



Research article

The doubling time analysis for modified infectious disease Richards model with applications to COVID-19 pandemic

Alexandra Smirnova¹, Brian Pidgeon¹, Gerardo Chowell² and Yichuan Zhao^{1,*}

¹ Department of Mathematics & Statistics, Georgia State University, 25 Park Place, Atlanta, GA 30303, USA

² Department of Population Health Sciences, Georgia State University, 140 Decatur St SE, Atlanta, GA 30303, USA

* **Correspondence:** Email: asmirnova@gsu.edu; Tel: +14044136409; Fax: +14044136403.

Abstract: In the absence of reliable information about transmission mechanisms for emerging infectious diseases, simple phenomenological models could provide a starting point to assess the potential outcomes of unfolding public health emergencies, particularly when the epidemiological characteristics of the disease are poorly understood or subject to substantial uncertainty. In this study, we employ the modified Richards model to analyze the growth of an epidemic in terms of 1) the number of times cumulative cases double until the epidemic peaks and 2) the rate at which the intervals between consecutive doubling times increase during the early ascending stage of the outbreak. Our theoretical analysis of doubling times is combined with rigorous numerical simulations and uncertainty quantification using synthetic and real data for COVID-19 pandemic. The doubling-time approach allows to employ early epidemic data to differentiate between the most dangerous threats, which double in size many times over the intervals that are nearly invariant, and the least transmissible diseases, which double in size only a few times with doubling periods rapidly growing.

Keywords: epidemiology; doubling times; Richards model; incidence cases

1. Introduction

The modified Richards model describes epidemic dynamics in two phases of fast and slow infection spread with a transition point (also referred to as turning or inflection point), at which the maximum incidence occurs. In the slow phase of infection spread (after the turning point), the epidemic peaks and subsequently declines, while the cumulative number of cases eventually saturates. At this point, little can be done in terms of control and prevention. On the other hand, efficient mitigation measures implemented at the early ascending stage of the outbreak have the potential of significantly reducing

humanitarian cost and economic impacts of the emerging disease.

In this study, we employ the modified Richards model to analyze the growth of epidemic in terms of 1) the number of times cumulative cases double until the epidemic peaks and 2) the rate at which the intervals between consecutive doubling times increase during the emerging stage of the disease. It is our expectation that the doubling-time research will become a major tool in the study of infectious disease epidemics, and potentially in understanding the underlying processes by which emerging outbreaks quantify their spread in distinct ways. Furthermore, we believe that the doubling-time metrics for the modified Richards model could be used to systematically characterize recurrent epidemics (e.g., multiple influenza seasons) or spatial heterogeneity for epidemics that can be disaggregated at finer spatial resolutions (e.g., regions, districts).

In what follows, we combine theoretical evaluation of doubling times with rigorous numerical experiments using synthetic and real epidemiological observations stemming from the well-documented COVID-19 pandemic. These observations are crucial to test, guide and improve epidemic modeling. This data driven enhancement of the modified Richards model leads to substantial improvements in our ability to capture epidemic trajectories and to a better characterization of the transmission potential of infectious diseases. Importantly, our results underscore how various underlying assumptions in epidemic models influence the doubling-time mathematical analysis and its numerical estimation. More generally, our study highlights the importance of confronting crucial modeling assumptions with real epidemic data. These assumptions undoubtedly play a key role in projecting possible outcomes of an outbreak, and estimating the intensity of intervention steps required for minimizing its impact on the population. Understanding the complete effect of parameters quantifying these assumptions and their variations on epidemic doubling times will require further methodological advancements.

2. Derivation of doubling times for modified Richards model in terms of cumulative number of cases

We analyze the growth of epidemics as a function of the number of times the cumulative incidence doubles before the epidemic peaks, and the rate at which the doubling time increases (hence the epidemic slows down), owing to interventions, behavior changes and depletion of susceptible individuals. The most difficult to control are epidemic threats that double in size many times by the time the peak has been reached, while doubling times increase at a relatively low rate. Conversely, the least difficult to control are epidemic threats that only double in size very few times, while the sequence of doubling times rapidly increases [1]. When $C'(t)$ represents the number of new infected cases at time t , the progression of the epidemic is often modeled by an autonomous differential equation tracking the cumulative number of infections, $C(t)$:

$$\frac{dC}{dt} = F(C). \quad (2.1)$$

Given this equation, the doubling time, Δt , can be expressed in terms of the cumulative number of cases as follows

$$\Delta t = \int_{C(t)}^{2C(t)} \frac{ds}{F(s)}. \quad (2.2)$$

Epidemics that span over a short period of time relative to the average lifetime of the host population can be routinely described by S-shaped logistic curves [2–6]. These differential equations rely on

a small number of parameters and assume initial exponential or sub-exponential growth phase that saturates as the number of cases accumulates, a pattern that implicitly captures a gradual decrease in at-risk susceptible population. In particular, the modified Richards model with four parameters has been used to fit a range of logistic-type epidemic curves [6–10]. For this model,

$$F(s) := rs^p \left[1 - \left(\frac{s}{K} \right)^a \right], \quad 0 < s < K, \quad (2.3)$$

where $r > 0$ represents the intrinsic growth rate in the absence of any limitation to disease spread, p , $0 \leq p \leq 1$, is a deceleration of growth parameter, $K > 0$ is the size of the epidemic, and $a > 0$ is a parameter that measures the extent of deviation from the S-shaped dynamics of the classical logistic growth model. At the early stages of the epidemic, this model enables us to capture different growth profiles ranging from constant incidence ($p = 0$), polynomial growth ($0 < p < 1$), to exponential growth ($p = 1$). As shown in [11], the solution of the indefinite integral in (2.2) for $F(s)$ defined in (2.3) can be expressed in terms of Gaussian Hypergeometric Function, which indicates that solution to (2.2)-(2.3) will converge under mild conditions. To derive the solution, let $u := (\frac{s}{K})^a$ be a new variable. Then the doubling time, Δt , for the modified Richards model (2.1), (2.3) is given by

$$\Delta t = \frac{K^{1-p}}{ar} \int_{(\frac{C}{K})^a}^{(\frac{2C}{K})^a} \frac{du}{u^{1-\frac{1-p}{a}}(1-u)} = b \int_{(\frac{C}{K})^a}^{(\frac{2C}{K})^a} \frac{du}{u^\delta(1-u)}, \quad (2.4)$$

where $\delta := 1 - \frac{1-p}{a}$ and $b := \frac{K^{1-p}}{ar}$. Since the cumulative number of cases cannot exceed the threshold level, K , the new variable, u , is less than 1. Thus, one can view $\frac{1}{1-u}$ as the sum of a geometric series and, therefore,

$$\Delta t = b \int_{(\frac{C}{K})^a}^{(\frac{2C}{K})^a} \sum_{n=0}^{\infty} u^{n-\delta} du = b \sum_{n=0}^{\infty} \frac{\left(\frac{2C}{K}\right)^{a(n+1-\delta)} - \left(\frac{C}{K}\right)^{a(n+1-\delta)}}{n+1-\delta}. \quad (2.5)$$

From (2.5) and the definitions of b and δ , one concludes that

$$\Delta t = \frac{K^{1-p}}{r} \sum_{n=0}^{\infty} \frac{\left(\frac{2C}{K}\right)^{na+1-p} - \left(\frac{C}{K}\right)^{na+1-p}}{na+1-p} = \frac{C^{1-p}}{r} \sum_{n=0}^{\infty} \frac{\left(\frac{C}{K}\right)^{na} [2^{na+1-p} - 1]}{na+1-p}. \quad (2.6)$$

3. The sequence of doubling times

Consider the sequence of doubling times, $\{\Delta t_j\}$, for the modified Richards model (2.1), (2.3) such that

$$\Delta t_0 := 0, \quad 2C \left(\sum_{k=0}^j \Delta t_k \right) = C \left(\sum_{k=0}^{j+1} \Delta t_k \right), \quad j = 0, 1, 2, \dots, N, \quad (3.1)$$

where N is the finite number of times cumulative cases double until they reach half of the threshold level, $K/2$. Let $C(0) = C_0$. Then clearly,

$$C(\Delta t_0) = C_0, \quad C \left(\sum_{k=0}^1 \Delta t_k \right) = 2C_0, \quad C \left(\sum_{k=0}^2 \Delta t_k \right) = 2^2 C_0, \dots \quad C \left(\sum_{k=0}^j \Delta t_k \right) = 2^j C_0, \quad (3.2)$$

and the $(j + 1)$ th element of the sequence, Δt_{j+1} , can be evaluated as

$$\Delta t_{j+1} = \int_{2^j C_0}^{2^{j+1} C_0} \frac{ds}{F(s)}, \quad j = 0, 1, 2 \dots N. \quad (3.3)$$

For $F(s)$ defined in (2.3), one concludes that

$$\begin{aligned} \Delta t_{j+1} &= \frac{K^{1-p}}{r} \sum_{n=0}^{\infty} \frac{\left(\frac{2^{j+1} C_0}{K}\right)^{na+1-p} - \left(\frac{2^j C_0}{K}\right)^{na+1-p}}{na + 1 - p} \\ &= \frac{(2^j C_0)^{1-p}}{r} \sum_{n=0}^{\infty} \frac{\left(\frac{2^j C_0}{K}\right)^{na} [2^{na+1-p} - 1]}{na + 1 - p}, \quad j = 0, 1, 2 \dots N. \end{aligned} \quad (3.4)$$

Identities (3.4) yield the following breakdown for Δt_{j+1} :

$$\Delta t_{j+1} = \frac{(2^j C_0)^{1-p} [2^{1-p} - 1]}{r(1-p)} + \frac{(2^j C_0)^{1-p}}{r} \sum_{n=1}^{\infty} \frac{\left(\frac{2^j C_0}{K}\right)^{na} [2^{na+1-p} - 1]}{na + 1 - p}, \quad j = 0, 1, 2 \dots N. \quad (3.5)$$

4. Important special cases

In what follows, we consider some important special cases, for which the expression for the doubling time, Δt , admits a closed form representation. First, we explore the classical Richards model where

$$F(s) := rs \left[1 - \left(\frac{s}{K} \right)^a \right], \quad 0 < s < K, \quad (4.1)$$

that is, a special case of (2.3) that corresponds to $p = 1$. According to (2.4), for $p = 1$,

$$\Delta t = \frac{1}{ar} \int_{\left(\frac{C}{K}\right)^a}^{\left(\frac{2C}{K}\right)^a} \frac{du}{u(1-u)} = \frac{\ln 2}{r} + \frac{1}{ar} \ln \left\{ \frac{1 - \left(\frac{C}{K}\right)^a}{1 - \left(\frac{2C}{K}\right)^a} \right\}, \quad (4.2)$$

or, alternatively, $\Delta t = \frac{1}{ar} \ln \left[\frac{(2K)^a - (2C)^a}{K^a - (2C)^a} \right]$. In (4.2), the term $\frac{\ln 2}{r}$ is the reflection of the exponential growth at the early ascending stage of the outbreak. The second term, $\frac{1}{ar} \ln \left\{ \frac{1 - \left(\frac{C}{K}\right)^a}{1 - \left(\frac{2C}{K}\right)^a} \right\}$ is due to the contribution of the factor $1 - \left(\frac{C}{K}\right)^a$ in the model, which, as opposed to a simple exponential growth, takes into account the carrying capacity of the environment. As expected, the expression for Δt implies that the doubling of cases occurs until $C < \frac{K}{2}$. Taking into account (3.3) and (4.1), one arrives at the following equality for the sequence of doubling times, Δt_{j+1} , in the case of the classical Richards model:

$$\Delta t_{j+1} = \frac{\ln 2}{r} + \frac{1}{ar} \ln \left\{ \frac{1 - \left(\frac{2^j C_0}{K}\right)^a}{1 - \left(\frac{2^{j+1} C_0}{K}\right)^a} \right\}, \quad j = 0, 1, 2 \dots N. \quad (4.3)$$

Applying elementary formula $\ln(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} (x-1)^n}{n}$, $0 < x \leq 2$, with $x = 1 - \left(\frac{2^j C_0}{K}\right)^a$ and then with $x = 1 - \left(\frac{2^{j+1} C_0}{K}\right)^a$, one gets

$$\ln \left\{ \frac{1 - \left(\frac{2^j C_0}{K}\right)^a}{1 - \left(\frac{2^{j+1} C_0}{K}\right)^a} \right\} = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} \left(-\left(\frac{2^j C_0}{K}\right)^a\right)^n}{n} - \sum_{n=1}^{\infty} \frac{(-1)^{n-1} \left(-\left(\frac{2^{j+1} C_0}{K}\right)^a\right)^n}{n}$$

$$= \sum_{n=1}^{\infty} \frac{\left(\frac{2^j C_0}{K}\right)^{na} [2^{na} - 1]}{n}, \quad j = 0, 1, 2, \dots, N. \quad (4.4)$$

Thus, for the classical Richards model (2.1), (4.1), one obtains

$$\Delta t_{j+1} = \frac{\ln 2}{r} + \frac{1}{ar} \sum_{n=1}^{\infty} \frac{\left(\frac{2^j C_0}{K}\right)^{na} [2^{na} - 1]}{n}, \quad j = 0, 1, 2, \dots, N. \quad (4.5)$$

Comparing identities (3.5) and (4.5), one can easily notice that the second term in (4.5) coincides with the second term in (3.5) when $p = 1$. At the same time, the first term in (4.5), $\frac{\ln 2}{r}$, is equal to $\lim_{p \rightarrow 1} \frac{(2^j C_0)^{1-p} [2^{1-p} - 1]}{r(1-p)}$. In other words, it is the limit (as p approaches 1) of the first term in (3.5).

Another special case that also gives rise to a Bernoulli differential equation is the one where

$$F(s) := rs^p \left[1 - \left(\frac{s}{K} \right)^{1-p} \right], \quad 0 < s < K. \quad (4.6)$$

For $F(s)$ defined in (4.6) and $0 < p < 1$, one obtains a wide spectrum of early growth rates. At the same time, just like in the case of the classical Richards model, the doubling time here is a simple elementary function. From (2.4) one has

$$\Delta t = b \int_{\left(\frac{C}{K}\right)^a}^{\left(\frac{2C}{K}\right)^a} \frac{du}{1-u} = \frac{K^{1-p}}{(1-p)r} \ln \left\{ \frac{1 - \left(\frac{C}{K}\right)^{1-p}}{1 - \left(\frac{2C}{K}\right)^{1-p}} \right\}, \quad (4.7)$$

and, therefore,

$$\Delta t_{j+1} = \frac{K^{1-p}}{(1-p)r} \ln \left\{ \frac{1 - \left(\frac{2^j C_0}{K}\right)^{1-p}}{1 - \left(\frac{2^{j+1} C_0}{K}\right)^{1-p}} \right\}, \quad j = 0, 1, 2, \dots, N. \quad (4.8)$$

We now consider the generalized-growth model, that is, $a \rightarrow \infty$ in (2.3), which results in

$$F(s) = rs^p, \quad s > 0. \quad (4.9)$$

Similar to (2.3) and (4.6), the generalized-growth model enables us to capture the diversity of epidemic growth profiles during the first few generations of disease transmission, including constant ($p = 0$), polynomial ($0 < p < 1$), and exponential ($p = 1$) growth rates. However, as opposed to (2.3), the doubling time for (4.9) admits the closed form expression for any $0 \leq p \leq 1$. Indeed, according to (2.2), in the case of sub-exponential growth, i.e., when $0 < p < 1$, for $F(s)$ defined in (4.9) one concludes

$$\Delta t = \int_C^{2C} \frac{ds}{rs^p} = \frac{1}{r(1-p)} \left[(2C(t))^{1-p} - (C(t))^{1-p} \right] = \frac{C^{1-p}(t) [2^{1-p} - 1]}{r(1-p)}. \quad (4.10)$$

Thus, for an emerging outbreak, in the absence of reliable information about transmission mechanisms, the generalized-growth model could provide a starting point for the assessment of the potential outcomes, particularly when the epidemiological characteristics of the disease are poorly understood or subject to substantial uncertainty. This implies that the sequence of doubling times, $\{\Delta t_j\}$, increases according to

$$\Delta t_{j+1} = \frac{(2^j C_0)^{1-p} [2^{1-p} - 1]}{r(1-p)}, \quad \Delta t_0 = 0, \quad j = 0, 1, 2, \dots, \quad (4.11)$$

which is precisely the first term in (3.5). As mentioned before, $\lim_{p \rightarrow 1} \frac{(2^j C_0)^{1-p} [2^{1-p} - 1]}{r(1-p)} = \frac{\ln 2}{r}$, the doubling time for the simple exponential growth, $C' = rC$.

5. Doubling times in terms of time variable

In this section, we continue to explore the progression of doubling times for the classical Richards model

$$\frac{dC}{dt} = rC \left[1 - \left(\frac{C}{K} \right)^a \right]. \quad (5.1)$$

The special S-shaped logistic curve, given by (5.1), is described by the following equation:

$$C(t) = \frac{K}{\left\{ 1 + \left[\left(\frac{K}{C_0} \right)^a - 1 \right] e^{-art} \right\}^{\frac{1}{a}}}. \quad (5.2)$$

Substituting this expression for $C(t)$ into (4.2), one arrives at

$$\Delta t = \frac{1}{ar} \ln \left\{ \frac{\left[\left(\frac{2K}{C_0} \right)^a - 2^a \right] e^{-art} + 1 - 2^a}{\left[\left(\frac{K}{C_0} \right)^a - 1 \right] e^{-art}} \right\} = \frac{\ln 2}{r} - \frac{1}{ar} \ln \left\{ 1 - \frac{(2^a - 1)C_0^a e^{art}}{K^a - C_0^a} \right\}. \quad (5.3)$$

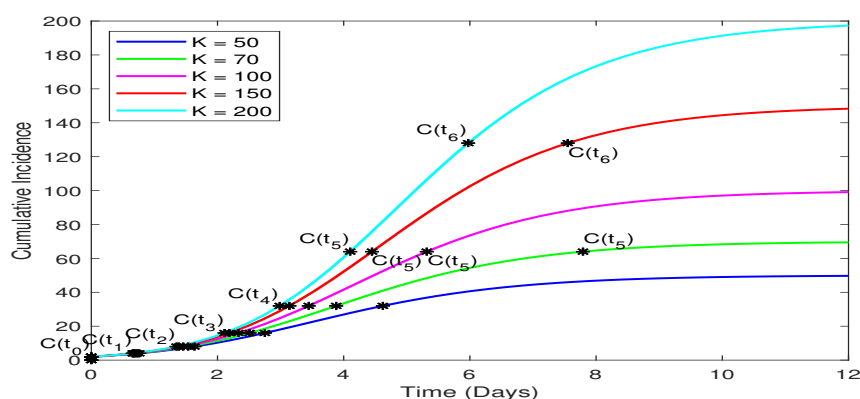


Figure 1. Cumulative incidence with doubling times for $K = 50, 70, 100, 150, 200$ with $C_0 = 2$, $r = 1.2$, and $a = 0.5$.

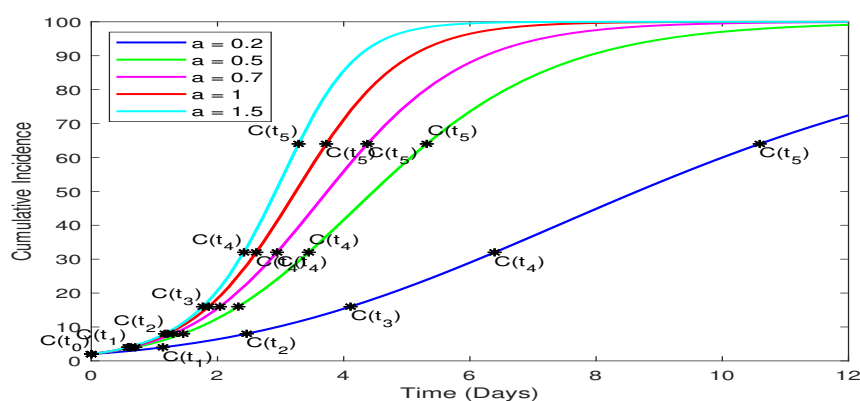


Figure 2. Cumulative incidence with doubling times for $a = 0.2, 0.5, 0.7, 1, 1.5$ with $C_0 = 2$, $r = 1.2$, and $K = 100$.

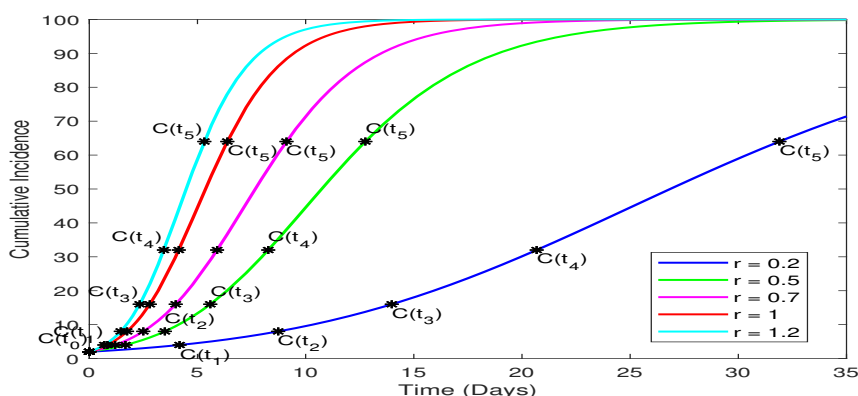


Figure 3. Cumulative incidence with doubling times for $r = 0.2, 0.5, 0.7, 1, 1.2$ with $C_0 = 2$, $a = 0.5$, and $K = 100$.

This identity reinforces the fact that cumulative cases would double unless $C_0 \geq \frac{K}{2}$ (and this is unlikely for most outbreaks). The cumulative cases will continue to double until the moment

$$\hat{t} = \frac{1}{ar} \ln \mu, \quad \mu := \frac{K^a - C_0^a}{(2^a - 1)C_0^a}, \quad (5.4)$$

that is, until the moment $C(t)$ reaches the value $K/2$. At that same moment, the argument of the logarithm in (5.3) becomes negative. As t approaches $\hat{t} = \frac{1}{ar} \ln \mu$ (and $C(t)$ approaches $K/2$), the term $\frac{1}{ar} \ln \left\{ 1 - \frac{(2^a - 1)C_0^a e^{art}}{K^a - C_0^a} \right\}$ in (5.3) goes to negative infinity, which means the doubling time approaches infinity.

For comparison, the doubling time for the generalized-growth model, $C' = rC^p$, (as a function of time passed from the start of the outbreak) takes the form

$$\Delta t = (2^{1-p} - 1) \left\{ t + \frac{C_0^{1-p}}{(1-p)r} \right\}, \quad C(0) = C_0. \quad (5.5)$$

Obviously, $\lim_{p \rightarrow 1} (2^{1-p} - 1) \left\{ t + \frac{C_0^{1-p}}{(1-p)r} \right\} = \frac{\ln 2}{r}$, which is the doubling time for $C' = rC$.

6. Numerical simulations

We now investigate how the doubling time depends on various values of the parameters r , K , and a for classical Richards model (4.1). In the first numerical example, we set $C_0 = 2$, $r = 1.2$, and $a = 0.5$. Our goal is to visualize how various values of K alter the doubling times and the sequence of doubling times. From a theoretical standpoint (see (4.2)), the doubling time must be decreasing with respect to K . The numerical experiment confirms this theoretical finding as shown in Figures 1 and 4. As expected, the number of times cumulative cases double increases as the value of K goes up.

In the second numerical example, we set $C_0 = 2$, $r = 1.2$, and $K = 100$. Here the goal is to observe how various values of a alter the doubling times and the sequence of doubling times. Again, from a theoretical standpoint (see (4.2)), the doubling time decreases with respect to a . Thus, as a increases,

the doubling of cases occurs more quickly and it takes less time for the cumulative cases to approach the carrying capacity as shown in Figures 2 and 5.

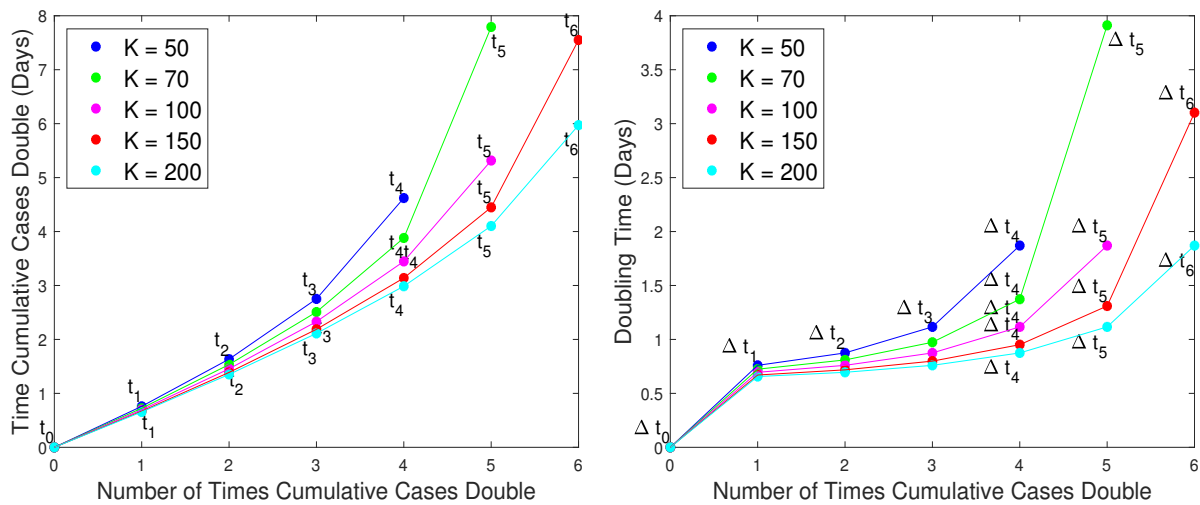


Figure 4. Moments of time where doubling occurs (left) and intervals of time between consecutive doubling of cases (right) for various values of K with $C_0 = 2$, $r = 1.2$, and $a = \frac{1}{2}$.

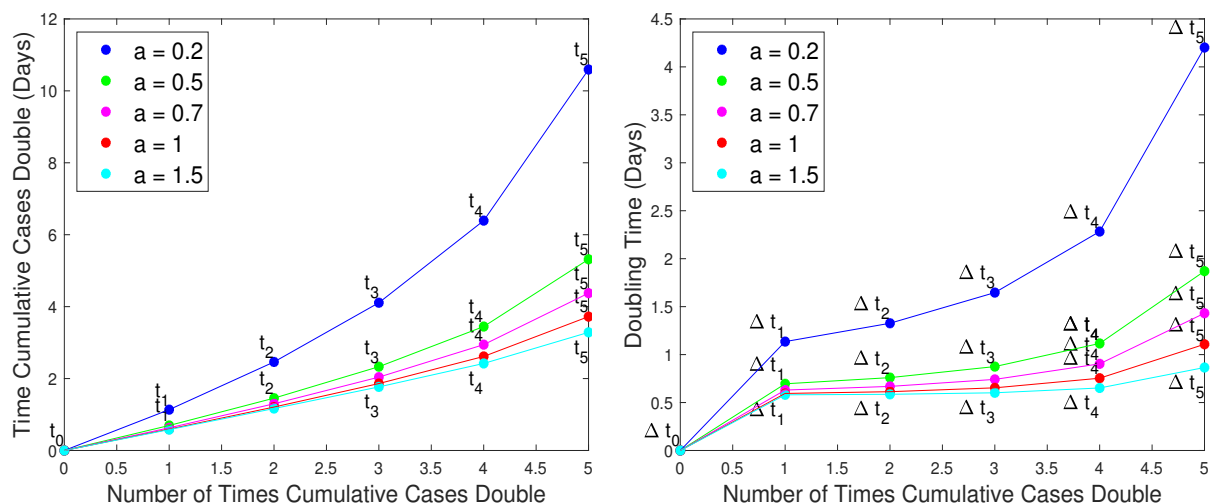


Figure 5. Moments of time where doubling occurs (left) and interval of time between consecutive doubling of cases (right) for various values of a with $C_0 = 2$, $r = 1.2$, and $K = 100$.

In the third example, we fix $C_0 = 2$, $a = 0.5$, and $K = 100$ and we study the dependence of doubling times on the parameter r . Numerical simulations illustrate that the dependence is very similar to the one of a , that is, the doubling time decreases as r goes up and, as r increases, the cumulative cases double more quickly and it takes less time to reach the threshold (Figures 3 and 6).

7. Comparison of doubling times

Here our goal is to compare the sequences of doubling times across various models. In the first experiment, we examine how the sequences of doubling times compare for Richards model and the

modified Richards model. In the case of both models, we set $a = 0.5$, $r = 1.2$, $C_0 = 1$, and $K = 100$, and our goal is to see how the sequence of doubling times for the modified Richards model behaves as p approaches 1. To that end, we look at the sequence of doubling times for $p = 0.5, 0.7, 0.8, 0.9$, and 0.95 . In line with our theoretical study, as p approaches 1, the sequence of doubling times for the modified Richards model approaches that of the Richards model as illustrated in Figure 7.

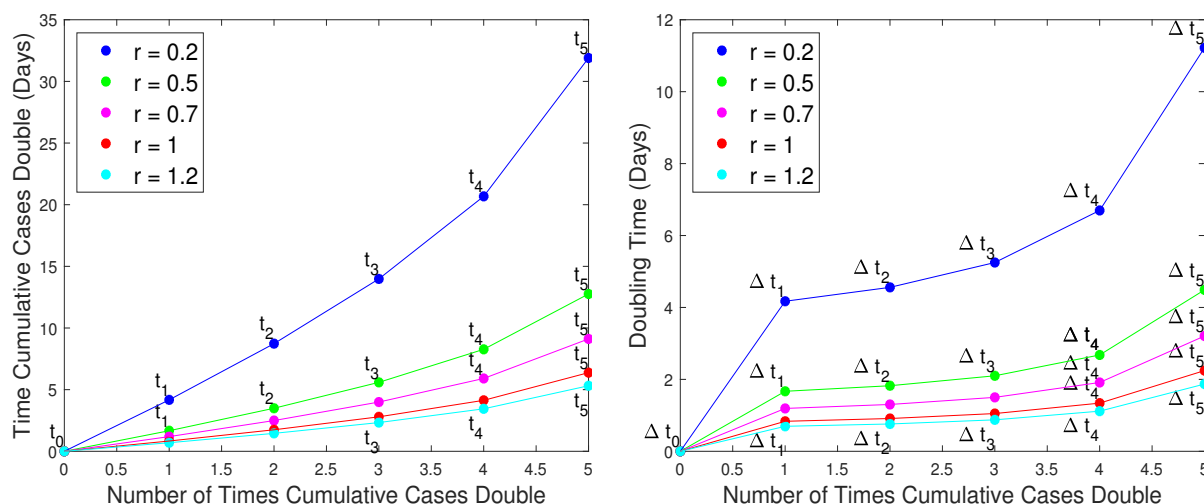


Figure 6. Moments of time where doubling occurs (left) and interval of time between consecutive doubling of cases (right) for various values of a with $C_0 = 2$, $a = 0.5$, and $K = 100$.

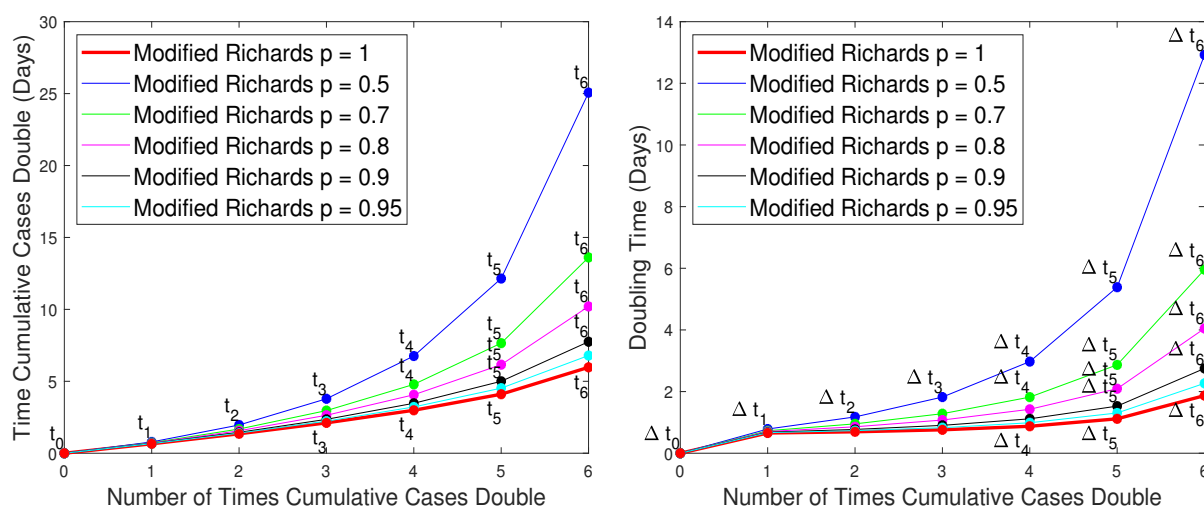


Figure 7. Comparing time cumulative cases double (left) and time between consecutive doubling of cases (right) between Richards model and the modified Richards model.

In the second experiment, we compare the generalized growth model to the modified Richards model. Clearly, the modified Richards model becomes the generalized growth model as $a \rightarrow \infty$. In order to simulate this theoretical result, for both models we set $K = 100$, $r = 1.2$, $p = 0.5$, and $C_0 = 1$, and the modified Richards model is considered for $a = 0.5, 0.9, 1, 1.5$, and 2.5 . From Figure 8, one can see that as a increases, the sequences of doubling times for the modified Richards model approaches that of the generalized growth model.

In the third numerical example, we compare the sequence of doubling times for the generalized growth model and the simple exponential model. For this numerical example, we set $r = 1.2$ and $C_0 = 1$, and then the generalized growth model is studied for $p = 0.5, 0.7, 0.8, 0.9$, and 0.95 . Figure 9 confirms that as p approaches 1, the sequence of doubling times of the generalized growth model approaches that of the exponential model.

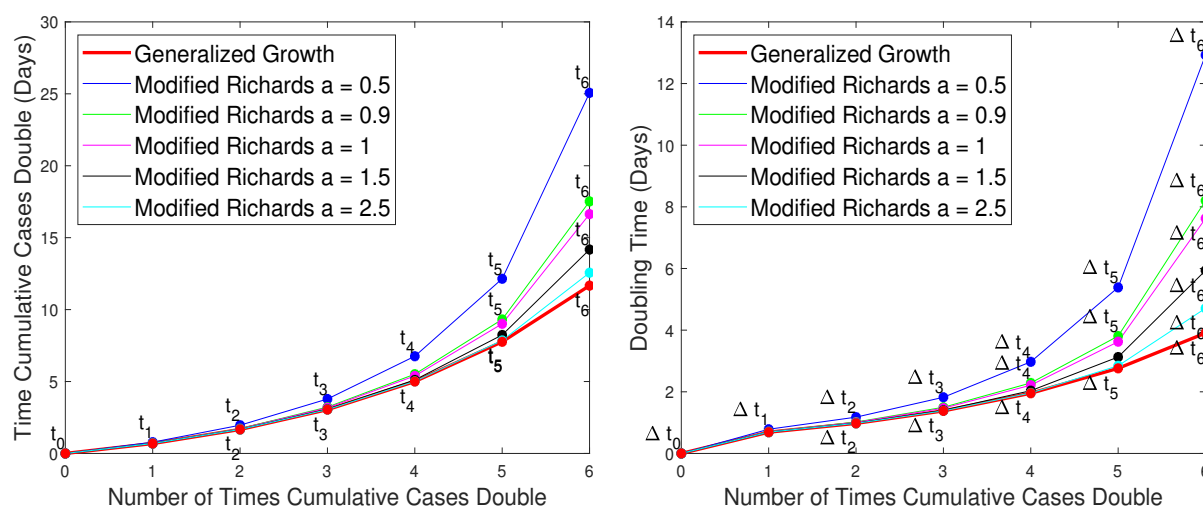


Figure 8. Comparing time cumulative cases double (left) and time between consecutive doubling of cases (right) between generalized growth model and the modified Richards model.

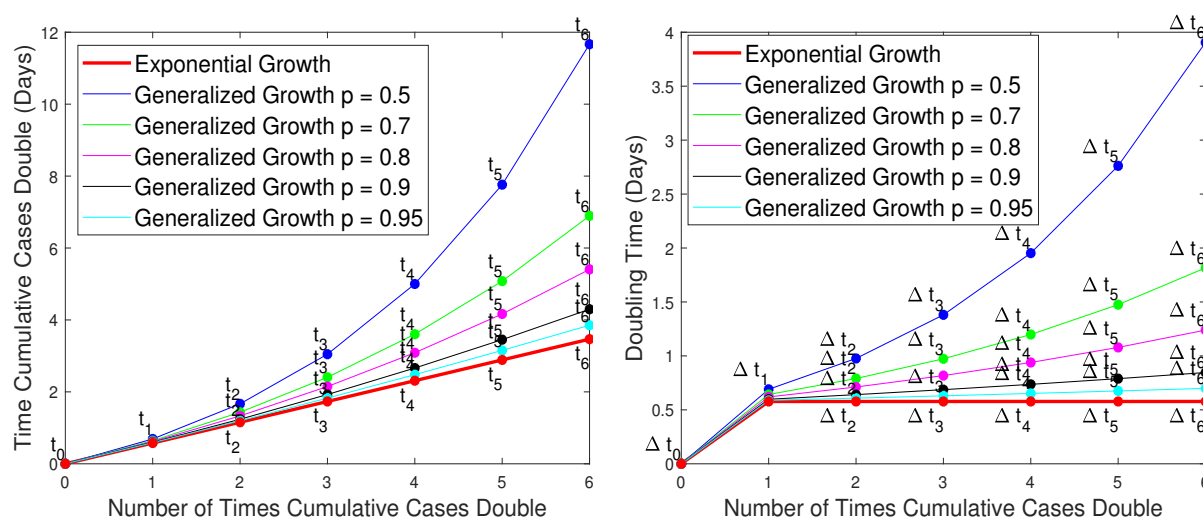


Figure 9. Comparing time cumulative cases Double (left) and time between consecutive doubling of cases (right) between exponential growth and generalized growth.

8. Synthetic data analysis

Our goal here is to estimate unknown parameters r , a , p , and K for the modified Richards model using least squares minimization, specifically through the lsqcurvefit algorithm in MATLAB. The parameters are extracted from early wave of the outbreak. Then the doubling times are calculated from these extracted parameters.

We start by using synthetic data in order to assess if three methods for uncertainty quantification, Normal Approximation, Empirical Likelihood, and Jackknife Empirical Likelihood (see Appendix for details), employed to calculate 95% confidence intervals (CIs) for the mean doubling time, will give rise to the CIs that contain true doubling times' values. Note that the Normal Approximation (NA) method provides a formulaic representation for calculating confidence intervals, and the calculation of confidence intervals is fast. On the flip side, NA method relies on the use of the normal family of distributions which can be applied when the sample size is large (typically when $n > 30$). In the case of limited emerging data, the method likely won't be applicable for accurate uncertainty quantification since the normal distribution may not be appropriate especially if the emerging incidence data is heavily skewed. Empirical Likelihood (EL) method doesn't require the use of this assumption and is more robust against non-normality and skewed data sets. Inferences are done using only the empirical data. The EL method can be easily implemented in most statistical software. However, if the sample size is small, the EL algorithm may diverge or fail to produce accurate confidence intervals depending on the variability of the data. And as the sample size increases, the computational burden increases over the NA method.

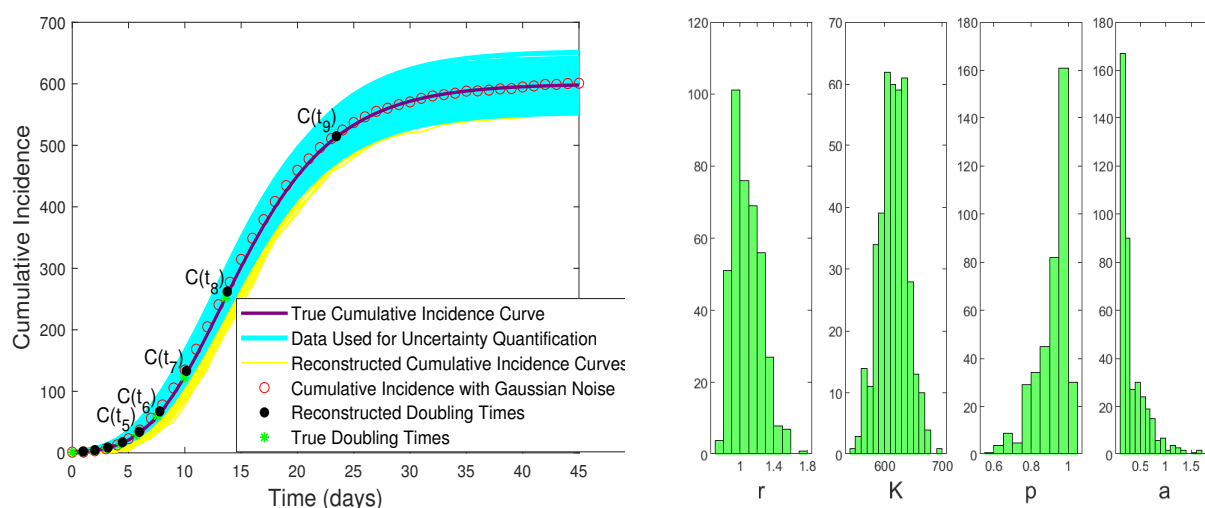


Figure 10. Simulated cumulative incidence curves with Gaussian noise with true doubling times (left) and histograms of extracted parameters for modified Richards model (right) for synthetic data with $r = 1.2$, $K = 600$, $p = 1$, and $a = 0.15$.

The Jackknife Empirical Likelihood (JEL) method makes no use of the underlying distribution of data, and inferences are done using only the empirical data. It applies to both normal and skewed distributions. But, unlike the EL method, the JEL performs well with both small and large samples. Unfortunately, the computational cost of JEL is greater than the cost of EL and NA methods. The results take longer to obtain. So, having consistent results with all three algorithms, allows to confirm that the CIs are reliable regardless of whether the incidence data is normal or heavily skewed and whether the sample size is large or limited.

To quantify uncertainty in the reconstructed doubling times' values, we conduct two numerical experiments. In the first experiment, we set

$$r = 1.2, \quad K = 600, \quad p = 1, \quad a = 0.15.$$

These parameters are used to solve the forward ODE problem and to generate cumulative data on a given time interval $[t_1, t_n]$. Afterwards, random Gaussian noise (with 0 mean and standard deviation equal to 1) is added to the generated cumulative curve to get "real" epidemic data. Then MATLAB *lsqcurvefit* is employed to fit the model and to extract estimated parameters a , p , K , and r . Since incidence data is known to be positive in a real-life setting, we use uniform noise in the case where the incidence becomes negative.

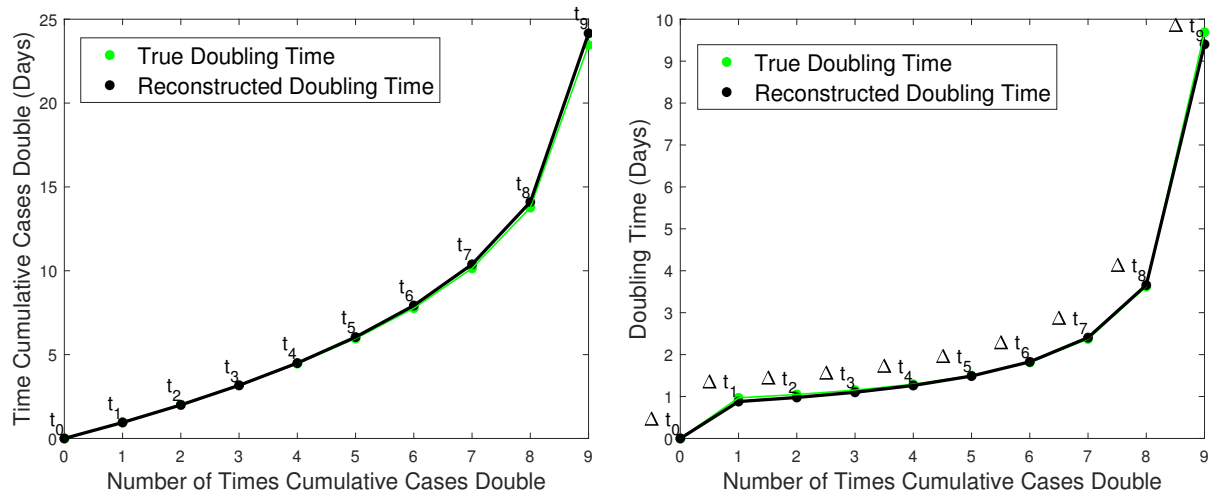


Figure 11. Comparison of true time doubling occurs with reconstructed time (left) and comparison of true interval of time between doubling of cases with reconstructed interval of time (right) for synthetic data with $r = 1.2$, $K = 600$, $p = 1$, and $a = 0.15$.

Table 1. 95% confidence intervals for extracted parameters for synthetic data with $r = 1.2$, $K = 600$, $p = 1$, and $a = 0.15$.

r	K	p	a
1.1 (0.81,1.4)	620 (560,650)	0.96 (0.77,1.3)	0.3 (0.13,0.96)

We perform uncertainty quantification for the extracted parameters by assuming Poisson error structure. A total of 400 bootstrap iterations are carried out and the extracted parameters are used to calculate the sequence of doubling times for all 400 bootstrap iterations. Displayed in Figure 10 are the graphs containing the reconstructed cumulative incidence curves with noisy data and the extracted parameters a , p , K , and r (see also Table 1). Table 4 (see Appendix) shows the true moments when the doubling of time occurs given the original values of a , p , K , and r , and the 95% confidence intervals obtained by Normal Approximation, Empirical Likelihood, and Jackknife Empirical Likelihood methods (see Appendix for details). In Table 5 (see Appendix), the true intervals of time between consecutive doublings of cases and the corresponding CIs (estimated by using the Normal Approximation, Empirical Likelihood, and Jackknife Empirical Likelihood methods) are presented. Figure 11 shows the comparison of true time when the doubling occurs with reconstructed time (left) and the comparison of true intervals of time between doubling of cases with reconstructed intervals of time (right).

One can clearly see that in the first numerical simulation, the 95% confidence intervals, obtained from the parameter estimation inverse problem, correctly capture the true parameters of the modified

Richards model (see Table 1) as well as the true doubling times (Tables 4 and 5). Looking at the constructed intervals, the Normal Approximation method tends to produce the widest intervals.

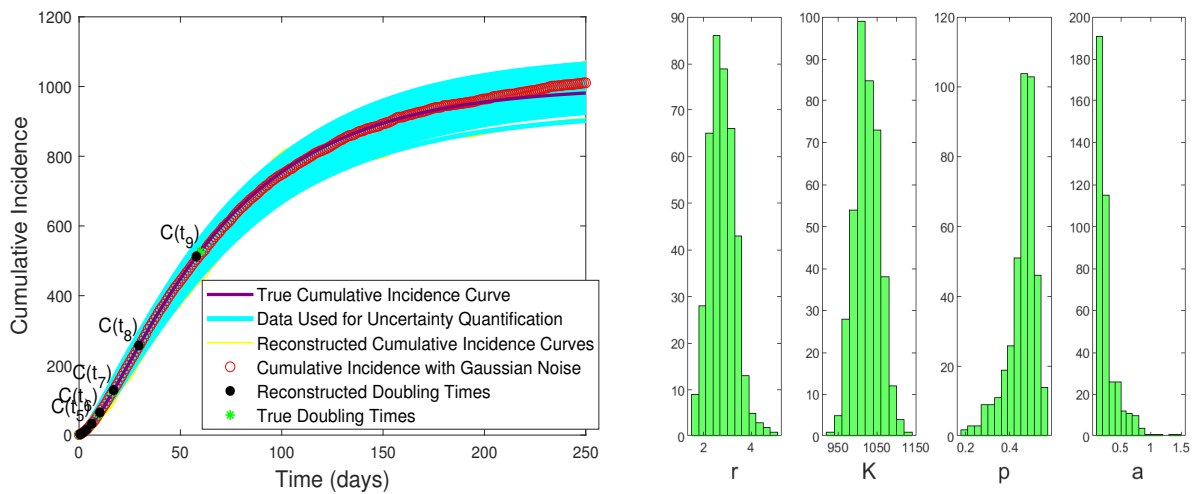


Figure 12. Simulated cumulative incidence curves with Gaussian noise with true doubling times (left) and histograms of extracted parameters for modified Richards model (right) for synthetic data with $r = 2.2$, $K = 1000$, $p = 0.5$, and $a = 0.25$.

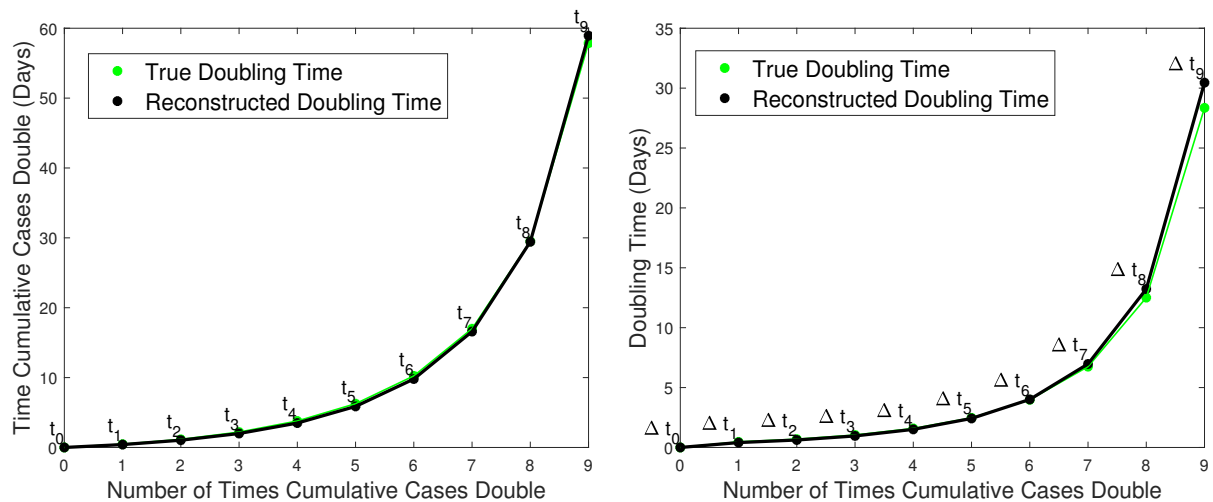


Figure 13. Comparing true time doubling occurs with reconstructed time (left) and comparing true interval of time between doubling of cases with reconstructed interval of time (right) for synthetic data with $r = 2.2$, $K = 1000$, $p = 0.5$, and $a = 0.25$.

Table 2. Extracted parameter results with 95% confidence intervals using Gaussian noise $r = 2.2$, $K = 1000$, $p = 0.5$, and $a = 0.25$.

r	K	p	a
2.9 (1.9,4.2)	1000 (9500, 11000)	0.47 (0.31,0.55)	0.25 (0.15,0.70)

The Empirical Likelihood and Jackknife Empirical Likelihood methods produce narrower intervals.

Among the last two, the Jackknife Empirical likelihood produces narrower intervals which is to be expected. In the second experiment, we set the following original values

$$r = 2.2, \quad K = 1000, \quad p = 0.5, \quad a = 0.25.$$

The simulations are repeated using the three uncertainty quantification methods described above, and the results are displayed in Figures 12 and 13 and in Tables 2, 6, and 7 (see Appendix for table results).

In the second numerical simulation, the 95% confidence intervals accurately capture the true parameters and the true doubling times for the modified Richards model. The Empirical Likelihood and Jackknife Empirical Likelihood methods both produced narrower intervals over the NA method but from a practical standpoint, they're both roughly the same up to a certain decimal point as was seen in our simulation studies. Since narrower intervals tend to be more precise, the JEL method overall outperformed the NA and EL methods.

9. Real data analysis

In the following examples, we look to estimate the doubling times using real data for the states of New Hampshire and Hawaii.

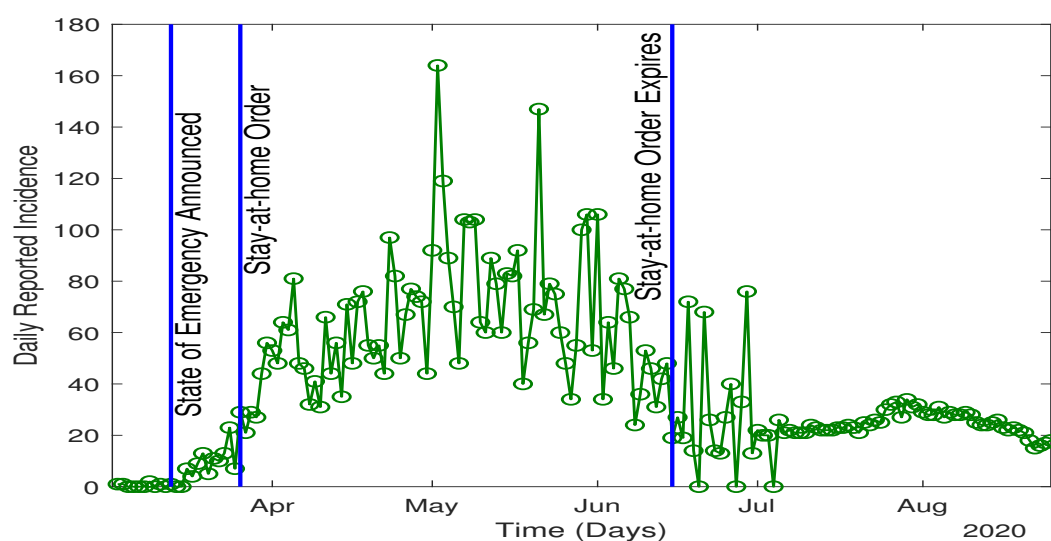


Figure 14. Reported daily incidence for New Hampshire from March 2 to August 25, 2020.

Assume that a wave of an outbreak of COVID-19 originates on day t_1 and ends on day b (which is unknown at the early stage of disease transmission). Let cumulative data, d_δ , be reported at $t = t_1 < t_2 < \dots < t_n$ with t_n being smaller than b . In that case, given limited data, d_δ , at the start of a new wave one has to solve the following constrained minimization problem

$$\min_{r,p,a,K} \frac{1}{2} \|C - d_\delta\|_2^2 = \frac{1}{2} \sum_{i=1}^n (C(t_i) - d_\delta(t_i))^2 \quad (9.1)$$

To solve this minimization problem in a stable manner, we use the MATLAB built-in function *lsqcurvefit*. This function solves the nonlinear curve fitting problem using Levenberg - Marquardt

optimization algorithm, which allows to extract the system parameters of r , a , K , and p that best model the early wave of the outbreak as illustrated in Table 7. After p , K , a , and r have been recovered from the early epidemic data, 400 additional cumulative bootstrap curves are generated by adding Poisson error structure to the series of reported cases in order to quantify uncertainty in the reconstructed parameters.

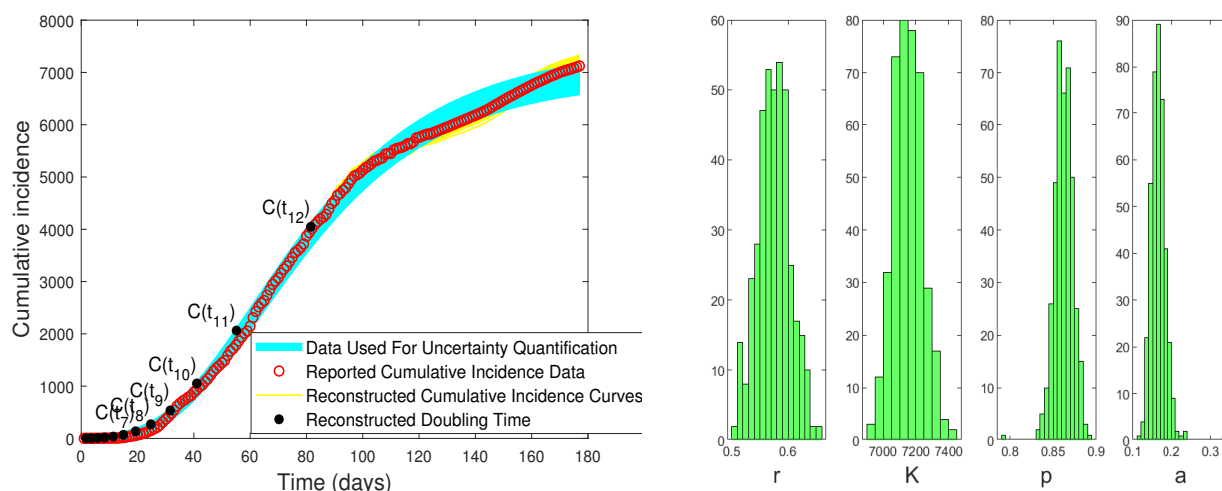


Figure 15. Reconstructed cumulative incidence cases for New Hampshire with doubling times (left) and recovered parameters (right).

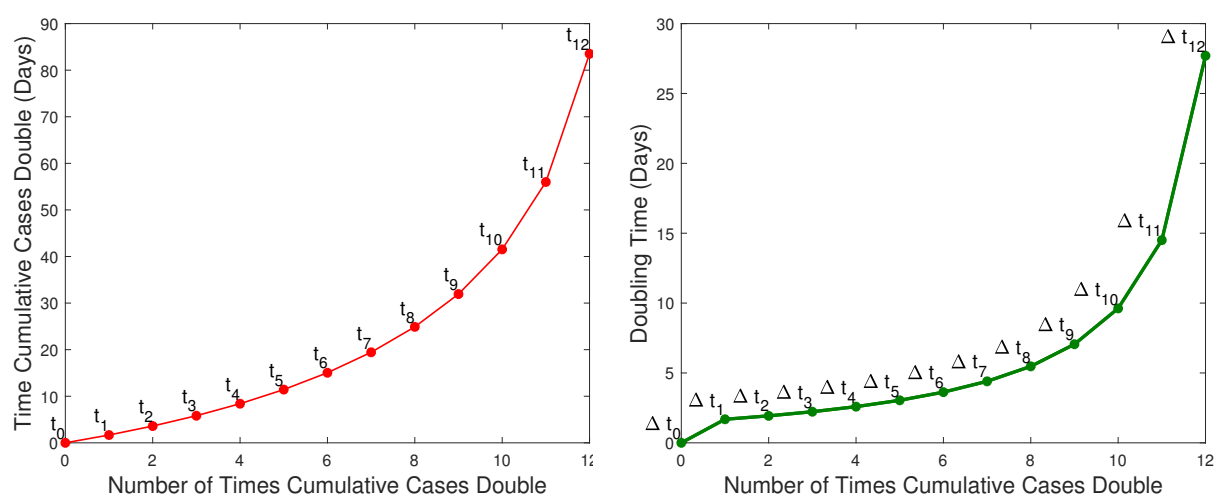


Figure 16. Time where doubling occurs (left) and time between consecutive doubling of cases (right) for New Hampshire.

In Figures 15 and 18, we show the reconstructed cumulative incidence for the two states with real cumulative incidence data for comparison (along with the histograms of the obtained values of those parameters along with their 95% confidence intervals). We also calculate the doubling times for the extracted values of the parameters r , a , K , and p . Figures 16 and 19 include the moments of time the doubling occurs and the intervals of time between consecutive doubling of cases. Expression (3.5) and the extracted parameter values are used to calculate the sequence of doubling times until cumulative cases stop doubling.

In the first example, we show the results obtained for the state of New Hampshire between March 2 to August 25, 2020 (See Figures 14–16), and in the second example, we present the results obtained for Hawaii from March 5 to May 28, 2020 (Figures 17–19) [12].

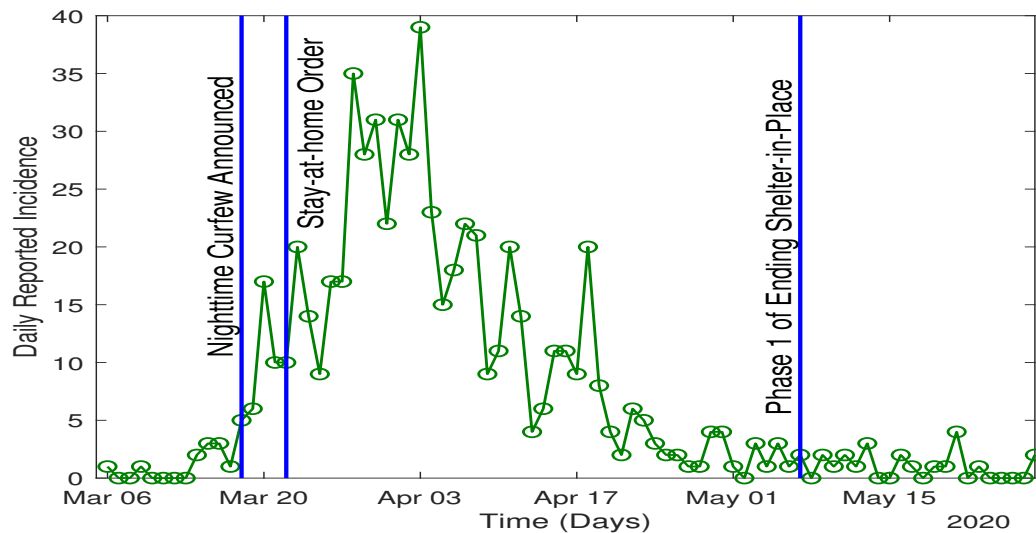


Figure 17. Reported daily incidence for Hawaii from March 5, 2020 to May 28, 2020.

To quantify uncertainty in the extracted doubling times, we refit the model to $M = 400$ additional data sets for cumulative cases (assuming Poisson error structure). This yields 400 sequences of doubling times for each state. Tables 8–11 display the Normal Approximation, Empirical Likelihood, and Jackknife Empirical Likelihood methods used to construct the 95% confidence intervals for the estimated doubling times (see Appendix).

Table 3. Extracted parameter results with 95% confidence intervals.

State	r	K	p	a
New Hampshire	0.57 (0.52, 0.63)	7200 (7000, 7300)	0.86 (0.84, 0.88)	0.17 (0.13, 0.2)
Hawaii	0.71 (0.54, 0.91)	640 (590, 690)	1.1 (1, 1.1)	0.094 (0.069, 0.13)

The constructed 95% confidence intervals for the sequences of doubling times for the states of New Hampshire and Hawaii are displayed in Tables 8–11 (see Appendix). These tables show the resulting confidence intervals for the doubling times of cumulative cases during the ascending phase of the first wave of the outbreak as well as the times between consecutive doubling of cases during the same phase. By looking at the CIs, we can see that the intervals for doubling times are very narrow which is not surprising because the intervals constructed for our recovered parameters are also quite narrow. Thus, the recovered doubling times will be less variable as the iterative process is carried out. In addition, the NA confidence intervals are generally wider than the non-parametric EL and JEL methods which is consistent with the simulation studies conducted above for synthetic data. The non-parametric tests make no assumptions about the distribution of the doubling times at each time point so they are constructed from the data only which is important when we cannot make any claims about

the distribution. Therefore they are considered more robust due to their reliance on fewer assumptions. As such, the JEL confidence intervals may be the most applicable to our purposes since they do not violate any major assumptions about the population distribution and they are generally the narrowest among all the intervals constructed.

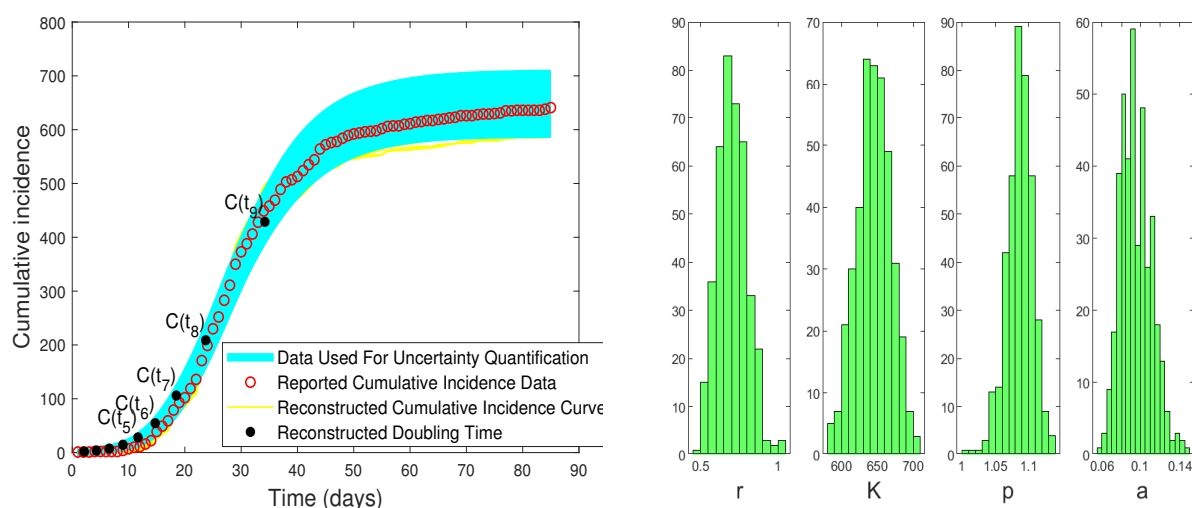


Figure 18. Reconstructed cumulative incidence cases for Hawaii with doubling times (left) and recovered parameters (right).

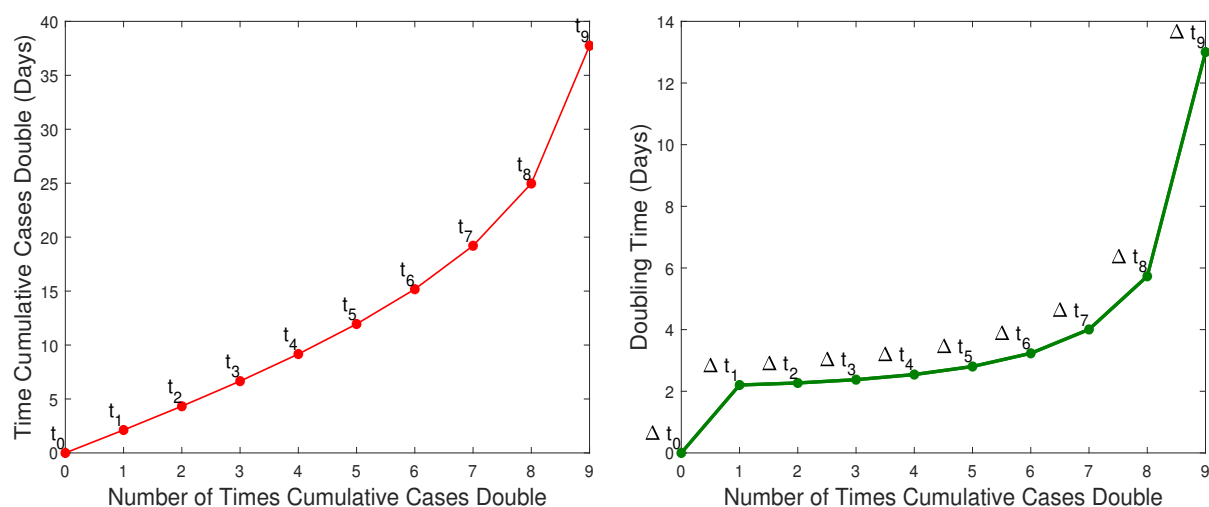


Figure 19. Time where doubling occurs (left) and time between consecutive doubling of cases (right) for Hawaii.

10. Conclusions

In this paper, we derived theoretical formulas for calculating the doubling times for the modified Richards model and important special cases. Using these theoretical calculations, we analyzed the growth of epidemics in terms of the number of times cumulative cases double until the epidemic peaks and the rate at which intervals between consecutive doubling times change. We explored the relationships between the doubling times of the modified Richards model, the exponential growth model, the

generalized growth model, and the classical Richards model.

The derived formulas have been investigated numerically using, first, synthetic data sets with predetermined values of the parameters governing the modified Richards model to assess model accuracy and validate our use of various uncertainty quantification methods. We employed three methods to quantify uncertainty: Normal Approximation, Empirical Likelihood, and Jackknife Empirical Likelihood. Normal Approximation tended to always produce the widest intervals due to the assumed structure of the probability distribution underlying the doubling times at each moment cumulative cases double. Empirical Likelihood and Jackknife Empirical Likelihood produced narrower intervals and are generally considered more precise due to this property. The Normal approximation method can be calculated quickly using common statistical functions built into R. Empirical Likelihood and Jackknife Empirical Likelihood computationally take a bit longer to construct as there are no built in functions to calculate them quickly. Nevertheless, due to their non-parametric nature, whenever the calculations are entirely based on the data, we recommend using Jackknife Empirical Likelihood since the constructed intervals are the most precise, and in practice have better small sample performance.

In the future, we would like to try other forms of Empirical Likelihood methods to assess their accuracy and to see how they perform in determining the true doubling time for each time point. Another important topic that needs to be studied next is the extension of the modified Richards model to the case of elaborate epidemic trajectories aggregating multiple asynchronous sub-epidemics, since a large number of COVID-19 incidence data curves do not have a simple bell-shape behavior for each phase of the outbreak [13, 14].

Acknowledgments

This study is supported by NSF awards 1818886, 2011622, 1610429, and 1633381, by NIH grant R01 GM 130900, and by Simons Foundation grant 638679.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. A. Smirnova, L. DeCamp, G. Chowell, Mathematical and statistical analysis of doubling times to investigate the early spread of epidemics: Application to the COVID-19 pandemic, *Mathematics*, **9** (2021), 625. <https://doi.org/10.3390/math9060625>
2. G. Chowell, L. Simonsen, C. Viboud, Y. Kuang, Is west Africa approaching a catastrophic phase or is the 2014 Ebola epidemic slowing down? Different models yield different answers for Liberia, *PLoS Curr.*, **6** (2014). <https://doi.org/10.1371/currents.outbreaks.b4690859d91684da963dc40e00f3da81>
3. B. Hau, E. Kosman, Comparative analysis of flexible two-parameter models of plant disease epidemics, *Phytopathology*, **97**(10), (2007), 1231–1244. <https://doi.org/10.1094/PHYTO-97-10-1231>

4. Y. H. Hsieh, Pandemic influenza A (H1N1) during winter influenza season in the southern hemisphere, *Influenza Other Respi. Viruses*, **4** (2010), 187–197. <https://doi.org/10.1111/j.1750-2659.2010.00147.x>
5. A. N. Tsoularis, J. Wallace, Analysis of Logistic Growth Models, *Math. Biosci.*, **179** (2002), 21–55. [https://doi.org/10.1016/S0025-5564\(02\)00096-2](https://doi.org/10.1016/S0025-5564(02)00096-2)
6. M. E. J. Turner, E. L. J. Bradley, K. Kirk, K. M. Pruitt, A theory of growth *Math. Biosci.*, **29** (1976), 367–373. [https://doi.org/10.1016/0025-5564\(76\)90112-7](https://doi.org/10.1016/0025-5564(76)90112-7)
7. S. A. Colgate, E. A. Stanley, J. M. Hyman, S. P. Layne, C. Qualls, Risk behavior-based model of the cubic growth of acquired immunodeficiency syndrome in the United States, *Proc. Natl Acad. Sci. U. S. A.*, **86** (1989), 4793–4797. <https://doi.org/10.1073/pnas.86.12.4793>
8. J. Ma, J. Dushoff, B. M. Bolker, D. J. Earn, Estimating initial epidemic growth rates, *Bull. Math. Biol.*, **76** (2014), 245–60. <https://doi.org/10.1007/s11538-013-9918-2>
9. B. Szendroi, G. Csanyi, Polynomial epidemics and clustering in contact networks, *Proc. Biol. Sci.*, **271** (2004), S364–S366. <https://doi.org/10.1098/rsbl.2004.0188>
10. C. Viboud, L. Simonsen, G. Chowell, A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks epidemics, *Epidemics*, **15** (2016), 27–37. <https://doi.org/10.1016/j.epidem.2016.01.002>
11. C. Jan, *Gradually-varied Flow Profiles in Open Channels. Analytical Solutions by Using Gaussian Hypergeometric Function*, Springer-Verlag, 2014. <https://doi.org/10.1007/978-3-642-35242-3>
12. *Trends in Number of COVID-19 Cases and Deaths in the US Reported to CDC, by State/Territory*, 2022. Available from: <https://stacks.cdc.gov/view/cdc/102187>.
13. G. Chowell, A. Tariq, J. M. Hyman, A novel sub-epidemic modeling framework for short-term forecasting epidemic waves, *BMC Med.*, **17** (2019), 164. <https://doi.org/10.1186/s12916-019-1406-6>
14. K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J. M. Hyman, et al., Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China, *J. Clin. Med.*, **9** (2020), 596. <https://doi.org/10.3390/jcm9020596>
15. N. Mukhopadhyay, Asymptotic normality of sequential stopping times with applications: Confidence intervals for an exponential mean, *Calcutta Stat. Assoc. Bull.*, **72** (2020). <https://doi.org/10.1177/0008068320923895>
16. W. Zhu, N. Zeng, N. Wang, *Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations*, 2010. Available from: <http://www.cpdm.ufpr.br/documentos/ROC.pdf>
17. A. Owen, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, **75** (1988), 237–249. <https://doi.org/10.1093/biomet/75.2.237>
18. A. Owen, Empirical likelihood ratio confidence regions, *Ann. Stat.*, **18** (1990), 90–120. <https://doi.org/10.1214/aos/1176347494>
19. H. Huang, Y. Zhao, Empirical likelihood for the bivariate survival function under univariate censoring, *J. Stat. Plann. Inference*, **194** (2018), 32–46. <https://doi.org/10.1016/j.jspi.2017.10.002>

20. G. Cheng, Y. Zhao, B. Li, Empirical likelihood inferences for the semiparametric additive, isotonic regression *J. Multivar. Anal.*, **112** (2012), 172–182. <https://doi.org/10.1016/j.jmva.2012.06.003>
21. J. Zhang, J. Zhang, X. Zhu, T. Lu, Testing symmetry based on empirical likelihood, *J. Appl. Stat.*, **45** (2018), 2429–2445. <https://doi.org/10.1080/02664763.2017.1421917>
22. B. Y. Jing, J. Yuan, W. Zhou, Jackknife empirical likelihood, *J. Am. Stat. Assoc.*, **104** (2009), 1224–1232. <https://doi.org/10.1198/jasa.2009.tm08260>
23. Y. Zhao, X. Meng, H. Yang, Jackknife empirical likelihood inference for the mean absolute deviation, *Comput. Stat. Data Anal.*, **91**, (2015), 92–101. <https://doi.org/10.1016/j.csda.2015.06.001>
24. Y. Sang, X. Dang, Y. Zhao, Jackknife empirical likelihood methods for Gini correlations and their equality testing, *J. Stat. Plann. Inference*, **199** (2019), 45–59. <https://doi.org/10.1016/j.jspi.2018.05.004>
25. H. Lin, Z. Li, D. Wang, Y. Zhao, Jackknife empirical likelihood for the error variance in linear models, *J. Nonparametr. Stat.*, **29** (2017), 151–166. <https://doi.org/10.1080/10485252.2017.1285028>
26. Y. Cheng, Y. Zhao, Bayesian jackknife empirical likelihood, *Biometrika*, **106** (2019), 981–988. <https://doi.org/10.1093/biomet/asz031>

Appendix

A. Generalized logistic growth model

All epidemic models considered above are nested in the generalized logistic model [5] defined as

$$\frac{dC}{dt} = F(C), \quad F(s) := rs^p \left[1 - \left(\frac{s}{K} \right)^a \right]^\gamma, \quad 0 < s < K, \quad (\text{A.1})$$

where $r > 0$ represents the intrinsic growth rate in the absence of any limitation to disease spread, $0 \leq p \leq 1$, is a "deceleration of growth" parameter, $K > 0$ is the size of the epidemic, $a > 0$ is a parameter that measures the extent of deviation from the S-shaped dynamics of the classical logistic growth model, and $\gamma > 0$ controls the flexibility of the inflection point. Just like the modified Richards model, the generalized logistic model (A.1) enables us to capture different growth profiles ranging from constant incidence ($p = 0$), polynomial growth ($0 < p < 1$), to exponential growth ($p = 1$) at the early accenting stage of the outbreak. While it is unclear if the addition of extra parameter, γ , gives any considerable advantage from biological standpoint, mathematically the study of doubling times can easily be extended to the generalized logistic model (A.1). Indeed, using the same substitute, $u := (\frac{s}{K})^a$, for the doubling time, Δt , one obtains

$$\Delta t = \frac{K^{1-p}}{ar} \int_{(\frac{C}{K})^a}^{(\frac{2C}{K})^a} \frac{du}{u^{1-\frac{1-p}{a}}(1-u)^\gamma} = b \int_{(\frac{C}{K})^a}^{(\frac{2C}{K})^a} \frac{du}{u^\delta(1-u)^\gamma}, \quad (\text{A.2})$$

where $\delta := 1 - \frac{1-p}{a}$ and $b := \frac{K^{1-p}}{ar}$. Given that

$$\frac{1}{(1-u)^\gamma} = \sum_{n=0}^{\infty} \frac{\gamma(\gamma+1)(\gamma+2)\dots(\gamma+n-1)}{n!} u^n = \sum_{n=0}^{\infty} \frac{\Gamma(n+\gamma)}{\Gamma(\gamma)n!} u^n,$$

one arrives at the following expression for the doubling times

$$\Delta t = b \int_{\left(\frac{C}{K}\right)^a}^{\left(\frac{2C}{K}\right)^a} \sum_{n=0}^{\infty} \frac{\Gamma(n+\gamma)}{\Gamma(\gamma)n!} u^{n-\delta} du = \frac{b}{\Gamma(\gamma)} \sum_{n=0}^{\infty} \frac{\Gamma(n+\gamma) \left[\left(\frac{2C}{K}\right)^{a(n+1-\delta)} - \left(\frac{C}{K}\right)^{a(n+1-\delta)} \right]}{n!(n+1-\delta)}. \quad (\text{A.3})$$

In the above, $\Gamma(\gamma+1) := \int_0^\infty e^{-t} t^\gamma dt$ is the gamma function. From (A.3) and the definitions of b and δ , one concludes that

$$\Delta t = \frac{K^{1-p}}{r\Gamma(\gamma)} \sum_{n=0}^{\infty} \frac{\Gamma(n+\gamma) \left[\left(\frac{2C}{K}\right)^{na+1-p} - \left(\frac{C}{K}\right)^{na+1-p} \right]}{n!(na+1-p)} = \frac{C^{1-p}}{r\Gamma(\gamma)} \sum_{n=0}^{\infty} \frac{\Gamma(n+\gamma) \left(\frac{C}{K}\right)^{na} [2^{na+1-p} - 1]}{n!(na+1-p)}.$$

This brings about the doubling times sequence in the form:

$$\Delta t_{j+1} = \frac{(2^j C_0)^{1-p}}{r\Gamma(\gamma)} \sum_{n=0}^{\infty} \frac{\Gamma(n+\gamma) \left(\frac{2^j C_0}{K}\right)^{na} [2^{na+1-p} - 1]}{n!(na+1-p)}, \quad j = 0, 1, 2, \dots, N. \quad (\text{A.4})$$

B. Uncertainty quantification

In practice, in order to estimate the doubling times of an outbreak, one has to extract the unknown parameters by fitting model predictions to the epidemic data. Using these extracted parameters, one employs (2.6) to calculate the sequence of doubling times until cumulative cases cannot double anymore.

Table 4. 95% confidence intervals for true moment of time cumulative cases double (days) for synthetic data with $r = 1.2$, $K = 600$, $p=1$, and $a = 0.15$.

True Moment of time cumulative cases double	n	NA	EL	JEL
0.969	1	(0.957748,0.977638)	(0.958551,0.976556)	(0.958579,0.976554)
2.016	2	(1.992155,2.028633)	(1.993715,2.026684)	(1.993718,2.026680)
3.165	3	(3.129453,3.179136)	(3.131634,3.176540)	(3.131644,3.176530)
4.455	4	(4.414135,4.473760)	(4.416851,4.470673)	(4.416836,4.470694)
5.946	5	(5.905395,5.972138)	(5.908445,5.968766)	(5.908472,5.968749)
7.752	6	(7.714451,7.786448)	(7.717759,7.782801)	(7.717767,7.782793)
10.113	7	(10.074112,10.151846)	(10.077642,10.147870)	(10.077629,10.147860)
13.719	8	(13.673559,13.768391)	(13.677840,13.763624)	(13.677850,13.763615)
23.350	9	(23.162364,23.534539)	(23.184927,23.522837)	(23.184957,23.522801)

In order to quantify uncertainty in the extracted doubling times, we refit the model to $M = 400$ additional data sets for cumulative cases (assuming Poisson error structure), which results in M approximate sequences of doubling times. Our goal is to construct 95% confidence intervals for the mean doubling time at every moment where cases double, i.e., such confidence intervals where the true doubling time falls within the interval 95% of the time. Define the true doubling time at each

point where cases double as μ_{double} . We calculate and compare three different confidence intervals for estimating μ_{double} : Normal Approximation, Empirical Likelihood, and Jackknife Empirical Likelihood.

Table 5. 95% confidence intervals for interval of time between consecutive doubling of cases (days) for synthetic data with $r = 1.2$, $K = 600$, $p=1$, and $a = 0.15$.

True Interval of time Between Consecutive Doubling	n	NA	EL	JEL
0.969	1	(0.957748,0.977638)	(0.958551,0.976556)	(0.958579,0.976554)
1.047	2	(1.034359,1.051043)	(1.035081,1.050180)	(1.035091,1.050165)
1.149	3	(1.137063,1.150738)	(1.137697,1.150032)	(1.137689,1.150040)
1.290	4	(1.283906,1.295400)	(1.284455,1.294837)	(1.284426,1.294810)
1.491	5	(1.489251,1.500387)	(1.489739,1.499857)	(1.489765,1.499835)
1.806	6	(1.805016,1.818351)	(1.805718,1.817770)	(1.805707,1.817778)
2.361	7	(2.352960,2.372099)	(2.353951,2.371263)	(2.353953,2.371261)
3.606	8	(3.589092,3.626900)	(3.590984,3.625175)	(3.590991,3.625160)
9.631	9	(9.474994,9.779960)	(9.494552,9.771645)	(9.494577,9.771616)

Table 6. 95% confidence intervals for true moment of time cumulative cases double (days) for synthetic data with $r = 2.2$, $K = 1000$, $p=0.5$, and $a = 0.25$.

True Moment of time cumulative cases doubles	n	NA	EL	JEL
0.468	1	(0.462279,0.476864)	(0.463066,0.476257)	(0.463062,0.476260)
1.162	2	(1.138896,1.172211)	(1.140686,1.170805)	(1.140680,1.170821)
2.202	3	(2.177769,2.234873)	(2.180796,2.232448)	(2.180820,2.232471)
3.788	4	(3.746427,3.833379)	(3.751046,3.829665)	(3.751051,3.829685)
6.261	5	(6.189960,6.313948)	(6.196485,6.308599)	(6.196486,6.308594)
10.244	6	(10.139300,10.309129)	(10.148049,10.301598)	(10.148059,10.301585)
16.992	7	(16.853192,17.082037)	(16.864593,17.071438)	(16.864613,17.071418)
29.502	8	(29.323407,29.646882)	(29.339147,29.631349)	(29.339174,29.631322)
57.856	9	(57.507775,58.133745)	(57.540453,58.106540)	(57.540501,58.106465)

For the Normal Approximation method [15, 16], which is based on the Central Limit Theorem, with $M = 400$, we have that the distribution of the mean doubling time at each point is approximately normal. Let $X_1, \dots, X_M$ be an i.i.d random sample of doubling times at each moment cumulative cases double. Let $\bar{x}_{double}$ be the sample mean of $X_1, \dots, X_M$ at each moment cumulative cases double. Therefore, we calculate 95% confidence intervals for the mean doubling time, μ_{double} , by adding and subtracting 1.96 standard errors from the mean. In other words, our 95% confidence intervals for the

mean doubling time is calculated by:

$$(\bar{x}_{double} - 1.96 \sqrt{\text{var}(\bar{x}_{double})}, \bar{x}_{double} + 1.96 \sqrt{\text{var}(\bar{x}_{double})}) \quad (\text{B.1})$$

Empirical Likelihood is a non-parametric method for constructing confidence intervals without having to assume the form of the underlying distribution introduced by Owen [17,18]. Empirical likelihood enables us to successfully incorporate the advantages of the likelihood methods. Unnecessary assumption of a family of distributions can be avoided in empirical likelihood inference in the sense that the object is driven by data.

Table 7. 95% confidence intervals for interval of time between consecutive doubling of cases (days) for synthetic data with $r = 1.2$, $K = 600$, $p=1$, and $a = 0.15$.

True Interval of time Between Consecutive Doubling				
	n	NA	EL	JEL
0.468	1	(0.462279,0.476864)	(0.463066,0.476257)	(0.463062,0.476260)
0.694	2	(0.676606,0.695358)	(0.677604,0.694572)	(0.677607,0.694570)
1.041	3	(1.038819,1.062716)	(1.040080,1.061703)	(1.040086,1.061696)
1.586	4	(1.568465,1.598699)	(1.570054,1.597358)	(1.570040,1.597375)
2.473	5	(2.442930,2.481171)	(2.444849,2.479447)	(2.444861,2.479426)
3.982	6	(3.947514,3.997008)	(3.949911,3.994645)	(3.949921,3.994635)
6.748	7	(6.708214,6.778586)	(6.711649,6.775220)	(6.711656,6.775212)
12.510	8	(12.452887,12.582173)	(12.459753,12.576669)	(12.459761,12.576658)
28.354	9	(28.143750,28.527481)	(28.165397,28.513089)	(28.165422,28.513042)

Table 8. Time cumulative cases double with 95% confidence intervals for New Hampshire.

n	NA	EL	JEL
1	(1.296990,1.325518)	(1.298529,1.324277)	(1.298505,1.324302)
2	(2.900348,2.956805)	(2.903265,2.954363)	(2.903280,2.954331)
3	(4.884038,4.966410)	(4.888216,4.962692)	(4.888211,4.962693)
4	(7.340890,7.445314)	(7.346032,7.440461)	(7.346034,7.440458)
5	(10.388441,10.508826)	(10.394176,10.503061)	(10.394188,10.503050)
6	(14.177891,14.305781)	(14.183782,14.299462)	(14.183791,14.299455)
7	(18.909441,19.034630)	(18.915030,19.028286)	(18.915037,19.028282)
8	(24.862485,24.975516)	(24.867473,24.969776)	(24.867484,24.969766)
9	(32.465413,32.564050)	(32.469979,32.559271)	(32.469989,32.559260)
10	(42.492654,42.589418)	(42.497416,42.584906)	(42.497426,42.584896)
11	(56.776063,56.886604)	(56.781603,56.881411)	(56.781586,56.881409)
12	(82.431949,82.618957)	(82.441320,82.610516)	(82.441336,82.610500)

Empirical Likelihood has become very popular in recent years. Huang and Zhao [19] used empirical likelihood inference on a bivariate survival function under univariate censoring. Cheng et al. [20] introduced empirical likelihood inference for semi-parametric additive isotonic regression. In addition, Zhang et. al. [21] used empirical likelihood for testing the symmetry of statistical distributions. Let $X_1, \dots, X_M$ be an i.i.d. random sample of doubling times from an unknown distribution function $F(x) = P(X \leq x)$. For notation, define $F(x-) = P(X < x)$. Thus, $P(X = x) = F(x) - F(x-)$. For simplicity, we denote $P(X = X_i) = p_i$, $i = 1, \dots, M$, where $p_i \geq 0$ and $\sum_{i=1}^M p_i = 1$. The empirical cumulative distribution function (ECDF) of $X_1, \dots, X_M$ is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^M 1_{X_i \leq x}.$$

The notation $1_{X_i \leq x}$ outputs a value of 1 if $X_i \leq x$ is true, and 0 otherwise. In addition, given $X_1, \dots, X_M$, which are assumed independent with common CDF F_0 , the non-parametric likelihood of the CDF F is

$$L(F) = \prod_{i=1}^M (F(X_i) - F(X_i-)) = \prod_{i=1}^M p_i. \quad (\text{B.2})$$

The empirical maximum likelihood estimator (EMLE) based on the empirical likelihood, puts equal probability mass $1/M$ on the M observed values $X_1, X_2, \dots, X_M$. This allows us to calculate the empirical likelihood ratio. Using Lagrange multipliers, we can show that the p_i 's which maximize $L(F) = \prod_{i=1}^M p_i$ subject to $p_i \geq 0$, $\sum_{i=1}^M p_i X_i = \mu_{\text{double}}$, and $\sum_{i=1}^M p_i = 1$ are given by:

$$p_i(\mu_{\text{double}}) = M^{-1} (1 + \lambda(X_i - \mu_{\text{double}}))^{-1}, \quad i = 1, \dots, M. \quad (\text{B.3})$$

The empirical likelihood ratio is therefore:

$$R(F) = \frac{L(F)}{L(F_n)} = \frac{\prod_{i=1}^M p_i}{(1/M)^M} = \prod_{i=1}^M M p_i = \prod_{i=1}^M (1 + \lambda(X_i - \mu_{\text{double}}))^{-1}, \quad (\text{B.4})$$

where λ is the unique solution of

$$\sum_{i=1}^M \frac{X_i - \mu_{\text{double}}}{1 + \lambda(X_i - \mu_{\text{double}})} = 0.$$

Let $E(X_1^2) < \infty$. Then, the empirical loglikelihood ratio for the true mean doubling time is

$$l(\mu_{\text{double}}) = -2 \log(R(F)) = -2 \sum_{i=1}^M \log(M p_i) = 2 \sum_{i=1}^M \log(1 + \lambda(X_i - \mu_{\text{double}})). \quad (\text{B.5})$$

Lastly, let μ_0 be the true doubling time at a certain moment in time. We have as $n \rightarrow \infty$

$$-2 \log(l(\mu_0)) \xrightarrow{\mathcal{D}} \chi_1^2.$$

According to Owen [18], the $100(1 - \alpha)\%$ confidence interval for μ_{double} is given by

$$\{\mu_{double} : l(\mu_{double}) \leq \chi_{1,1-\alpha}^2\}, \quad (\text{B.6})$$

where $\chi_{1,1-\alpha}^2$ is the $1 - \alpha$ quantile of the χ_1^2 distribution.

There are some computational challenges for the application of empirical likelihood methods in practice. To avoid these, Jing et al. [22] introduced the “jackknife”, as a kind of re-sample method, and combined it with empirical likelihood and called it the Jackknife Empirical Likelihood (JEL). This method can greatly simplify the optimization procedures of traditional empirical likelihood methods. Various research articles have been published using jackknife empirical likelihood. Zhao et al. [23] used JEL inference for the mean absolute deviation. Sang et al. [24] proposed JEL method for estimating Gini correlations. Lin et al. [25] published the method for the error variance using JEL in linear regression models. In addition, the JEL method can also be applied to Bayesian inference, which was published by Cheng and Zhao [26].

Table 9. Time between consecutive doubling of cases with 95% confidence intervals for New Hampshire.

n	NA	EL	JEL
1	(1.296990,1.325518)	(1.298529,1.324277)	(1.298505,1.324302)
2	(1.603345,1.631299)	(1.604792,1.630013)	(1.604765,1.630041)
3	(1.983635,2.009660)	(1.984865,2.008437)	(1.984885,2.008417)
4	(2.456660,2.479097)	(2.457661,2.477939)	(2.457663,2.477955)
5	(3.046878,3.064184)	(3.047593,3.063270)	(3.047599,3.063260)
6	(3.786726,3.799679)	(3.787323,3.799023)	(3.787311,3.799037)
7	(4.721794,4.738605)	(4.722481,4.737708)	(4.722489,4.737693)
8	(5.932581,5.961348)	(5.933821,5.959773)	(5.933798,5.959794)
9	(7.574728,7.616734)	(7.576687,7.614647)	(7.576684,7.614652)
10	(10.001532,10.051077)	(10.004001,10.048830)	(10.004011,10.048820)
11	(14.265515,14.315081)	(14.267936,14.312761)	(14.267946,14.312752)
12	(25.628555,25.759683)	(25.635179,25.753758)	(25.635190,25.753718)

Table 10. Time cumulative cases double with 95% confidence intervals for Hawaii.

n	NA	EL	JEL
1	(2.198740,2.226115)	(2.200083, 2.224878)	(2.200113,2.224848)
2	(4.465612,4.515581)	(4.468100 ,4.513272)	(4.468110,4.513261)
3	(6.841237,6.909199)	(6.844618,6.906041)	(6.844624,6.906035)
4	(9.385318,9.466835)	(9.389377,9.463027)	(9.389371,9.463032)
5	(12.192842,12.283751)	(12.197337,12.279508)	(12.197345,12.279500)
6	(15.431831,15.528894)	(15.436600,15.524349)	(15.436610,15.524339)
7	(19.450330,19.554213)	(19.455392,19.549313)	(19.455394,19.549311)
8	(25.200992,25.331885)	(25.207450,25.325901)	(25.207461,25.325874)
9	(38.314156,38.720210)	(38.338153,38.706650)	(38.338161,38.706624)

Let $\hat{x}_{i,double}$ denote the mean doubling time calculated with the i^{th} observation deleted where $i = 1, \dots, M$. We define $\bar{x}_{double}$ as the mean doubling time of the entire i.i.d random sample $X_1, \dots, X_M$. Suppose that $\hat{V}_i$ denotes the jackknife pseudo-value, which is obtained by

$$\hat{V}_i = M\bar{x}_{double} - (M-1)\hat{x}_{i,double}, \quad i = 1, \dots, M. \quad (B.7)$$

Thus, the jackknife empirical likelihood evaluated at μ_{double} can be defined as

$$J(\mu_{double}) = \prod_{i=1}^M \left(1 + \lambda(\hat{V}_i - \mu_{double})\right)^{-1}, \quad (B.8)$$

where λ is the solution of the following equation

$$\sum_{i=1}^M \frac{\hat{V}_i - \mu_{double}}{1 + \lambda(\hat{V}_i - \mu_{double})} = 0. \quad (B.9)$$

With λ , we can now calculate the $-2 \log$ of the empirical likelihood ratio as

$$-2 \log(J(\mu_{double})) = 2 \sum_{i=1}^M \log \left(1 + \lambda(\hat{V}_i - \mu_{double})\right).$$

Let $E(X_1^2) < \infty$. Then, by Jing et al. by [22] the empirical loglikelihood ratio for the true mean doubling time is

$$-2 \log(J(\mu_{double})) = 2 \sum_{i=1}^M \log \left(1 + \lambda(\hat{V}_i - \mu_{double})\right). \quad (B.10)$$

Lastly, let μ_0 be the true doubling time at a certain moment in time. We have as $n \rightarrow \infty$

$$-2 \log J(\mu_0) \xrightarrow{\mathcal{D}} \chi_1^2.$$

Thus, the Jackknife Empirical Likelihood confidence interval for μ_{double} is obtained by

$$\{\mu_{double} : -2 \log(J(\mu_{double})) \leq \chi_{1,1-\alpha}^2\}, \quad (B.11)$$

where $\chi_{1,1-\alpha}^2$ is the $1 - \alpha$ quantile of χ_1^2 .

Table 11. Time between consecutive doubling of cases with 95% confidence intervals for Hawaii.

n	NA	EL	JEL
1	(2.198740,2.226115)	(2.200083, 2.224878)	(2.200113,2.224848)
2	(2.266848,2.289491)	(2.267973,2.288438)	(2.267975,2.288435)
3	(2.375497,2.393745)	(2.376394,2.392893)	(2.376398,2.392889)
4	(2.543588,2.558130)	(2.544302,2.557439)	(2.544298,2.557443)
5	(2.805798,2.818642)	(2.806437,2.818023)	(2.806423,2.818036)
6	(3.234070,3.250063)	(3.234861,3.249375)	(3.234882,3.249360)
7	(4.008355,4.035462)	(4.009760,4.034378)	(4.009787,4.034351)
8	(5.735666,5.792668)	(5.738761,5.790458)	(5.738777,5.790478)
9	(13.099598,13.401891)	(13.118536,13.393273)	(13.118561,13.393238)



AIMS Press

© 2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)