

SARS incubation and quarantine times: when is an exposed individual known to be disease free?

V. T. Farewell^{1,*}, A. M. Herzberg², K. W. James², L. M. Ho³ and G. M. Leung³

¹*MRC Biostatistics Unit, Cambridge, U.K.*

²*Queen's University, Kingston, Canada*

³*Department of Community Medicine, University of Hong Kong, Hong Kong Special Administrative Region, China*

SUMMARY

The setting of a quarantine time for an emerging infectious disease will depend on current knowledge concerning incubation times. Methods for the analysis of information on incubation times are investigated with a particular focus on inference regarding a possible maximum incubation time, after which an exposed individual would be known to be disease free. Data from the Hong Kong SARS epidemic are used for illustration. The incorporation of interval-censored data is considered and comparison is made with percentile estimation. Results suggest that a wide class of models for incubation times should be considered because the apparent informativeness of a likelihood depends on the choice and generalizability of a model. There will usually remain a probability of releasing from quarantine some infected individuals and the impact of early release will depend on the size of the epidemic. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: incubation times; log-gamma distribution; truncated distributions; SARS

1. INTRODUCTION

Control of infectious diseases is a major public health concern. After an individual's exposure to infection, opposing biological processes take place both in the infecting organism and in the host and these result in either that individual's development of clinical evidence of the disease or in an imperceptible host-victory. During this variable period of time, the individual

*Correspondence to: Vernon T. Farewell, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB4 2AP, U.K.

†E-mail: vern.farewell@mrc-bsu.cam.ac.uk

Contract/grant sponsor: Medical Research Council (UK)

Contract/grant sponsor: National Science and Engineering Research Council (Canada)

Contract/grant sponsor: Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau of the Hong Kong SAR Government

may in turn become infectious to others and thus play a part in generating or perpetuating an epidemic. Historically, attempts have been made to prevent and control epidemics by isolating, for an arbitrary period of time after which the biological struggle could be assumed complete, any individuals who might be incubating the disease.

The word 'quarantine', derived from the Latin word *quarisma*, means 40 and reflects the origin of the practice in the 40-day period of compulsory isolation of ships arriving in Venice in the 14th century. As more has been learned about the different infections, quarantine periods have varied, but when a hitherto unknown disease appears it is extremely difficult to decide what arbitrary period should be applied. And yet this is especially important if there should be no effective treatment for the disease or its infectious state. Controlling or preventing an epidemic then depends solely on releasing no infectious individuals into the general community. But, as was noted earlier, the period of unperceived changes in the individual is variable.

Quarantine was one of the key aspects of infection control introduced during the recent severe acute respiratory syndrome (SARS) epidemic. Individuals who may have been exposed to the SARS virus were quarantined for a fixed period of time, most commonly 10 days. The premise was that those who may have been exposed, but who showed no signs of illness after 10 days, were unlikely to come down with the disease. Since SARS was previously unknown, a quarantine policy offered the only control.

An important paper on epidemiological aspects of SARS was that of Donnelly *et al.* [1] which made use of data from the Hong Kong experience with SARS. The estimation of the incubation period in this paper was based on only '57 patients with only one exposure to SARS over a limited time scale with recorded start and end dates'. Donnelly *et al.* [1] assumed a gamma distribution for the incubation times, implicitly therefore assuming the possibility of very long incubation periods.

The work reported here arose from a question related to the confidence a community should have that an individual who has passed through the SARS quarantine period is disease-free and how long the quarantine period should be to make the probability of this very high. The concept of a maximum incubation time could be relevant to these considerations.

There are many issues to be considered in setting a quarantine time, for example the extent of disruption to individuals' lives. Also, and quite sensibly, it can be argued that there is unlikely to be a 'true' maximum incubation time. However, one motivation for a quarantine policy is the assumption that there is a reasonably well-behaved distribution of incubation times and some maximum time beyond which it is biologically quite implausible that symptoms may arise. This time could be the basis for setting a quarantine time.

Whether it is helpful to think about quarantine in this way is debatable. To inform this debate, we investigated what might reasonably be inferred about such a maximum incubation time based on the moderately sized samples that would typically be available in the early course of an epidemic. For comparison, brief consideration is also given to the estimation of tail behaviour in untruncated distributions. Our general premise is that careful specification of the available knowledge concerning the incubation distribution must be central to public health decisions to control epidemics. The work reported here should be viewed primarily as an exploration of statistical methodology that might be useful for this purpose, not as a critique of other approaches or specific estimates, such as those for SARS.

2. STATISTICAL MODELS

2.1. Gamma model

To illustrate the general principles involved, we follow Donnelly *et al.* [1] and consider a gamma distribution for incubation times. Thus if T is the random variable representing an incubation time, with an observed value of $T = t$, then a gamma distribution for T is specified by the probability density function

$$g(t) = \frac{1}{s^a \Gamma(a)} t^{(a-1)} e^{-(t/s)}$$

where $t > 0, a > 0$ and $s > 0$.

The expectation of this distribution is as and the variance is as^2 . However, we now introduce the assumption that this distribution is truncated at some time M , so that $0 < T < M$, and that the density function for T now becomes

$$f(t) = g(t)/G(M) \quad \text{where } G(M) = \int_0^M g(t) dt$$

Assume that data are available on n incubation times t_1, t_2, \dots, t_n . Maximum likelihood estimation of the parameters a, s and M can then be based on the likelihood function

$$L(a, s, M) = \frac{1}{G(M)^n} \prod_1^n g(t_i)$$

The profile likelihood for the parameter M is defined by

$$L_P(M) = L(\tilde{a}(M), \tilde{s}(M), M)$$

where $\tilde{a}(M)$ and $\tilde{s}(M)$ are the maximum likelihood estimates of a and s for a fixed value of M . It is clear that, for any fixed values of a and s , the likelihood is maximized when $G(M)$ is minimized and that therefore the maximum likelihood estimate (MLE) of M is at the largest t_i , usually denoted $t_{(n)}$.

Standard asymptotic distributional results for MLEs will not be applicable for the parameter M . In the consideration of inferential statements concerning M , there are parallels with Jeffreys' 'Bus problem' or more accurately, 'tramcar problem', raised in a letter to Fisher on 10 April 1934 [2, p. 163]. A brief summary is that in a town it is known that tramcars are numbered consecutively and that a new arrival in the town observes a tramcar numbered 100. Can the new arrival infer anything about the number of tramcars, say N , in the town? The problem can be extended by allowing the observation of more than one tramcar.

Jeffreys' considered the use of a prior proportional to $1/N$, after showing that a constant prior leads to no useful inferential statements. A very similar problem is the estimation of N in binomial (N, p) models. In both situations, the choice of the prior can be shown to be highly influential inferentially.

In the tramcar problem, the maximum observed number is the MLE for N and is sufficient for its estimation if a uniform distribution is assumed for the observed numbers. It is, however, a biased estimate. A unique unbiased estimate can be derived but the question of optimal interval estimation remains. For the binomial problem, it has been shown that no unbiased estimator of N exists [3].

For the purpose of this paper, we simply define the MLE of M and make no claims for its optimality in any sense. For public health purposes, the upper end-point of some interval of plausible values is more likely to be useful for decision making than a point estimate of the parameter. We consider the likelihood function simply as representing the information *available from the data* for inference concerning the unknown parameters. Comparison of the shape of the likelihoods is sufficient for the issues considered here and the likelihood function, particularly through providing ratios of likelihoods, is simply regarded as giving the relative plausibilities of parameter values [4, p. 50].

Since, by definition, it is true that for $M > t_{(n)}$, $L_P(M)$ can be defined for $t_{(n)} \leq M < \infty$, it will thus provide some indication of the values of M which are plausible given the observed data. It is frequently convenient to standardize this function so that the maximum value is one by dividing by the value of the likelihood function at the MLEs. This function can then be defined as

$$L_P^*(M) = \frac{L_P(M)}{L_P(\hat{M})}$$

where \hat{M} is the MLE of M .

2.2. Log-normal model

While the MLE of M will be the same irrespective of the distributional assumption made concerning T , the shape of the profile likelihood for M , and therefore the range of plausible values for M , will depend on the assumption and, in particular, on assumptions about the tails of the truncated distribution. While the gamma model is well known in epidemic theory, motivated by regarding the incubation period as a fixed number of independent and successive stages of infection, each exponentially distributed, alternatives to the gamma distribution should be considered from a model fitting perspective at least. For illustration, we consider the log-normal distribution. A log-normal regression model can be written as a location scale model $y = \log(t) = \alpha + \sigma e$, where e follows a standard normal distribution

$$f(e) = \frac{1}{(2\pi)^{0.5}} \exp(-0.5e^2)$$

and where $\alpha \in R$ and $\sigma > 0$.

The development of a truncated log-normal model follows the development for the truncated gamma given in Section 2.1 as does the likelihood development with (α, σ, M) replacing (a, s, M) as the set of model parameters. The use of this model is also considered in Section 3.

2.3. Log-gamma model

More general distributions than the truncated gamma and the truncated log-normal can also be considered. A convenient choice is the so-called log-gamma distribution of Farewell and Prentice [5] which represents a reparameterization and extension of a generalized gamma distribution.

With $\alpha, q \in R$ and $\sigma > 0$, the log-gamma model can be written as the location scale model $y = \log(t) = \alpha + \sigma w$, where the density $f(w; q)$ for w is

$$|q| (q^{-2})^{q^{-2}} \exp[q^{-2}\{qw - \exp(qw)\}]/\Gamma(q^{-2})$$

if $q \neq 0$ and, when $q=0$, is the standard normal distribution. The cumulative distribution function can be written as $1 - I(v; q)$ if $q > 0$ and $I(v; q)$ if $q < 0$, where $v = qw - 2 \log(|q|)$ and

$$I(v; q) = \{\Gamma(q^{-2})\}^{-1} \int_0^{\exp(v)} x^{q^{-2}-1} e^{-x} dx$$

The log-gamma distribution includes the Weibull ($q=1$) and exponential ($q=\sigma=1$) distributions as special cases as well as the gamma ($q=1/\sigma$) and log-normal ($q=0$). The distribution of W is negatively skewed for $q > 0$ and positively skewed for $q < 0$.

Another alternative truncated distribution for incubation times is, therefore, the truncated log-gamma. The development of a profile likelihood for M will follow as in Sections 2.1 and 2.2 with maximization over α, σ and q . The use of this more general distribution is also illustrated in Section 3.

3. EXAMPLE

We consider data from 128 SARS cases, a subset of 1755 cases in a Hong Kong Hospital Authority database, for which some information was available on time of infection. The data consist of the date of the appearance of the symptoms of SARS and an earliest and latest possible date of exposure. Initially, we restrict attention to 67 cases whose interval of possible exposure times is less than 5 days and also exclude 10 cases recorded as having first symptoms on the date of exposure. These may represent questionable records or cases related to an unusually high level of exposure, possibly hospital acquired, not of general relevance for setting quarantine times for controlling community outbreaks.

Relatively short intervals of exposure times are used to provide some reasonably precise information concerning incubation, as is done in AIDS seroconverter cohorts [6]. Table I provides some comparison of the 67 cases with infection intervals less than 5 days with the cases with longer intervals. The variables examined were age, sex, health care worker status, vital status on hospital discharge and lactate dehydrogenase (LDH) level, where higher values of LDH reflect more severe disease. It can be seen that while the cases are similar in age, sex and worker status, there is a higher death rate and some evidence of more severe disease in the cases with the longer possible infection intervals. This may reflect the fact that more severe cases arriving at a hospital might well have had a longer period with the disease and be less able to characterize precisely their possible time of infection. The impact of extending the allowed interval size is examined later. There remains, of course, the implicit assumption that the cases with some information on infection time are a random sample of the entire distribution of cases. However, the possibility of biases in reporting, heterogeneity in routes of transmission or varying infectious doses of the SARS coronavirus remains.

Table II presents the longest and shortest possible incubation times for these patients as well as the average of these two times, rounded to the nearest day since that is how the data would normally be recorded. We consider first the averaged times.

For the data set of averaged times, Figure 1 presents the profile likelihoods, $L_p^*(M)$, based on the gamma, log-normal and log-gamma models discussed in Section 2. Figure 1 allows the comparison of the apparent information in the data set under the different modelling assumptions.

Table I. Comparison of SARS cases.

	Censoring interval <5 days	Censoring interval ≥ 5	Significance level*
Health care worker			
No	49	39	0.84
Yes	18	13	
Sex			
Female	39	29	0.85
Male	28	23	
Survival			
Yes	59	35	0.007
No	8	17	
Mean age (years)	45.0	42.9	0.44
Mean LDH ($\mu\text{kat/L}$)	2.85	3.19	0.19

*Two sample comparison based on Fisher's exact test or *t*-test.

Table II. Incubation times.

Lower limits														
2.5	4.5	0.0	0.5	5.5	2.5	1.5	3.5	7.5	1.5	2.5	7.5	4.5	4.5	3.5
0.5	0.5	2.5	2.5	0.0	2.5	0.0	11.5	0.5	0.5	1.5	5.5	3.5	2.5	12.5
1.5	8.5	7.5	1.5	5.5	4.5	3.5	1.5	0.5	0.5	3.5	7.5	5.5	2.5	1.5
1.5	0.5	1.5	1.5	1.5	0.5	8.5	0.5	0.0	2.5	5.5	9.5	2.5	7.5	4.5
3.5	1.5	3.5	3.5	3.5	4.5	0.5								
Upper limits														
6.5	7.5	2.5	3.5	6.5	5.5	4.5	5.5	11.5	4.5	5.5	12.5	5.5	5.5	6.5
1.5	4.5	4.5	4.5	1.5	4.5	4.5	13.5	2.5	1.5	2.5	10.5	5.5	4.5	14.5
6.5	11.5	9.5	2.5	7.5	9.5	7.5	2.5	4.5	1.5	4.5	8.5	6.5	7.5	2.5
2.5	1.5	2.5	2.5	3.5	1.5	12.5	1.5	4.5	4.5	6.5	10.5	3.5	8.5	7.5
5.5	5.5	8.5	7.5	7.5	9.5	3.5								
Rounded average times														
5	6	1	2	6	4	3	5	10	3	4	10	5	5	5
1	3	4	4	1	4	2	13	2	1	2	8	5	4	14
4	10	9	2	7	7	6	2	3	1	4	8	6	5	2
2	1	2	2	3	1	11	1	2	4	6	10	3	8	6
5	4	6	6	6	7	2								

For the truncated gamma model, the profile likelihood never drops below 60 per cent suggesting that any value of M greater than the maximum time observed, 14, is plausible. Thus it appears that the data is uninformative with respect to the maximum incubation time if a gamma distribution is assumed. However, the situation is different for the truncated log-normal model. While the MLE for M is again 14 days for this model, any value for M greater than 19.5 days makes the data more than 10 times less plausible than does the MLE

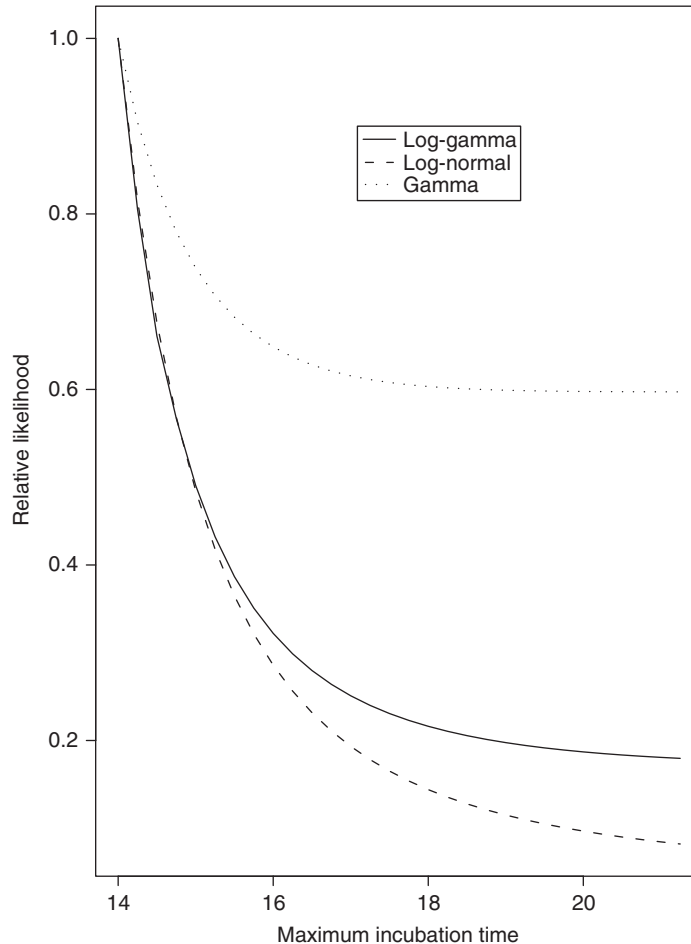


Figure 1. Profile likelihoods based on 67 SARS cases.

of 14 days. For public health purposes, it could therefore be argued that, based only on such data and an assumed log-normal model, that a quarantine time of 20 days might be necessary to ensure that SARS cases were not released ‘too early’. Recall that the focus here is on the upper limit of an interval of plausible values rather than any specific estimator for the maximum incubation time.

A possible reason for the widely different behaviour of the profile likelihood for the two models is a difference in model fit. If we consider the more general log-gamma model that includes both of the other models as special cases, the profile likelihood for M is more informative than that based on a gamma model, but it never falls below a value of 20 per cent. Thus the apparent ability to rule out larger values of M under the log-normal model is not present if a less restrictive model assumption is made. This is true even though the maximum likelihood estimate of q is -0.13 , a value close to the value $q=0$ corresponding

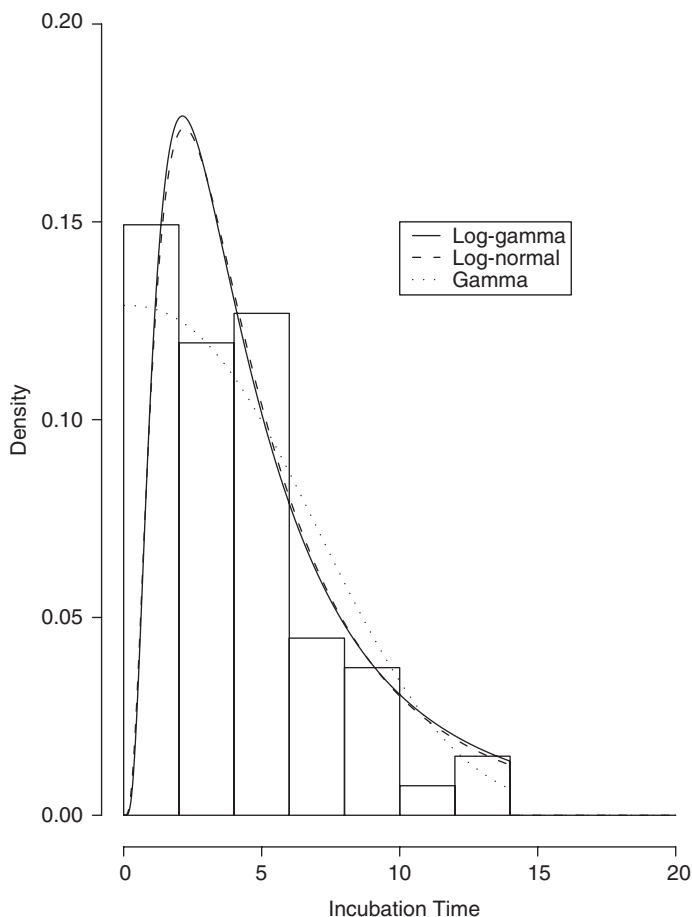


Figure 2. Maximum likelihood fits for truncated distributions.

to the log-normal model. The hypothesis of a truncated gamma distribution would not be supported within this class of models.

The maximum likelihood estimates of the various models are given in Figure 2 along with a histogram of the data. The estimated log-normal and log-gamma distributions are quite similar while the truncated gamma does not appear to fit the data very well. All the models fail to some extent in reflecting the preponderance of short incubation times.

Since the use of the log-gamma model suggests there is little information for the estimation of a maximum incubation time, this may raise doubts about the assumption of a truncated distribution. The reason for this is illustrated by the values of $\tilde{q}(M)$ over the range of the plotted profile likelihood for M . The MLE for q is -0.13 corresponding to $\tilde{q}(\hat{M})$, where $\hat{M} = 14$ days. However, $\tilde{q}(15) = 0.15$ and $\tilde{q}(20) = 0.55$ indicating that incubation distributions with less weight in the tails and with less effect therefore of truncation times are as consistent with the data as distributions with more weight in the tails and a truncation time.

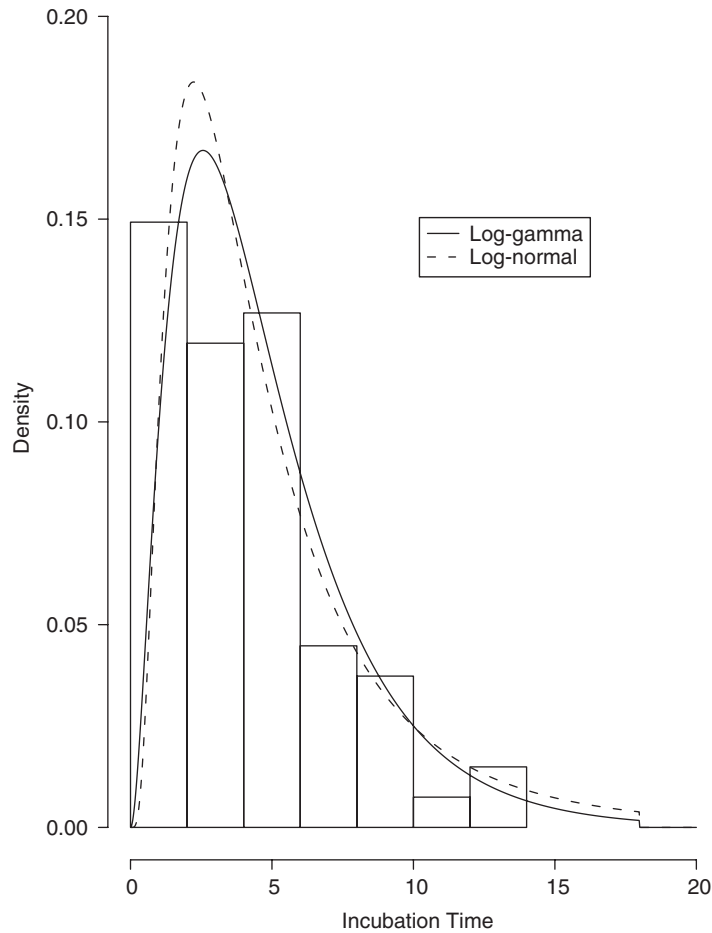


Figure 3. Maximum likelihood fits for $M = 18$.

To illustrate the different behaviour of the profile likelihoods for the log-normal and log-gamma models, Figure 3 plots the estimated log-gamma and log-normal models fit when the truncation time is taken to be $M = 18$ days. This shows that the lack of fit is much more pronounced for the log-normal distribution than for the log-gamma thus reducing the plausibility of $M = 18$ under the log-normal model.

4. INTERVAL-CENSORED INCUBATION TIMES

In general, as for the SARS cases in Hong Kong, it will be difficult to specify precisely when the exposure leading to a case occurred. As with many other diseases therefore, the usual data on incubation times will derive from cases in which the exposure is known to be within

a small window of time. This will generate interval-censored incubation time information for each case.

Assume that such information leads to a set of data $\{(t_{Li}, t_{Ui}), (i = 1, n)\}$ where, for individual i , the incubation time is known to lie between a lower limit, t_{Li} , and an upper limit, t_{Ui} . If we assume that $f(t) = g(t)/G(M)$ is the probability density function for the actual incubation time, as in Section 2 but where $g(t)$ is not restricted to be a gamma distribution, then, with only minor modifications, the development given there can be followed for interval-censored data.

The assumption of a maximum possible incubation time M creates some complication because it will limit the intervals of possible incubation times. It is also convenient to assume that all incubation times are interval-censored and that M is only allowed to take values greater than $\max(t_{Li})$. This avoids any possibility of a case contributing to the likelihood via its probability density function for the smallest value of M and via a probability value otherwise. In principle, other cases could be taken to have a known incubation time if such times were below any plausible values for M , but in practice such accuracy does not exist in any event. This type of consideration arises in other non-standard likelihood inference problems [7].

The likelihood function for the estimation of the parameters of $g(t)$ and M can then be written

$$\frac{1}{G(M)^n} \prod_1^n \{G(\min[M, t_{Ui}]) - G(t_{Li})\}$$

A profile likelihood for M can be defined in the usual manner. However, it is not possible to determine immediately the MLE, \hat{M} , which will lie somewhere between the lowest allowed value, $\max(t_{Li}) + \varepsilon$, and $\max(t_{Ui})$.

To illustrate the effect of interval-censoring, we consider the data in Table II which show the lower and upper limits of the incubation times for the 67 SARS cases, for which average rounded times were used in Section 3. We have subtracted 0.5 from the lowest time in days and added 0.5 to the highest to give appropriate intervals in continuous time and to make all observations interval-censored. As outlined earlier, it is convenient mathematically to make all observations interval-censored. Observations with a single day of presumed exposure are given an interval of width 1 day in our analysis but, in principle, a much narrower interval could be used if the precision could be justified. A brief exploration suggests that this would have little impact on the likelihood.

Figure 4 presents profile likelihoods for M based on the gamma, log-normal and log-gamma models of Section 2. These plots are based on calculations of the likelihoods for values of M at intervals of 0.25 and beginning at $\min(t_{Li}) + 0.5 = 13$. For convenience, the MLE of M has been taken to be the value among these which gives the largest likelihood. Further precision could be achieved but is not likely to be important. It can be seen that while the general pattern of the likelihoods is similar to that in Figure 1, with interval-censoring not even the log-normal likelihood drops to less than the 10 per cent level. This is, of course, reasonable in the sense that much less precise information is being assumed about the incubation times and this must impact the precision of inferences.

In spite of this slight, but perhaps important, change in the likelihoods, the fitted distributions are not much altered by the interval-censoring. For example, with the log-gamma model and

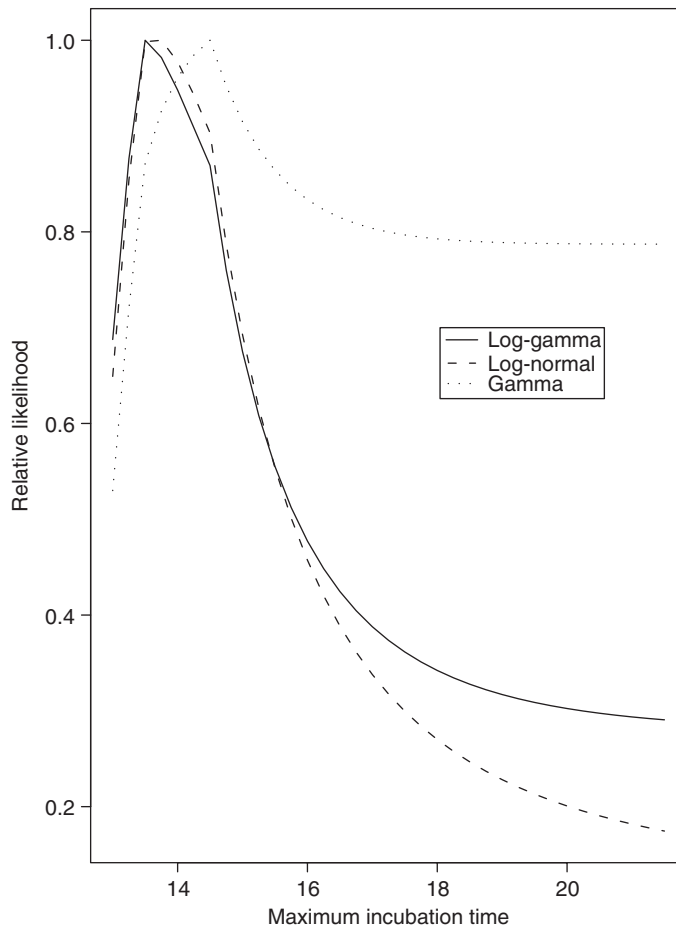


Figure 4. Profile likelihoods for interval censored data.

with $M = 14$, the estimated value for (α, σ, q) is $(1.42, 0.84, -0.13)$ for the rounded average data in Table II and $(1.33, 0.81, -0.10)$ the interval censored data in Table II.

Finally, to show the effect of more extreme interval-censoring, we consider extending the set of data in Table II by including additional SARS cases from Hong Kong whose period of possible exposure, which defines the width of the interval within which their incubation time lies, is thought to be less than 10 days rather than 5 days. This produces a set of data of 86 cases and Figure 5 presents the relative likelihoods for the three models based on these data. The profile likelihoods are seen to be substantially less informative with the gamma likelihood being virtually flat for M values greater than 16. Note that one of the additional cases has an interval of $(13.5, 19.5)$ for their incubation time in days. The use of censoring intervals of width 10 days is quite large in the context of SARS and could not be recommended in practice.

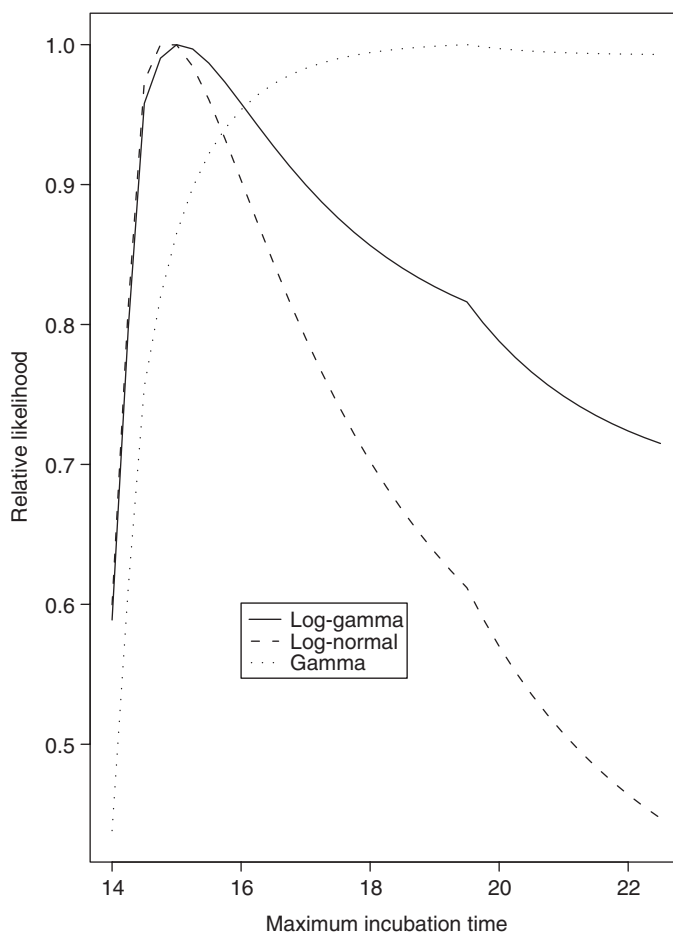


Figure 5. Profile likelihoods for extended interval censored data.

5. PERCENTILE ESTIMATION

Consideration of models for incubation times which incorporate truncation may provide valuable information for public health purposes. Nevertheless, as is illustrated in the earlier sections, there might often be insufficient evidence to be very confident about a maximum incubation time, even within the context of a particular model.

In this situation, an alternative approach is to set a quarantine time on the basis of percentile estimation, i.e. a quarantine time might be set as the time below which 95 per cent of cases are expected to develop. For comparison with the analyses presented earlier, the use of parametric models for this purpose is considered here.

Model choice will be important since the behaviour of a distribution in the tail is very model dependent. Thus, the log-gamma model which incorporates a significant component of model choice through the parameter q might be recommended. A more *ad hoc* approach to

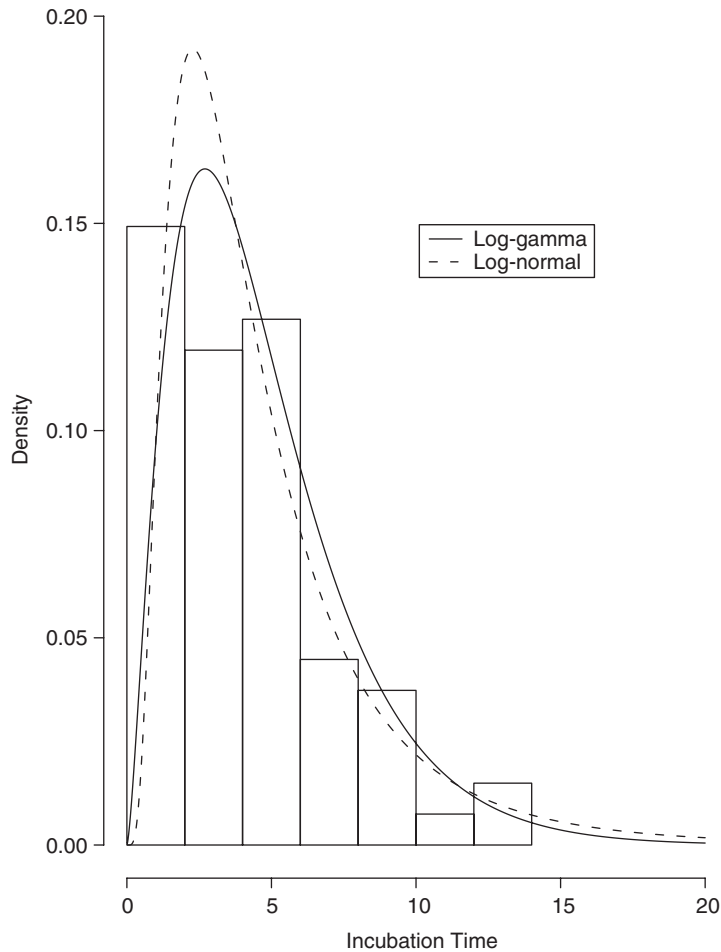


Figure 6. Maximum likelihood fits of untruncated distributions.

model choice could be adopted although the uncertainty involved in the choice might be more difficult to incorporate into inferences.

Figure 6 illustrates the best fitting log-gamma and log-normal distributions, not involving truncation, to the average incubation time data in Table II. The slightly better fit of the log-gamma at shorter times can be seen and there is some difference in the tails. For the log-gamma, the probability of an incubation time greater than 14 days is 0.013 while, for the log-normal, it is 0.032. The MLE for the log-gamma, in contrast to the case with truncated distributions, is further from the log-normal model with $\hat{q} = 0.61$. Essentially this reflects the need for the distribution to drop more quickly at larger values of T .

The estimated 95th percentiles for the log-gamma and log-normal distributions are 10.66 and 12.09. Confidence intervals for these values can be derived by simulating from the estimated asymptotic distribution of the MLEs to produce an interval within which 95 per cent, say,

of the corresponding simulated percentiles lie. This methodology has been compared with a delta method and a non-parametric bootstrap and performed well for the estimation of a complicated function of MLEs [8]. Based on a simulation of 1000 values, the corresponding 95 per cent intervals are (9.24, 13.68) and (9.95, 15.34). Interestingly, these values suggest the commonly adopted quarantine time for SARS of 10 days is associated with the possibility of 'releasing' approximately 5 per cent of patients 'too early'. In fact, to ensure that this is the maximum fraction released, consideration should be given to longer quarantine times reflecting the upper endpoint of the estimation intervals.

Note that if the interval-censored data in Table II is used to fit the log-gamma model, then the estimated 95th percentile is 10.2 with a confidence interval of (8.64, 13.68), an interval 14 per cent longer than that for the average data.

6. DISCUSSION

The present paper explores methodology to characterize the available knowledge on incubation times early in an infectious epidemic. Issues such as different routes of infection or different subsets of infectious individuals have not been discussed. In principle, the models used could be extended to incorporate explanatory variables defined by such factors. Preliminary investigations of possible explanatory variables in the Hong Kong data did not reveal any strong relationships.

We have made pragmatic decisions as to which data to include for model fitting. These might warrant revisiting in a more comprehensive analysis. Also, since infection events cannot be observed, some data on incubation times will inevitably be 'guesses'. Many aspects of the comparison of methodologies will not be altered by this but such data will naturally give rise to interval-censoring which the methodologies discussed here do allow. A further extension is to consider individuals with more than one period of possible exposure prior to the development of symptoms. Meltzer [9] considers a simple simulation approach to this.

Definitive conclusions about the choice of statistical methodology are not warranted based on the investigations reported here. In the early days of an epidemic this will usually be the case. Thus, the range of inferences based on different methodologies will often be the basis of decisions. Nevertheless, some comments can be made.

Inference concerning a truncation parameter is apparently more informative the stronger the assumptions made about the form of the incubation distribution. In the absence of independent reasons to make such an assumption however, the use of a general model, such as the log-gamma, for inference should be considered, at least as part of a sensitivity analysis. The key aspect to such inferences will be the shape of the tails encompassed in the model for the incubation times.

In the absence of precise information on a truncation time, estimation of percentiles provides a natural way to fix quarantine times. It can also be argued that this approach is less risky, and more realistic, than making the assumption of a truncated distribution. Because of its flexibility in the tails, the log-gamma can also be recommended for percentile estimation. Investigation of other methods is warranted. Possibilities would include the use of sample quantiles to define non-parametric confidence intervals for population quantiles [10, Chapter XI, Section 3.1] or the asymptotic distribution of sample quantiles [4, Appendix A.2.3]. Whatever method is adopted, the uncertainty involved in any estimation of percentiles should be incorporated into public health decisions.

In the setting of quarantine times, other factors must also be considered. Meltzer [9] presents evidence for some SARS incubation times greater than 10 days. It appears based on the data presented here that a quarantine time of 10 days for SARS might release one infectious patient in twenty. Therefore, for a quarantined population of 200, this would correspond to 10 individuals but the larger the quarantined population, the larger the number of released infectious individuals. Thus the length of a quarantine period might well be set in light of the expected number of quarantined individuals. Also consideration of the psychological and economic impact of quarantine on individuals and the population as a whole must be balanced against the risks associated with early release of infected individuals.

Finally, note that the implicit assumption in setting a quarantine time is that quarantine is isolation of x days from the *supposed* day of contact whereas it is often implemented as isolation of x days from the first day on which an individual is identified as having been exposed to the disease. This may build in an additional margin of safety from the public health perspective.

ACKNOWLEDGEMENTS

We thank the referees for their comments that led to an improved presentation. This work was supported by the Medical Research Council (U.K.), the National Science and Engineering Research Council (Canada) and the Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau of the Hong Kong SAR Government.

REFERENCES

1. Donnelly CA, Ghani AC, Leung GM, Hedley AJ, Fraser C, Riley S, Abu-Raddad LJ, Ho LM, Thach T-Q, Chau P, Chan K-P, Lam T-H, Tse L-Y, Tsang T, Liu S-H, Kong JHB, Lau EMC, Ferguson NM, Anderson RM. Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 2003; **361**:1761–1766.
2. Bennett JH (ed.). *Statistical Inference and Analysis. Selected Correspondence of R.A. Fisher*. Clarendon Press: Oxford, 1990.
3. Gupta AK, Nguyen TT, Wang Y. On maximum likelihood estimation of the binomial parameter n . *Canadian Journal of Statistics* 1999; **27**:599–606.
4. Cox DR, Hinkley DV. *Theoretical Statistics*. Chapman & Hall: London, 1974.
5. Farewell VT, Prentice RL. A study of distributional shape in life testing. *Technometrics* 1977; **19**:69–75.
6. Gore SM, Farewell VT. Extending public health surveillance of HIV disease. *Statistics in Medicine* 1993; **12**:2061–2064.
7. Matthews DE, Farewell VT. On a singularity in the likelihood for a change-point hazard rate model. *Biometrika* 1985; **72**:703–704.
8. Aalen OO, Farewell VT, De Angelis D, Day NE, Gill ON. A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Statistics in Medicine* 1997; **16**:2191–2210.
9. Meltzer MI. Multiple contact dates and SARS incubation periods. *Emerging Infectious Diseases* 2004; **10**: 207–209.
10. Mood AM, Graybill FA, Boes DC. *Introduction to the Theory of Statistics* (3rd edn). McGraw-Hill Kogakusha: Tokyo, 1974.