval-censored observations, and let L be the index set for the individuals with left-censored observations (i.e., individuals with $t_{Li}=0$). Therefore, the likelihood function for the interval-censored data is given by

$$L(\theta,M|[t_{L1},t_{U1}],\cdots,[t_{Ln},t_{Un}]) = \frac{1}{\{F(M;\theta)\}^n}\left[\prod_{i\in I}\{F(t_{Ui};\theta)-F(t_{Li};\theta)\}\right]\left[\prod_{i\in L}F(t_{Ui};\theta)\right].$$

Maximizing the log-likelihood, $\log L(\theta,M|[t_{L1},t_{U1}],\cdots,[t_{Ln},t_{Un}])$, with respect to θ and M gives the MLE of θ and M, denoted $\hat{\theta}_{cens}$ and $\hat{M}_{cens}$, respectively.

Analogous to the discussion in Section 3.2, we consider the RPL function of M for interval-censored incubation times

$$\mathrm{RPL}(M) = \frac{L_p(M|[t_{L1},t_{U1}],\cdots,[t_{Ln},t_{Un}])}{L_p(\hat{M}_{cens}|[t_{L1},t_{U1}],\cdots,[t_{Ln},t_{Un}])}.$$

For a given c, the set

$$\mathscr{S}(c|[t_{L1},t_{U1}],\cdots,[t_{Ln},t_{Un}]) = \{M > M_1 : \mathrm{RPL}(M) \le c\} \tag{6}$$

is constructed to determine a plausible quarantine time, and we are interested in identifying

$$M_{cens} \triangleq \min \mathscr{S}(c|[t_{L1},t_{U1}],\cdots,[t_{Ln},t_{Un}])$$

if existing; otherwise, set M_{cens} to be M_1.

4.2 Determine Quarantine Time based on Interval-Censored Incubation Times

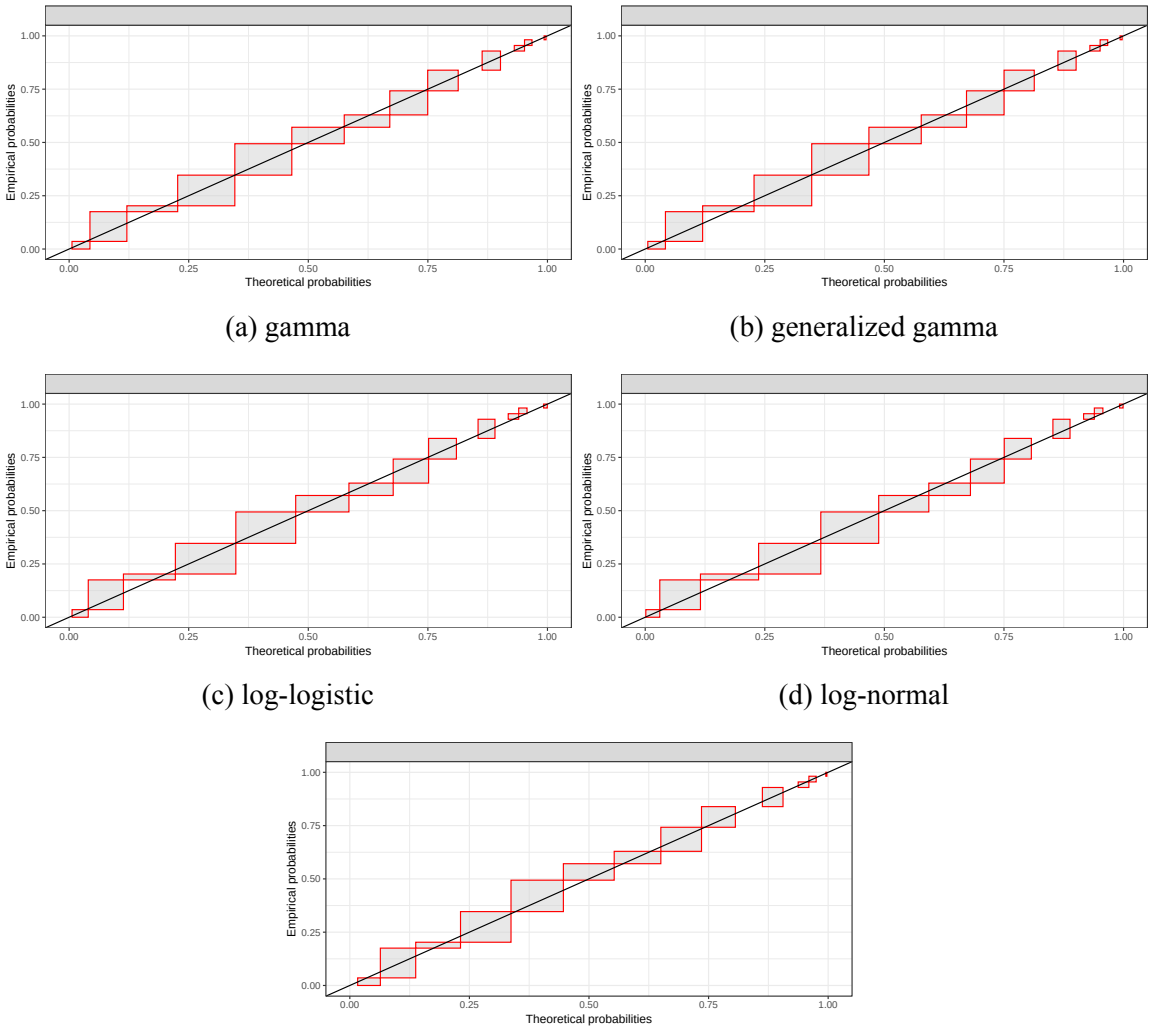
While using the midpoint of the incubation interval (t_{Li},t_{Ui}) in Section 3.3 allows us to easily model incubation times, this approach usually yields biased estimation of the incubation time distribution, which may tend to be right skewed [Virlogeux et al., 2016]. Here we use the development in Section 4.1 to conduct a more refined analysis by incorporating the feature of interval-censored data under the same models in Section 3.3.

To determine how well the models fit the data, we calculate the AIC and BIC of the fitted models using (5) with $L(\hat{\theta},\hat{M}|t_1^*,\cdots,t_n^*)$ replaced by $L(\hat{\theta}_{cens},\hat{M}_{cens}|[t_{L1},t_{U1}],\cdots,[t_{Ln},t_{Un}])$, and report the results in Table 3. Clearly, the gamma distribution shows the best fit in terms of both AIC and BIC values. Among other four distributions,

generalized gamma, log-normal, log-logistic and Weibull perform similarly in terms of AIC, whereas log-normal is slightly better than others if using the BIC. Using the R package *fitdistrplus* [Delignette-Muller et al., 2015], we produce probability-probability (P-P) plots for the models and report them in Figure 5, which shows similar patterns for the five models.

Table 3: Summary of the model fit for the interval-censored data

Distribution	gamma	generalized gamma	log-normal	log-logistic	Weibull
AIC	573.42	575.41	575.20	575.65	576.00
BIC	582.97	588.13	584.74	585.20	585.54



(a) gamma

(b) generalized gamma

(c) log-logistic

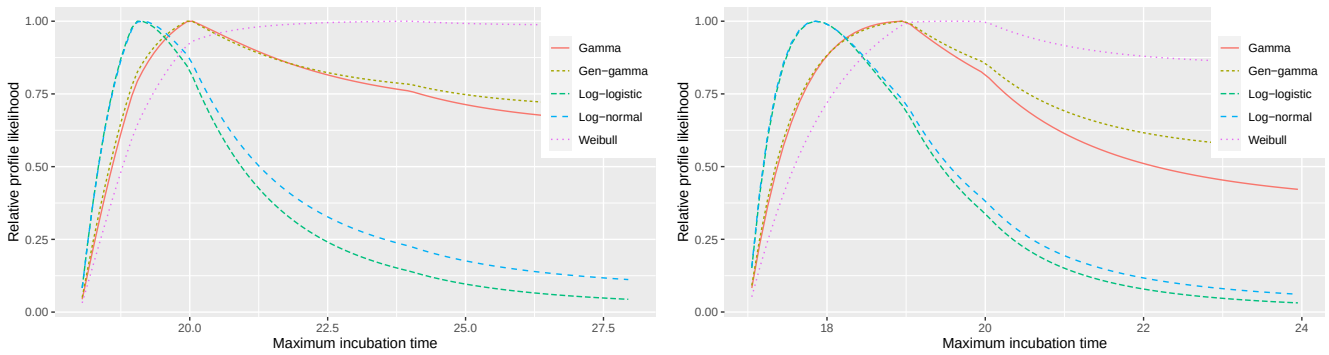
(d) log-normal

(e) Weibull

Figure 5: P-P plots for the truncated distributions fitted to the interval-censored data

In Figure 6a, we report the RPL for M derived from the five models assumed for the incubation times. The RPLs for log-normal and log-logistic distributions are maximized at a relatively same value of M. The RPL for Weibull is nearly flat around a value slightly smaller than 1 for $M \geq 22$, offering little information regarding a reasonable quarantine time. For the generalized gamma and gamma models, the RPL does not suggest any plausible quarantine time because they become flat around a large value (e.g., a value higher than 0.6). From the RPL for the log-logistic model, it is apparent that the profile likelihood is more than 10 times smaller for any value of $M \geq 26$ than that evaluated at $\hat{M}_{cens}$. This suggests that, under the log-logistic model, setting 26 days as a quarantine time may be necessary. The RPL under the log-normal model suggests a longer quarantine time than 26 days.

In contrast to excluding a possible outlier with the incubation interval of (18, 24) days in Section 3.3, we repeat the preceding analysis with this observation removed and plot the RPL in Figure 6b. It is clear that, the pattern of the RPL for M is consistent with that shown in Figure 6a. It can be seen that, the RPLs for the log-normal and the log-logistic models drop below 0.1 around $M = 22$, suggesting that 22 days may be a plausible quarantine time if potential outlier is removed in the analysis.



(a) with the possible influential point

(b) without the possible influential point

Figure 6: RPLs for the interval-censored data

5 Determine Quarantine time based on Percentile Estimation

Using the function RPL(M) offers us a convenient and intuitive way to determine a practical quarantine time, as illustrated in Sections 3 and 4. However, this approach has the limitation in its sensitivity to possible outliers. To alleviate the issue, we now alternatively employ the percentile estimation method to determine a quarantine time. A reasonable quarantine time may be specified as a time beyond which no large percentage of cases would develop symptoms.

Unlike the RPL methods discussed in Sections 3 and 4 which consider truncated distributions of incubation times, here we consider to directly use the five distributions in Section 2.2 to model incubation times of COVID-19. Taking the same data treatment as in Sections 3 and 4, we present two types of analysis in the following subsections using the percentile estimation method.

5.1 Analysis Using the Middle Points of the Incubation Intervals

Using the data $\{t_i^* : i = 1, \cdots, n\}$ considered in Section 3.3, we fit each of the five distributions described in Section 2.2. Specifically, we calculate the likelihood function

$$L(\theta | t_1^*, \cdots, t_n^*) = \prod_{i=1}^{n} f(t_i^*; \theta) \tag{7}$$

for the distribution function $f(\cdot)$, as described in Section 3.1. Then maximize (7) with respect to θ to obtain the MLE of θ. Figure 7 shows the resulting estimated density functions in contrast to a histogram of the surrogate data $\{t_i^* : i = 1, \cdots, n\}$, together with the estimated theoretical CDFs against the empirical CDFs. The five estimated distributions fit the data generally well, in fitting the right tail of the data, the gamma and generalized gamma distributions are slightly better than others. AIC and BIC values (not reported here) of the model fits, obtained from (5), indicate that, the gamma model results in the lowest AIC followed by the generalized gamma model.

Next, for each of the five estimated distributions, we calculate the 95% and 99% percentiles and obtain their corresponding 95% bootstrap confidence intervals. The results are reported in Table 4. The confidence intervals for 95% percentile under the gamma, generalized gamma and Weibull distributions suggest that the currently recommended quarantine time of 14 days is not long enough, because approximately 5% of infected patients may show symptoms after 14 days of quarantine. Assuming log-normal and log-logistic distributions for incubation times, a quarantine time of 14 days may release more than 5% of infected patients prior to the appearance of the symptoms. The 99% percentiles of the gamma, generalized gamma and Weibull models suggest that if we extend the quarantine time to be about 18 days, then only 1% of released individuals could be infected cases. On the contrary, the log-normal and log-logistic models require over 23 and 29 days, respectively, to reach this small percentage.

5.2 Analysis Using the Incubation Intervals

In contrast to surrogate incubation times $\{t_i^* : i = 1, \cdots, n\}$ considered in Section 5.1, we now turn our focus to the percentile estimation with incubation intervals. To be specific, we fit each of the distributions described in Section

2.2 to the incubation intervals $\{[t_{Li}, t_{Ui}], (i = 1, \cdots, n)\}$, shown in Figure 1. Modifying the discussion in Section 4.1, we calculate the likelihood function

$$L(\theta | [t_{L1}, t_{U1}], \cdots, [t_{Ln}, t_{Un}]) = \left[\prod_{i \in I} \{F(t_{Ui}; \theta) - F(t_{Li}; \theta)\}\right] \left[\prod_{i \in L} F(t_{Ui}; \theta)\right] \tag{8}$$

to obtain the MLE of θ associated with $f(\cdot)$. Figure 8 displays P-P plots of the estimated distributions, showing that those five distributions seem to provide fairly reasonable fit to the data though there are minor differences from distribution to distribution.

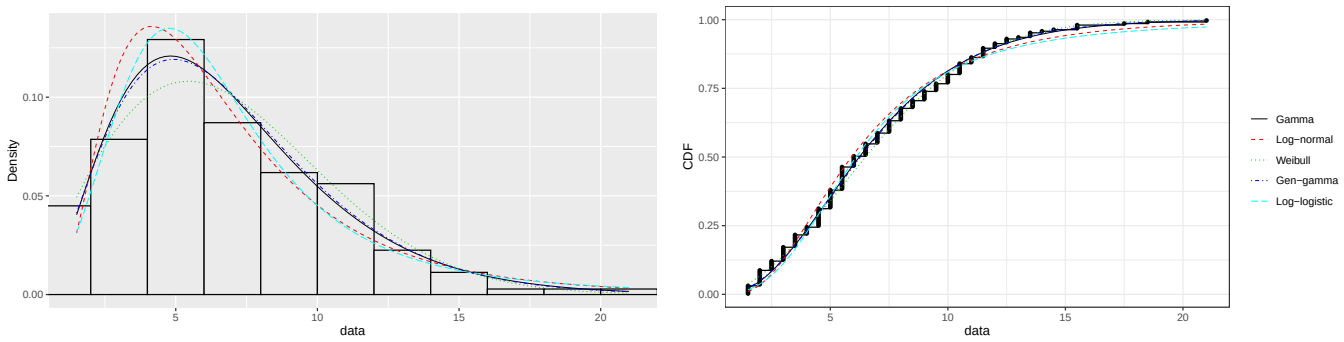
Next, for each of the estimated distributions, we calculate the 95% and 99% percentiles and obtain their corresponding 95% bootstrap confidence intervals, together with the length of each confidence interval. The results are reported in Table 5. The estimates of the percentiles obtained from the interval-censored data, reported in Table 5, are smaller than those obtained from the surrogate data, reported in Table 4. The lengths of the confidence intervals obtained from the different treatments of the data are fairly close, indicating the similar variability incurred in the estimation procedures.

Table 4: Estimates of 95% and 99% percentiles and their associated 95% bootstrap confidence intervals (in day), obtained from using the surrogate data $\{t_i^* : i = 1, \cdots, n\}$

Distribution	gamma	log-normal	generalized gamma	log-logistic	Weibull
95% percentile	14.15	15.65	14.03	16.73	13.71
95% C.I.	(12.90,15.42)	(14.07,17.23)	(12.73,15.41)	(14.73,18.63)	(12.53,14.94)
Length of C.I.	2.52	3.16	2.68	3.90	2.41
99% percentile	18.70	23.46	18.36	29.44	17.08
95% C.I.	(16.88,20.56)	(20.33,26.40)	(15.76,20.97)	(24.51,34.07)	(15.35,18.84)
Length of C.I.	3.68	6.07	5.21	9.56	3.49

Table 5: Estimates of 95% and 99% percentiles and their associated 95% bootstrap confidence intervals (in day), obtained from using the incubation intervals $\{[t_{Li}, t_{Ui}], (i = 1, \cdots, n)\}$

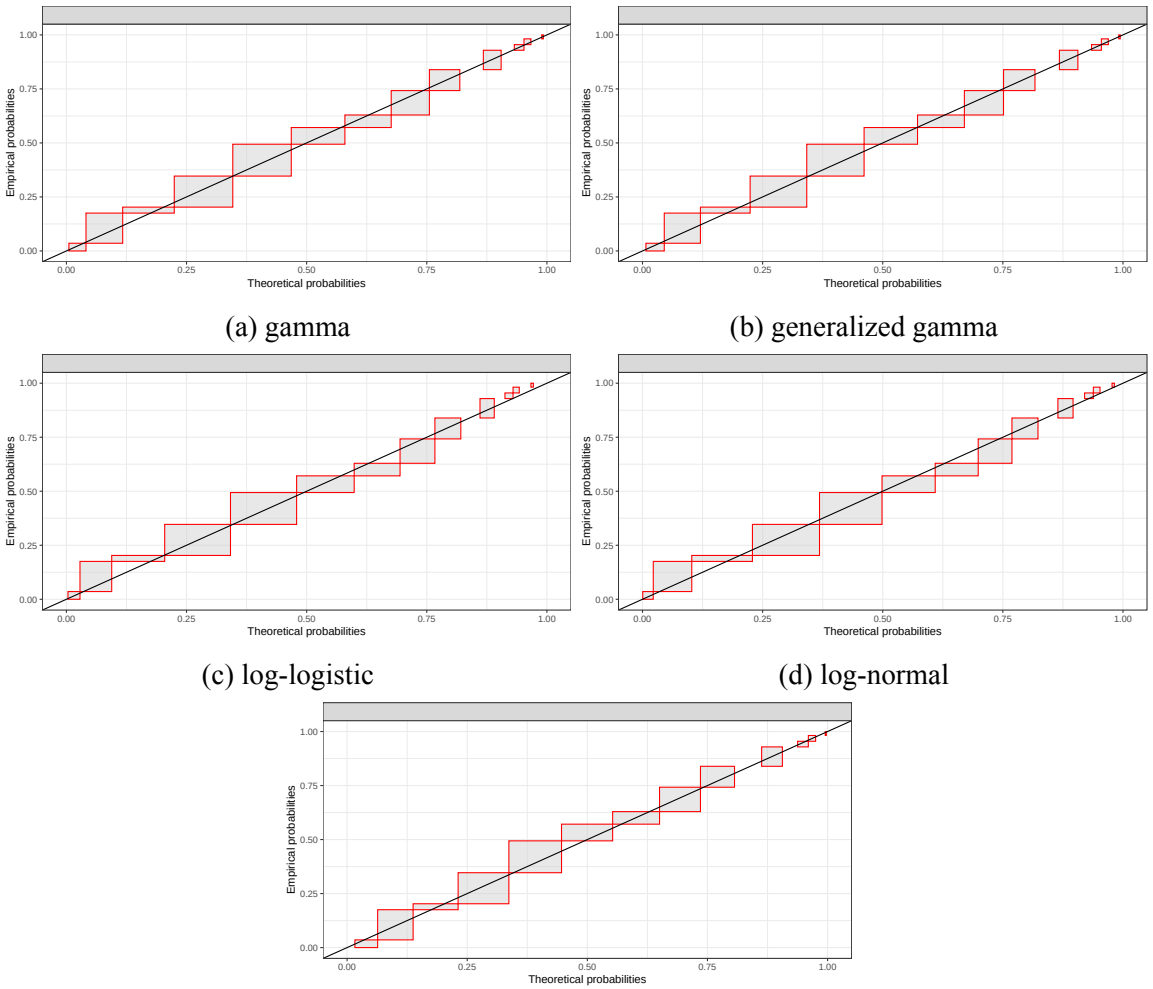
Distribution	gamma	log-normal	generalized gamma	log-logistic	Weibull
95% percentile	13.87	14.82	13.72	15.72	13.51
95% C.I.	(12.59,15.14)	(13.28,16.37)	(12.40,15.09)	(13.87,17.52)	(12.28,14.77)
Length of C.I.	2.55	3.09	2.69	3.65	2.49
99% percentile	18.20	21.55	17.71	26.58	16.75
95% C.I.	(16.31,20.12)	(18.49,24.33)	(15.17,20.20)	(21.88,30.79)	(14.98,18.54)
Length of C.I.	3.81	5.84	5.03	8.91	3.56



(a) Histogram and theoretical densities

(b) Empirical and theoretical CDFs

Figure 7: Two goodness-of-fit plots for the untruncated distributions fitted to the surrogate data



(a) gamma

(b) generalized gamma

(c) log-logistic

(d) log-normal

(e) Weibull

Figure 8: P-P plots for the untruncated distributions fitted to the interval-censored data

6 Discussion

In this paper, we analyze COVID-19 incubation times using a publicly available dataset and use different models to characterize the distribution of incubation times. Our findings suggest that the current recommended 14-day quarantine time is not long enough to control the probability of an early release of infected individuals to be small.

While the data are analyzed from multiple angles, we comment that certain aspects need further attention. Recall bias and reporting bias are likely to be present in the data, and de-biasing adjustments need to be employed in the inferential procedures to remove the bias. The data analyzed here are homogeneous in the sense that the study subjects come from the same city in a province of China. However, people with different races and demographic features may respond differently to infectious diseases. Therefore, the study sample must be representative of the whole population in order to set a more sensible quarantine time than the currently adopted 14-days. In most epidemiological studies of incubation times, the attention is focused on finding an ideal quarantine time merely to ensure infected individuals not to be released too early. However, this may not be the best criterion to reduce the pandemic impacts on the economy, social activities, and schooling, etc. Epidemiological-economic models can be carefully studied to set up an optimal quarantine time to take into account both the risk of early-release of cases and the impacts on the economy. Finally, we note that the size of the data analyzed is not large. With more data becoming available, it would be interesting to carry out a more comprehensive investigation to account for other possible risk factors such as age and gender.

Acknowledgements

YKC thanks Yuan Du for his initial interest in this project. The research is partially supported by funding from the Natural Sciences and Engineering Research Council of Canada (NSERC). Yi is Canada Research Chair in Data Science (Tier 1). Her research was undertaken, in part, thanks to funding from the Canada Research Chairs Program.

References

J. A. Backer, D. Klinkenberg, and J. Wallinga. Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from wuhan, China, 20–28 january 2020. *Eurosurveillance*, 25(5):2000062, 2020. doi: 10.2807/1560-7917.ES.2020.25.5.2000062.

P. Banka and C. M. Comiskey. The incubation period of COVID-19: A scoping review and meta-analysis to aid modelling and planning. 2020. doi: 10.1101/2020.10.20.20216143.

Y. K. Charvadeh and G. Y. Yi. Data visualization and descriptive analysis for understanding epidemiological characteristics of covid-19: A case study of a dataset from january 22, 2020 to march 29, 2020. *Journal of Data Science*, 18(3):526–535, 2020.

P. C. Cooley, L. E. Myers, and D. N. Hamill. A meta-analysis of estimates of the aids incubation distribution. *European Journal of Epidemiology*, 12(3):229–235, 1996.

B. J. Cowling, M. P. Muller, I. O. Wong, L.-M. Ho, M. Louie, A. McGeer, and G. M. Leung. Alternative methods of estimating an incubation distribution: examples from severe acute respiratory syndrome. *Epidemiology*, 18(2): 253–259, 2007.

C. Cox, H. Chu, M. F. Schneider, and A. Munoz. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in Medicine*, 26(23):4352–4374, 2007.

M. L. Delignette-Muller, C. Dutang, et al. fitdistrplus: An r package for fitting distributions. *Journal of Statistical Software*, 64(4):1–34, 2015.

V. T. Farewell, A. Herzberg, K. James, L. Ho, and G. Leung. SARS incubation and quarantine times: when is an exposed individual known to be disease free? *Statistics in Medicine*, 24(22):3431–3445, 2005.

W. He, G. Y. Yi, and Y. Zhu. Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for COVID-19: Meta-analysis and sensitivity analysis. *Journal of Medical Virology*, 92(11):2543–2550, 2020.

X. Jiang, Y. Niu, X. Li, L. Li, W. Cai, Y. Chen, B. Liao, and E. Wang. Is a 14-day quarantine period optimal for effectively controlling coronavirus disease 2019 (COVID-19)? 2020. doi: 10.1101/2020.03.15.20036533.

J. G. Kalbfleisch. *Probability and Statistical Inference*, volume 2. Springer, New York, 1979.

Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al. Early transmission dynamics in wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, 2020.

A. A. Lover, X. Zhao, Z. Gao, R. J. Coker, and A. R. Cook. The distribution of incubation and relapse times in experimental human infections with the malaria parasite plasmodium vivax. *BMC Infectious Diseases*, 14(1):539, 2014. doi: 10.1186/1471-2334-14-539.

C. McAloon, Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open*, 10(8), 2020. doi: 10.1136/bmjopen-2020-039652.

Public Health Ontario. COVID-19 incubation period and considerations for travellers' quarantine duration. "https://www.publichealthontario.ca/-/media/documents/ncov/main/2020/12/covid-19-incubation-travellers-quarantine-duration.pdf?la=en". Accessed: 2021-01-25.

J. Qin, C. You, Q. Lin, T. Hu, S. Yu, and X.-H. Zhou. Estimation of incubation period distribution of covid-19 using disease onset forward time: a novel cross-sectional and forward follow-up study. *Science Advances*, 6(33), 2020. doi: 10.1101/2020.03.06.20032417.

B. J. Quilty, S. Clifford, S. Flasche, A. J. Kucharski, W. J. Edmunds, C. C.-. W. Group, et al. Quarantine and testing strategies in contact tracing for sars-cov-2. 2020. doi: 10.1101/2020.08.21.20177808.

V. Virlogeux, J. Yang, V. J. Fang, L. Feng, T. K. Tsang, H. Jiang, P. Wu, J. Zheng, E. H. Lau, Y. Qin, et al. Association between the severity of influenza A (H7N9) virus infections and length of the incubation period. *PLoS One*, 11(2), 2016. doi: 10.1371/journal.pone.0148506.

WHO. Coronavirus disease (COVID-19) dashboard. "https://covid19.who.int/". Accessed: 2021-01-14.

L. Yang, J. Dai, J. Zhao, Y. Wang, P. Deng, and J. Wang. Estimation of incubation period and serial interval of COVID-19: analysis of 178 cases and 131 transmission chains in hubei province, China. *Epidemiology & Infection*, 148, 2020. doi: 10.1017/S0950268820001338.