Statistics in Biosciences (2022) 14:175–190
https://doi.org/10.1007/s12561-021-09320-8

Is 14-Days a Sensible Quarantine Length for COVID-19? Examinations of Some Associated Issues with a Case Study of COVID-19 Incubation Times

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

Received: 28 March 2021 / Revised: 16 August 2021 / Accepted: 22 August 2021 /
Published online: 9 September 2021

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, 14-days is commonly taken as a quarantine time to curb the virus spread in balancing the impacts of COVID-19 on diverse aspects of the society, including public health, economy, and humanity perspectives, etc. However, setting a sensible quarantine time is not trivial and it depends on various underlying factors. In this article, we take an angle of examining the distribution of the COVID-19 incubation time using likelihood-based methods. Our study is carried out on 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 describe incubation times of COVID-19. Our findings suggest that statistically, the 14-day quarantine time may not be long enough to control the probability of an early release of infected individuals to be small. While the size of the study data is not large enough to offer us a definitely acceptable quarantine time, and further in practice, the decision-makers may take account of other factors related to social and economic concerns to set up a practically acceptable quarantine time, our study demonstrates useful methods to determine a reasonable quarantine time from a statistical standpoint. Further, it reveals some associated complexity for fully understanding the COVID-19 incubation time distribution.

Keywords COVID-19 · Incubation times · Profile likelihood · Quantile estimation · Quarantine time

✉ Wenqing He
whe@stats.uwo.ca

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

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

1 Introduction

As of January 2021, more than 90 million confirmed COVID-19 cases and 1.9 million resulting deaths had been reported in more than 200 countries and regions [19]. To control the virus spread, the World Health Organization (WHO) suggested a 14-day quarantine time for potential COVID-19 infected cases [18]. The 14-day quarantine 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 started, many authors have studied the incubation information. For example, Li et al. [12] studied the incubation time using the data of 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. [1] 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 [3] 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 showed no signs of illness at their time of departure from Wuhan city, China, Qin et al. [16] 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: 7.67–8.90 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. [9] 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 4.77–5.39 days. Banka and Comiskey [2] studied the incubation period distribution for COVID-19 and highlighted the need for a longer quarantine time than the initially suggested 14-days by the WHO. Quilty et al. [17] 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%. Jiang et al. [10] 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 as well as the considerations of balancing the COVID-19 impacts on the other aspects of the society such as the economy and humanity concerns. Public health officials often use the tail end of the incubation range to determine a quarantine time [15]. A good understanding of the distribution of incubation times thereby becomes critical. 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 imperative to broadly examine the distribution of COVID-19 incubation times from different angles in order to better understand 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 COVID-19 incubation times. Our explorations are carried out on a dataset of 178 COVID-19 cases collected between January 20, 2020 and February 29, 2020, with the information of exposure periods and dates of symptom onset available for the study subjects. Our findings suggest that the current practice of setting 14-days as a quarantine time may not be long enough to control the probability of an early release of infected individuals to be small. Though it may be premature to recommend a sensible quarantine time based on this small-sized data we study, the explorations here offer us methods to examine the COVID-19 incubation time distribution, and they also demonstrate associated complexity in this process.

The remainder of the article is organized as follows. In Sect. 2 we describe the data and outline useful distributions for describing incubation times of an infectious disease. We present the methodology for inference and discuss modelling for the case with observed incubation times or interval-censored incubation times in Sects. 3 and 4. We investigate different analyses in the hope of revealing broad scenarios for assessing whether the period of 14 days is a sensible quarantine time for COVID-19. In Sect. 5 we examine the percentiles of the incubation times to see how a sensible quarantinetime may be considered. We conclude the manuscript with a 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 [20]. 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 an epidemiological survey. The date of exposure refers to either (a) entry and exit dates of 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 Fig. 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 S1 in the Supplementary Material reports t_{Li}, and t_{Ui} for each of the 178 cases, and Fig. S1 in Supplementary Material illustrates the interval incubation times by gender and three different age groups.

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 incubation times of SARS [5]. 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 COVID-19 incubation times (e.g., [14]).

Let T denote a continuous non-negative random variable representing the incubation time of a patient 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.

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 $g(z) = \exp\{z - \exp(z)\}$, one obtains 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

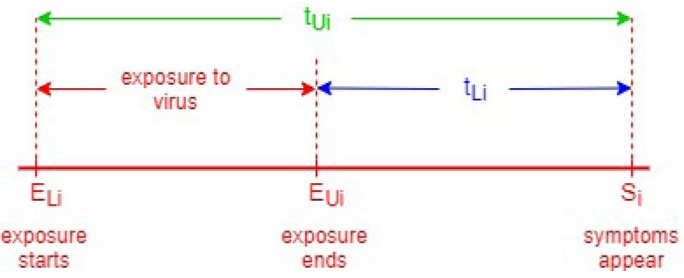


Fig. 1 Timeline of the incubation period

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

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

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 [6].

The log-logistic distribution is another choice for modeling incubation times [4, 13]. 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 Sect. 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 analyses, 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 such as those outlined in Sect. 2.2, and θ is the parameter whose value is unknown. Let $F(\cdot;\theta)$ denote the corresponding cumulative distribution function (CDF) of T. Let M_0 denote the *maximum* incubation time of *all* the individuals in the target population, thus, implicitly, we have $T \le M_0$. Suppose that we have a sample of observed incubation times, $\{t_1, \ldots, t_n\}$, for n randomly selected COVID-19 patients.

One may be interested in estimating M_0 using the sample information. To this end, we introduce a generic parameter M which represents a sufficiently large time so that it is unlikely for an observed incubation time to exceed it, and M_0 is regarded as the true value of the parameter M. For the clarity of exposition, we consider that the parameter M varies in the parameter space indicated by $[M_1, \infty)$, where M_1 is a specified finite positive value. For instance, setting M_1 to be 1 day reflects that incubation times are longer than one day.

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

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

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

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

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

Estimation of θ and M may be carried out using the likelihood method by maximizing $L(\theta, M|t_1, \ldots, t_n)$ 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. [8], we maximize $L(\theta, M|t_1, \ldots, t_n)$ 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

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

By the form of (2), for any given θ, $L(\theta, M|t_1, \ldots, t_n)$ is a non-increasing function in M, which suggests that M reaching its lower bound maximizes $L(\theta, M|t_1, \ldots, t_n)$ 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 a lot shorter incubation times. 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 adapting the idea of the relative likelihood, which provides a convenient metric ranging between 0 and 1 to rank all parameter values according to their plausibility in light of the data [11].

Given the data $\{t_1, \ldots, t_n\}$ and the resultant MLEs of θ and M, denoted, respectively, $\hat{\theta}$ and $\hat{M}$, the relative plausibility 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

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

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

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

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

$$\mathcal{S}(c|t_1, \ldots, t_n) = \left\{M \ge M_1 \,:\, \mathrm{RPL}(M) \le c\right\}. \tag{4}$$

The set $\mathcal{S}(c|t_1, \ldots, t_n)$ collects those values of M that are *implausible* in the sense that they yield small values of the profile likelihood, with the resulting profile likelihood no bigger than $100c\%$ of the profile likelihood evaluated at the MLE $\hat{M}$. As $L_p(M|t_1, \ldots, t_n)$ is a non-increasing function of M, it is immediate that if $M^* \in \mathcal{S}(c|t_1, \ldots, t_n)$, then all values larger than M^* belong to $\mathcal{S}(c|t_1, \ldots, t_n)$. It is thereby of interest to identify the smallest value in $\mathcal{S}(c|t_1, \ldots, t_n)$ for a given c, denoted $M_c = \min \mathcal{S}(c|t_1, \ldots, t_n)$. In the instance where the smallest value does not exist, we set M_c to be M_1. Apparently, the set $\mathcal{S}(c|t_1, \ldots, t_n)$ tends to become larger as c becomes bigger, and hence, $M_{c_1} \le M_{c_2}$ if $c_1 \ge c_2$. One needs, however, to note that $\mathcal{S}(c|t_1, \ldots, t_n)$ can be empty when c is smaller than a certain value.

We hope to set a possibly small quarantine time so that the chance of an incubation time exceeding it is slim. By controlling the value of c, we are able to collect

those implausible values of M so that the corresponding profile likelihood is upper bounded by a fraction of the maximum profile likelihood. In principle, the smaller we set a value for c, the more conservative it may yield a quarantine time. In spirit similar to the discussion about the likelihood region by Kalbeisch [11, p. 18], one may consider taking different threshold values for c to describe the varying degree of the plausibility for a value of the parameter M. While there is no gold standard for the choice of c, it is sensible to treat the values in $\mathcal{S}(c|t_1, \ldots, t_n)$ to be very implausible if c is close to 0. For instance, we may treat the M values in $\mathcal{S}(0.1|t_1, \ldots, t_n)$ as "implausible" and those in $\mathcal{S}(0.01|t_1, \ldots, t_n)$ as "very implausible" (if those sets are non-empty).

3.3 Determination of Quarantine Time Using Surrogate Incubation Times

A quick approach to analyzing interval-censored incubation times described in Sect. 2.1 is to take the midpoint 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 Sect. 3.2 in combination with the discussion in Sect. 2.2

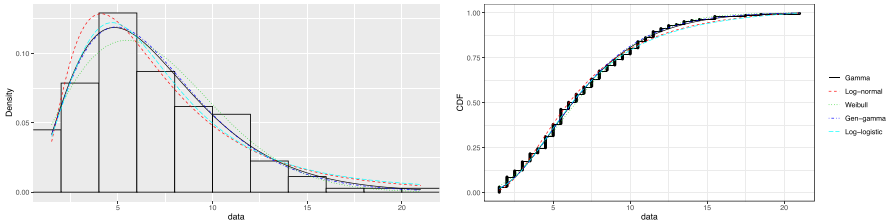
For $i = 1, \ldots, n$, let t_i^* be the midpoint of the interval $[t_{Li}, t_{Ui}]$, where $t_i^* = \frac{t_{Li}+t_{Ui}}{2}$. For the COVID-19 data we consider here, the midpoints of the incubation intervals $[t_{Li}, t_{Ui}]$ range from 1.5 to 21 days, with the mean 6.92 days and the standard deviation 3.75 days. We fit the surrogate data $\{t_i^* : i = 1, \ldots, n\}$ with each of the five distribution described in Sect. 2.2, and then estimate the model parameters using the method outlined in Sect. 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

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

where q is the number of the model parameters. The results are reported in Table 1. 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 1 Summary of the model fit for the surrogate data

Fit	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

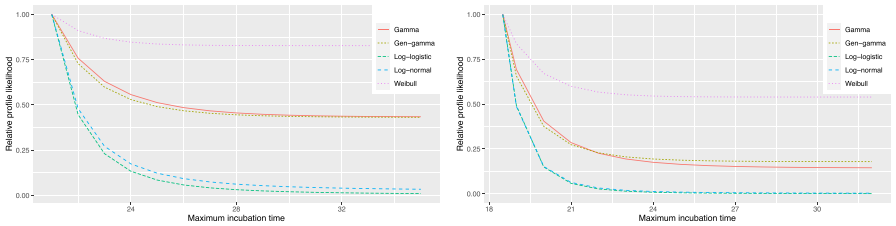


(a) Histogram and fitted densities **(b)** Empirical and fitted CDFs

Fig. 2 Two goodness-of-fit plots for the truncated distributions fitted to the surrogate data

In Fig. 2a, we plot the pdf for five distributions discussed in Sect. 2.2 with the parameters replaced by their estimates, in contrast to the histogram for the data. Figure 2b shows the corresponding CDFs 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.

Despite the fact that $\hat{M}$, being 21 days, 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 [8]. Figure 3a 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., $\text{RPL}(M) \approx 0.85$ for any $M \geq 24$. This implies that any values of $M \geq 24$ do not decrease the corresponding profile likelihood to be greatly lower than 0.85 of the profile likelihood at the MLE. Therefore, the truncated Weibull distribution does not provide much information about the choice of a quarantine time larger then 24. 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 0.40 for M greater than 28. The RPL for the 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 drop sharply towards 0, suggesting that any values of $M \geq 26$ make the profile likelihood $L_p(M|t_1^*, \ldots, t_n^*)$ 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 threshold for identifying implausible values for M.



(a) with the possible influential point **(b)** without the possible influential point

Fig. 3 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 merely on outlier, we repeat the same analysis by removing it to see how the results may change; the new $\hat{M}$ becomes 18.5 days. The results are displayed in Fig. 3b, 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 of potential outliers.

4 Analysis Methods with Interval-Censored Incubation Times

4.1 Estimation with Interval-Censored Data

In Sect. 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 the true incubation times may not be identical to the surrogate values. Furthermore, the method in Sect. 3.3 ignores the uncertainty induced from interval-censored data. In this section, we describe an inference method for interval-censored data.

Using the notation in Sect. 3, suppose the incubation times $\{t_1, \ldots, t_n\}$ for n study subjects are not directly observed, but instead, we observe a sequence of exposure windows $\left\{[t_{Li}, t_{Ui}] : i = 1, \ldots, n\right\}$, where $t_i \in [t_{Li}, t_{Ui}]$ for $i = 1, \ldots, n$. Again, similar to the idea in Sect. 3.1, we consider the distribution of incubation times truncated by M, a time point varying in $[M_1, \infty)$. 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

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

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

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

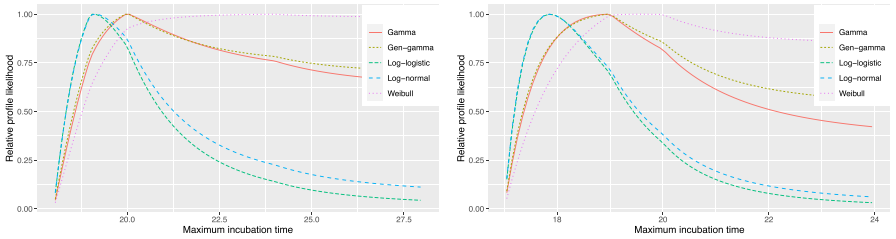
$$\text{RPL}(M) = \frac{L_p(M|[t_{L1}, t_{U1}], \ldots, [t_{Ln}, t_{Un}])}{L_p(\hat{M}_{\text{cens}}|[t_{L1}, t_{U1}], \ldots, [t_{Ln}, t_{Un}])}.$$

For a given c, the set

$$\mathcal{S}(c|[t_{L1}, t_{U1}], \ldots, [t_{Ln}, t_{Un}]) = \left\{M \geq M_1 : \text{RPL}(M) \leq c\right\} \tag{6}$$

Table 2 Summary of the model fit for the interval-censored data

Fit	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) with the possible influential point **(b)** without the possible influential point

Fig. 4 RPLs for the interval-censored data

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

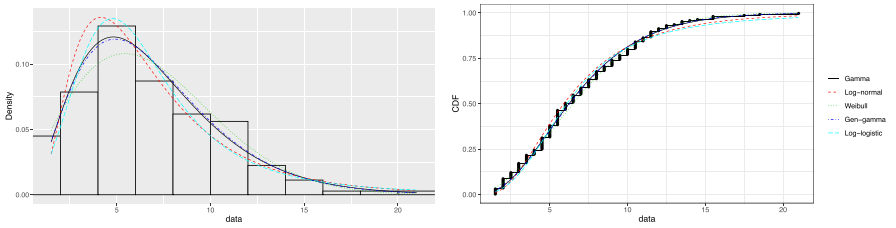
$$M_{\text{cens}} \triangleq \min \mathcal{S}(c|[t_{L1}, t_{U1}], \ldots, [t_{Ln}, t_{Un}])$$

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

4.2 Determination of Quarantine Time Using Interval-Censored Incubation Times

Here we use the development in Sect. 4.1 to analyze interval-censored data under the same models in Sect. 3.3. To determine how well the models fit the data, we calculate the AIC and BIC of the fitted models using (5) with $L(\hat{\theta}, \hat{M}|t_1^*, \ldots, t_n^*)$ replaced by $L(\hat{\theta}_{\text{cens}}, \hat{M}_{\text{cens}}|[t_{L1}, t_{U1}], \ldots, [t_{Ln}, t_{Un}])$, and report the results in Table 2. 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 BIC. Using the R package *fitdistrplus* [7], we produce probability–probability (P–P) plots for the models and report them in Fig. S2 in Supplementary Material, which shows similar patterns for the five models.

In Fig. 4a, we report the RPL for M derived from the five models assumed for the incubation times. The RPLs for the log-normal and log-logistic distributions are maximized at a relatively same value of M. The RPL for the Weibull is nearly flat, taking 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



(a) Histogram and fitted densities **(b)** Empirical and fitted CDFs

Fig. 5 Two goodness-of-fit plots for the untruncated distributions fitted to the surrogate data

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 $\hat{M}_{\text{cens}}$. 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 $(18, 24)$ in Sect. 3.3, we repeat the preceding analysis with this observation removed and plot the RPL in Fig. 4b. It is clear that the pattern of the RPL for M is consistent with that shown in Fig. 4a. 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 a potential outlier is removed in the analysis.

5 Determination of 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 Sects. 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 Sects. 3 and 4 which consider truncated distributions of incubation times, here we consider directly using the five distributions in Sect. 2.2 to model incubation times of COVID-19. Taking the same data treatment as in Sects. 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 $\{t_i^* : i = 1, \ldots, n\}$ considered in Sect. 3.3, we fit each of the five distributions described in Sect. 2.2. Specifically, we calculate the likelihood function

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

for the distribution function $f(\cdot)$, as described in Sect. 3.1. Then we maximize (7) with respect to θ to obtain the MLE of θ. Figure 5 shows the resulting estimated density functions in contrast to a histogram of the surrogate data $\{t_i^* : i = 1, \ldots, n\}$, together with the estimated CDFs against the empirical CDF. 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 a form modifying (5), indicate that, the Gamma model results in the lowest AIC followed by that for 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, where 10,000 bootstrap samples are considered. The results are reported in Table 3. 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 the 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.

Table 3 Estimates of 95% and 99% percentiles and their associated 95% bootstrap confidence intervals (in day)

Data		Gamma	Log-normal	Generalized Gamma	Log-logistic	Weibull
t_i^*	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
$[t_{Li}, t_{Ui}]$	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

5.2 Analysis Using the Incubation Intervals

We now turn our focus to the percentile estimation with incubation intervals. To be specific, we fit each of the distributions described in Sect. 2.2 to the incubation intervals $\left\{[t_{Li}, t_{Ui}] : i = 1, \ldots, n\right\}$. Modifying the discussion in Sect. 4.1, we calculate the likelihood function

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

to obtain the MLE of θ associated with $f(\cdot)$. Figure S3 in the Supplementary Material 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, where 10,000 bootstrap samples are considered. The results are reported in Table 3. The estimates of the percentiles obtained from the interval-censored data are smaller than those obtained from the surrogate data. 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.

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 currently 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 China. However, people of different races and demographic features may respond differently to infectious diseases. As pointed out by a referee, disease incubation is just one component related to the disease transmission process. In the case where the disease infectivity is negatively correlated with the length of the incubation period, aiming to quarantine the full length of incubation may not be reasonable. The incubation period may have a close association with many factors such as age, gender, chronic health diseases, family health history, and so on. We also note that the size of the data analyzed is not large. Having a large and representative study sample of the whole population is desirable

to determine a sensible quarantine time by facilitating the underlying heterogeneities of the population.

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. To set the optimal quarantine time, epidemiological-economic models can be carefully studied to take into account both the risk of early release of cases and the impacts on the economy.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12561-021-09320-8.

Acknowledgements The research is partially supported by funding from the Discovery Grant Program and the Emerging Infectious Disease Modeling Program of 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.

Data Availability Dataset can be found on the website https://zenodo.org/record/3898225.

Code availability Available upon request.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest/competing interest.

References

1. Backer JA, Klinkenberg D, Wallinga J (2020) Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. Euro Surveill 25(5):2000062. https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062
2. Banka P, Comiskey CM (2020) The incubation period of COVID-19: a scoping review and meta-analysis to aid modelling and planning. medRxiv https://doi.org/10.1101/2020.10.20.20216143
3. Charvadeh YK, Yi GY (2020) 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. J Data Sci 18(3):526–535
4. Cooley PC, Myers LE, Hamill DN (1996) A meta-analysis of estimates of the aids incubation distribution. Eur J Epidemiol 12(3):229–235
5. Cowling BJ, Muller MP, Wong IO, Ho LM, Louie M, McGeer A, Leung GM (2007) Alternative methods of estimating an incubation distribution: examples from severe acute respiratory syndrome. Epidemiology 18(2):253–259
6. Cox C, Chu H, Schneider MF, Munoz A (2007) Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Stat Med 26(23):4352–4374
7. Delignette-Muller ML, Dutang C (2015) fitdistrplus: an R package for fitting distributions. J Stat Softw 64(4):1–34
8. Farewell VT, Herzberg A, James K, Ho L, Leung G (2005) SARS incubation and quarantine times: when is an exposed individual known to be disease free? Stat Med 24(22):3431–3445
9. He W, Yi GY, Zhu Y (2020) Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for COVID-19: meta-analysis and sensitivity analysis. J Med Virol 92(11):2543–2550

10. Jiang X, Niu Y, Li X, Li L, Cai W, Chen Y, Liao B, Wang E (2020) Is a 14-day quarantine period optimal for effectively controlling coronavirus disease 2019 (COVID-19)? medRxiv https://doi.org/10.1101/2020.03.15.20036533
11. Kalbfleisch JG (1985) Probability and Statistical Inference, vol 2. Springer, Berlin
12. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY et al (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 382(13):1199–1207
13. Lover AA, Zhao X, Gao Z, Coker RJ, Cook AR (2014) The distribution of incubation and relapse times in experimental human infections with the malaria parasite plasmodium vivax. BMC Infect Dis 14(1):539. https://doi.org/10.1186/1471-2334-14-539
14. McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F, Casey M, Griffin J, Lane E, McEvoy D et al (2020) Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. BMJ Open 10(8):e039652. https://doi.org/10.1136/bmjopen-2020-039652
15. Public Health Ontario (2021) 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 25 Jan 2021
16. Qin J, You C, Lin Q, Hu T, Yu S, Zhou XH (2020) Estimation of incubation period distribution of covid-19 using disease onset forward time: a novel cross-sectional and forward follow-up study. Sci Adv 6(33). https://doi.org/10.1126/sciadv.abc1202
17. Quilty BJ, Clifford S, Flasche S, Kucharski AJ, Edmunds WJ, Group CCW, et al. (2021) Quarantine and testing strategies in contact tracing for SARS-CoV-2: a modelling study. Lancet Public Health https://doi.org/10.1016/S2468-2667(20)30308-X
18. WHO (2021a) Considerations for quarantine of contacts of COVID-19 cases. https://www.who.int/publications/i/item/WHO-2019-nCoV-IHR-Quarantine-2021.1. Accessed 07 Aug 2021
19. WHO (2021b) Coronavirus disease (COVID-19) dashboard. https://covid19.who.int/. Accessed 01 Jan 2021
20. Yang L, Dai J, Zhao J, Wang Y, Deng P, Wang J (2020) Estimation of incubation period and serial interval of COVID-19: analysis of 178 cases and 131 transmission chains in Hubei province, China. Epidemiol Infect 148. https://doi.org/10.1017/S0950268820001338