

# Significantly longer Covid-19 incubation times for the elderly, from a case study of 136 patients throughout China

Ally Bi-Zhu Jiang<sup>a</sup>, Richard Lieu<sup>b,\*</sup>, Siobhan Quenby<sup>c</sup>

<sup>a</sup>Shenzhen RAK wireless Technology Co., Ltd., China

<sup>b</sup>Department of Physics, University of Alabama, Huntsville, AL 35899, USA

<sup>c</sup>Division of Reproductive Health, Warwick Medical School, The University of Warwick, UK

---

## Abstract

Objective: To infer Covid-19 incubation time distribution from a large sample.

Method: Based on individual case data published online by 21 cities of China, we investigated a total of 136 COVID-19 patients who traveled to Hubei from 21 cities of China between January 5 and January 31, 2020, remained there for 48 hours or less, and returned to these cities with onset of symptoms between January 10 and February 6, 2020. Among these patients, 110 were found to be aged 15 – 64, 22 aged 65 – 86, and 4 aged under 15.

Findings: The differential incubation time histogram of the two age groups 15 – 64 and 65 – 86 are adequately fitted by the log normal model. For the 15 - 64 age group, the median incubation time of  $7.00_{-0.90}^{+1.10}$  days (uncertainties are 95 % CL) is broadly consistent with previous literature. For the 65-86 age group, the median is  $10.9_{-2.0}^{+2.7}$  days is statistically significantly longer. Moreover, for this group, the 95 % confidence contour indicates the data cannot constrain the upper bound of the log normal parameters  $\mu$ ,  $\sigma$  by failing to close there; this is because the sample has a maximum incubation time of 17 days, beyond which we ran out of data even though the histogram has not yet peaked. Thus there is the potential of a much longer incubation time for the 65-86 age group than 10 – 14 days. Only a much larger sample can settle this.

---

\*Corresponding author

Email address: [lieur@uah.edu](mailto:lieur@uah.edu) (Richard Lieu)

Preprint submitted to *Journal of L<sup>A</sup>T<sub>E</sub>X Templates*

April 15, 2020

---

## 1. Introduction

The incubation time of Covid-19 [1] and the closely related question of asymptomatic case numbers are two topics of major interest and concern. On the former, the research results presented here for the main age group of 15-64  
5 broadly corroborates previous studies [2, 3, 4, 5, 6], but for the elderly group of 65-86 years the incubation time we report below is significantly longer.

For nearly every city in China, daily information on list of COVID-19 cases is released officially to the Chinese social media WeChat accounts of respective cities. However, only a minority of cities include in their official release clear  
10 information on the day of symptom onset, which is required in estimating incubation period. We extracted the information so released between January 22 and February 15, 2020 and compiled a list of patients with COVID-19 who have traveled to Hubei, the origin of the COVID-19 pandemic, from cities which reported the day of symptom onset. The supplementary Excel file containing  
15 these raw data on COVID-19 cases reported from 21 cities of China outside Hubei: patient case number, age, sex, first and last day in Hubei, and first day with symptoms.

In the current investigation we included only those COVID-19 patients who stayed in Hubei for at most two calendar days. The day of exposure was taken  
20 as the first day to Hubei if the patient stayed in Hubei for one calendar day; or as the middle of the first and second day in Hubei if the patient stayed for two calendar days. By excluding COVID-19 patients who stayed in Hubei for more than two days, one can better define the the day of exposure. The incubation period for each COVID-19 patient is inferred as the number of days between  
25 exposure and symptom onset.

## 2. Log normal distribution

As will be shown below, the distribution of incubation times may adequately be fitted with a log normal distribution for the two age groups mentioned above,

suggesting that the incubation time  $\tau$  (in days) is a multiplicative variate. Specific to the problem of Covid-19, it seems reasonable to envisage an inverse proportionality relationship between the virus growth rate  $1+r$  and  $\tau$ , and the sample average growth rate is

$$\langle 1+r \rangle = \left[ \prod_{j=1}^N (1+r_j) \right]^{1/N} \quad (1)$$

where  $N$  is the sample size (or number of measurements of the incubation time  $\tau$ ). Since  $\tau_j \propto 1/(1+r_j)$  this means

$$\langle \tau \rangle \propto \left( \prod_{j=1}^N \tau_j \right)^{1/N} \quad (2)$$

or, equivalently, both  $\ln(1+r)$  and  $\ln \tau$  are additive variates. Moreover, if  $\ln(1+r_j)$  is normally distributed because  $1+r_j$  itself is a geometrically many times averaged growth rate of the virus inside the human body (*i.e.* the Central Limit Theorem may apply to  $\ln(1+r_j)$ ), the distribution of  $\tau_j$  would then be a log normal<sup>1</sup> of (arithmetic) mean  $\mu$  and standard deviation  $\sigma$ . Thus the expected number of cases within some incubation time interval  $k$ , or incubation time  $\tau_k = k\tau_0$ , is

$$n_k = \frac{n_0}{\sqrt{2\pi}\tau_k\sigma} e^{-(\ln \tau_k - \mu)^2 / (2\sigma^2)}, \quad (3)$$

with the coefficient  $1/\tau_k$  originating from the relationship between logarithmic and linear intervals, *viz.*  $d \ln \tau = d\tau/\tau$ .

The applicability of the log normal model to the Covid-19 incubation times distribution compels one to calculate the mean incubation time as the geometric mean (2) at least as an alternative, as we shall do in the following section.

---

<sup>1</sup>Under this assumption one can also derive (3) without enlisting the Central Limit Theorem, by consideration the viral interaction with the human body as a thermodynamic process with a fixed mean and variance, *i.e.* one which maximizes the Entropy  $-\sum_j p_j \ln p_j$  subject to the constraints  $p_j = 1$ ,  $\sum_j p_j \ln \tau_j = \mu$ , and  $\sum_j p_j (\ln \tau_j)^2 = \sigma^2 + \mu^2$ .

### 3. Fitting the data

The usual way of fitting a multi-parameter model to the data is by minimizing the  $\chi^2$  statistic

$$\chi^2 = \sum_k \frac{(n_k - \bar{n}_k)^2}{\sigma_k^2} \quad (4)$$

50 w.r.t. the model parameters  $\alpha, \beta, \gamma, \dots$ , where  $n_k \equiv n_k(\alpha, \beta, \gamma, \dots)$  is the number of cases for incubation time interval  $k$  as expected by the model,  $\bar{n}_k$  is the observed number of cases, and  $\sigma_k^2 = n_k$  is the expected model variance assuming Poisson fluctuation in the case counts. In the case of the currently scrutinized log normal model distribution consisting of three parameters  $\alpha = \mu$ ,  $\beta = \sigma$ , and  
55  $\gamma = n_0$  (where  $\mu$ ,  $\sigma$ , and  $n_0$  are as in (3)), 95 % confidence intervals  $\delta\mu$ ,  $\delta\sigma$ , and  $\delta n_0$  for the best-fit parameters may be inferred from the  $\Delta\chi^2$  criterion discussed below. For a log normal quantile  $q = q(m) = e^{\mu+m\sigma}$ ,  $\Delta\chi^2$  also yields  $\delta q$ . Yet an alternative method is to take advantage of the independence of  $\mu$  and  $\sigma$  in the model by writing

$$(\delta q)^2 = q^2[(\delta\mu)^2 + m^2(\delta\sigma)^2]. \quad (5)$$

60 It turns out, however, that the fitting technique of (4) leads to large uncertainties in some parameters of interest, including those computed within the dynamic range of incubation times as set by the data. This is especially the case for the second of the two age groups, consisting of age 65-86 years patients, where the incubation time span of the model is significantly wider than the data. Below  
65 we present a slight variation of the method in (4) which avoids the problem.

To facilitate introducing the modified model fitting algorithm we first remind the reader of the standard maximum likelihood method<sup>2</sup>, which relies upon the  $n_k \gg 1$  limit for all time intervals  $k$ , the limit where a Poisson distribution of counts tend to a Gaussian with equality between variance and mean. Thus the

---

<sup>2</sup>The maximum likelihood method only works when there is *independence* of measurement intervals. Thus it is *incorrect* to apply the model to fitting the data of a cumulative (or integral) distribution, in which the counts of previous intervals affect the later ones.

70 likelihood of a match between the number of counts  $n_k$  predicted by the model and observational data  $\bar{n}_k$  for *all* incubation time intervals  $k$  is given by the conditional probability

$$P(\text{data}|\text{model}) \propto \prod_{k=1}^m \exp \left[ -\frac{(n_k - \bar{n}_k)^2}{2n_k} \right], \quad (6)$$

and one's task is to maximize  $P(\text{data}|\text{model})$  w.r.t. the model parameters  $\alpha, \beta, \gamma, \dots$ , where  $m$  is the total number of time intervals spanned by the data. If  
75 the number of counts per bin does not satisfy  $n_k \gg 1$ , however, the distribution of counts will not be normal, *i.e.* it will not be genuinely Poisson, in which case the expression on the right side of (6) is not strictly valid. In this limit, which does apply to the second of our two age groups (*viz.* 65 – 86 of age), one should use likelihood ratios [7] instead of (6), but because the difference between the  
80 two output parameter values are small compared to their uncertainties there is no real advantage in deviating from (6); we simply feel that for completeness sake this subtle point should be mentioned.

The maximization of (6) w.r.t. model parameters is obviously the same as minimizing  $\chi^2$  as given in (4). Explicitly, if one writes, in the context of the log  
85 normal model (3),

$$\chi^2(\mu, \sigma, n_0) = \sum_k \frac{[n_k(\mu, \sigma, n_0) - \bar{n}_k]^2}{n_k(\mu, \sigma, n_0)}. \quad (7)$$

the procedure would be equivalent to solving

$$\left. \frac{\partial \chi^2}{\partial n_0} \right|_{\mu, \sigma} = \left. \frac{\partial \chi^2}{\partial \mu} \right|_{n_0, \sigma} = \left. \frac{\partial \chi^2}{\partial \sigma} \right|_{n_0, \mu} = 0. \quad (8)$$

Thus there are 3 equations in 3 unknowns, and the minimization process is fully deterministic.

If, as mentioned above, the total incubation time  $m\tau_0$  spanned by the entire  
90 database is not long enough to clinch the full extent of the log normal distribution, one will have to constrain the fitting procedure to ensure that the area under the log normal is *exactly* equal to the total number of cases  $N$  over the time  $m\tau_0$ . The specific question one seeks to answer here is: given there are

$N$  cases to be randomly distributed into  $m$  time intervals in accordance with a  
95 prescribed set of average proportions  $\{n_k\}$ ,  $k = 1, 2, \dots, m$  satisfying

$$n_1 + n_2 + \dots + n_m = N, \quad (9)$$

and  $n_k = n_k(n_0, \mu, \sigma)$ , how would one tune  $n_0$ ,  $\mu$ , and  $\sigma$  to maximize the likelihood of the hypothetical distribution agreeing with the data, when Poisson counting uncertainties in the latter are taken into account? Note that the model for  $n_k$  does not have to cut off at  $k = m$ , *i.e.* (9) is merely there to enforce  
100 the equality between expected and actual total number of cases within the full range of incubation times available to the study. In this way, one is obliged to respect only those ‘in range’ parameter values ensuing from the best fit model.

Thus one would now extremize the statistic

$$F(n_0, \mu, \sigma, \lambda) = \chi^2(\mu, \sigma, n_0) + \lambda \left[ \sum_k n_k(\mu, \sigma, n_0) - N \right], \quad (10)$$

by requiring

$$\left. \frac{\partial F}{\partial n_0} \right|_{\mu, \sigma, \lambda} = \left. \frac{\partial F}{\partial \mu} \right|_{n_0, \sigma, \lambda} = \left. \frac{\partial F}{\partial \sigma} \right|_{n_0, \mu, \lambda} = \left. \frac{\partial F}{\partial \lambda} \right|_{n_0, \mu, \sigma} = 0, \quad (11)$$

105 where the vanishing of the last partial derivative enforces the area ( $\lambda$  is a Lagrange multiplier). Once again there are 4 equations solving for 4 unknowns, and the number of free parameters is reduced from the previously 3 to currently 2 (note however that the area constraint is not as simple as fixing  $n_0$ ; note also that the degrees of freedom of the whole problem is increased from the  $m - 3$   
110 to  $m - 2$ ).

Turning to the confidence interval for an interested parameter  $\eta(n_k) = \eta(\mu, \sigma, n_0)$ , such as the expected (arithmetic) mean incubation time within the observation interval  $m\tau_0$ ,

$$\eta = \frac{1}{N} \sum_{k=1}^m n_k \tau_k, \quad (12)$$

one could re-extremize  $F$  subject to yet another additional constraint which  
115 ensures  $\eta$  equals some fixed value  $\eta_0$  by invoking one more Lagrange multiplier

$\nu$  to form the  $G$  statistic,

$$G = \chi^2(\mu, \sigma, n_0) + \lambda \left[ \sum_k n_k(\mu, \sigma, n_0) - N \right] + \nu[\eta(\mu, \sigma, n_0) - \eta_0] \quad (13)$$

and requiring

$$\left. \frac{\partial F}{\partial n_0} \right|_{\mu, \sigma, \lambda, \nu} = \left. \frac{\partial G}{\partial \mu} \right|_{n_0, \sigma, \lambda, \nu} = \left. \frac{\partial G}{\partial \sigma} \right|_{n_0, \mu, \lambda, \nu} = \left. \frac{\partial G}{\partial \lambda} \right|_{n_0, \mu, \sigma, \nu} = \left. \frac{\partial G}{\partial \nu} \right|_{n_0, \mu, \sigma, \lambda} = 0. \quad (14)$$

The resulting increase in  $\chi^2$  w.r.t. (11),  $\Delta\chi^2$ , is also  $\chi^2$  distributed with one degree of freedom, because the extra constraint enforced by  $\nu$  has likewise increased the degrees of freedom by one<sup>3</sup>. Thus, to obtain the 95 % confidence interval in  $\eta$ , one needs to find the value of  $\eta_0$  such that (14) leads to a  $\chi^2$  increase of  $\Delta\chi^2 = 3.8$ . This procedure applies if  $\eta$  is a mean, variance, or quantile, or any other attribute of the distribution.

By adopting the aforementioned procedure, we obtained the best-fit parameters as shown in Table 1, the goodness of fit in Figures 1 and 2, cumulative frequencies in Figures 3 and 4, and 95 % confidence contours in Figure 5, for the two age groups 15-64 and > 65.

#### 4. Conclusion

The two age groups being analyzed are clearly distinct samples. For the 15 - 64 age group, the median incubation time of  $7.00_{-0.90}^{+1.10}$  days (uncertainties are 95 % CL, see Table 1) is broadly consistent with previous measurements [2, 3, 4, 5, 6]. For the 65-90 age group, the median is  $10.9_{-2.0}^{+2.7}$  days is statistically significantly longer. The other equally importantly point is, as revealed by the open confidence contour in Figure 5b, the inability of the 65-86 age group data in constraining the upper bound of the model parameters  $\mu, \sigma$ . This is because the sample of  $N = 22$  cases here has a maximum incubation time of 17 days,

<sup>3</sup>To elaborate, the degrees of freedom equals the number of independent data points minus the number of free parameters in the model. If the last of the three quantities is reduced by one because of the incorporation of the  $\nu$ -related constraint, the first will be increased by one.

Parameter	Age 15-64	Age 65-86
Sample size $N$	110	22
$\mu$	1.95, 2.26, 2.95	2.50, 4.10, –
$\sigma$	0.68, 0.83, 1.20	0.55, 1.11, –
Mean (geometric)	6.09, 6.85, 7.62	8.57, 10.07, 11.54
Mean (arith.)	7.32, 8.08, 8.88	9.60, 11.05, 12.32
Standard deviation (arith.)	3.90, 4.22, 4.35	3.62, 4.09, 4.80
Median	6.10, 7.00, 8.10	8.9, 10.9, 12.6
0.05 quantile	1.5, 1.7, 2.1	1.65, 3.40, 7.30
Lower quartile ie	3.40, 4.10, 4.80	4.80, 7.40, 9.30
Upper quartile	9.5, 10.7, 11.9	12.6, 14.0, 15.0
0.9 quantile	12.9, 14.1, 14.8	15.1, 15.8, 16.3
0.95 quantile	14.7, 15.4, 15.9	16.3, 16.4, 16.6

Table 1: Parameters of the best log normal fit to the two age groups.  $\mu$  and  $\sigma$  are as defined in (3); while the other parameters are calculated by applying the best-fit model to the incubation time range of 17 days, which is the full range spanned by the data (true for both age groups). The information on each parameter comprises an expectation value sandwiched between the lower and upper uncertainty limit, both of which are 95 % confidence (note that for the 65+ age group their upper  $\mu$  and  $\sigma$  uncertainties are not constrained by the data, because the data have not revealed the other side of the peak of the differential probability distribution, see Figures 2 and 5).

beyond which one ran out of data even though the differential case histogram has not yet peaked (Figure 3). This indicates the potential of much longer Covid-19 incubation time for age 65-86 years old patients. It should also be pointed out that we attempted another model, *viz.* the Gumbel (or log Weibull) distribution, to see if the situation improves, but the goodness-of-fit turns out to be worse than the log normal (see the caption of Figure 3 for details). Indeed, only a larger sample for this age group can settle the question of the true median incubation time by clinching the tail of the distribution, equivalently by closing the 95 % CL contour of Figure 5b.



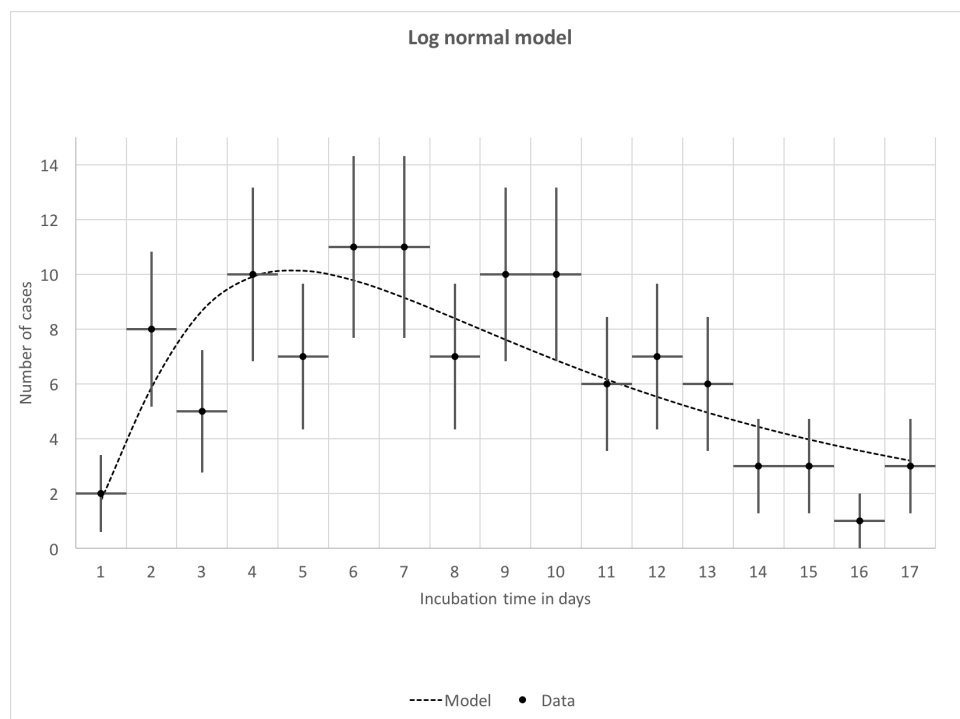


Figure 1: Best fit log normal distribution of the 15-64 age group is plotted against the data. Vertical error bars are the Poisson  $\sqrt{n_k}$  uncertainties in the counts, while horizontal bars mark the duration of 1 day for each time bin. The goodness-of-fit is given by  $\chi^2 = 9.44$  for 15 d.o.f., or a 85.3 % probability of rejecting the null hypothesis (namely the hypothesis that the match between data and model occurred purely by random chance).

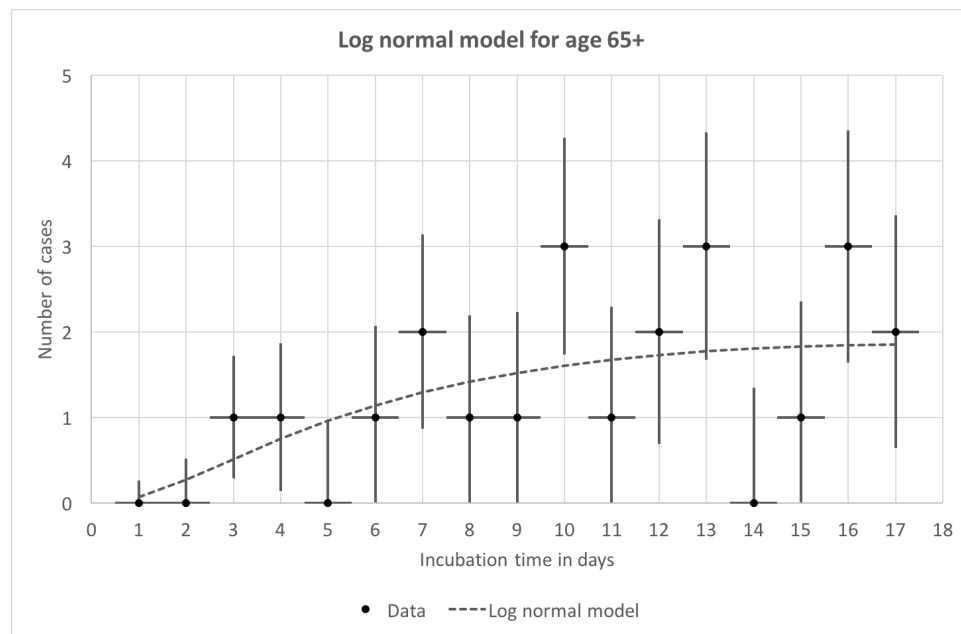


Figure 2: Best fit log normal distribution of the 65-86 age group is plotted against the data. Vertical error bars are the Poisson  $\sqrt{n_k}$  uncertainties in the counts, while horizontal bars mark the duration of 1 day for each time bin. The goodness-of-fit is given by  $\chi^2 = 7.85$  for 15 degrees of freedom, or a 93.0 % probability of null hypothesis rejection (for comparison the Gumbel (log Weibull) model scored, under the same fitting criteria,  $\chi^2 = 8.17$ , or 91.7 % probability of null hypothesis rejection). Note however that for this age group the data have not revealed the other side of the peak of the differential probability distribution.

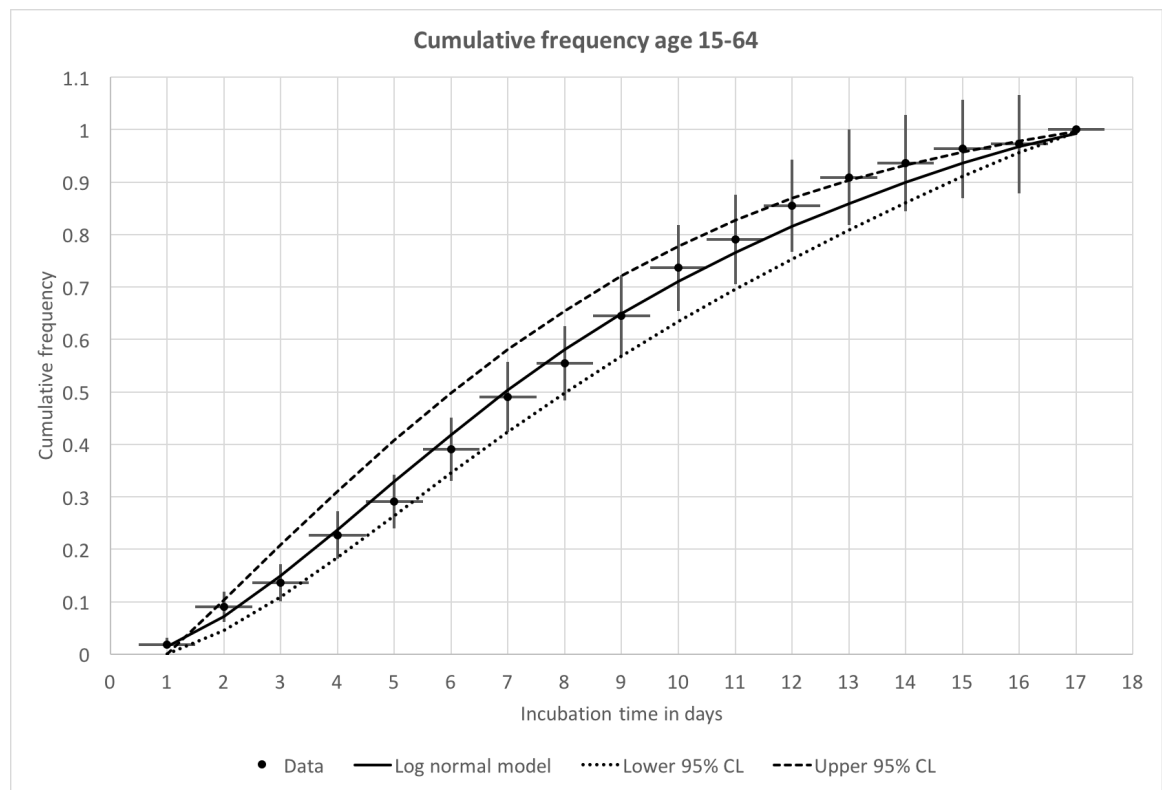


Figure 3: Cumulative frequency distribution for the 15-64 age group. Solid line is the best-fit log normal model for the differential distribution, while dotted lines mark the 95 % confidence (or  $\Delta\chi^2 = 3.8$  for one degree of freedom, see text) uncertainties of the model. Vertical error bars for the data are Poisson counting fluctuations ( $1-\sigma$ , and correlated among the time bins because of the accumulation of counts), while horizontal bars mark the duration of 1 day for each time bin.

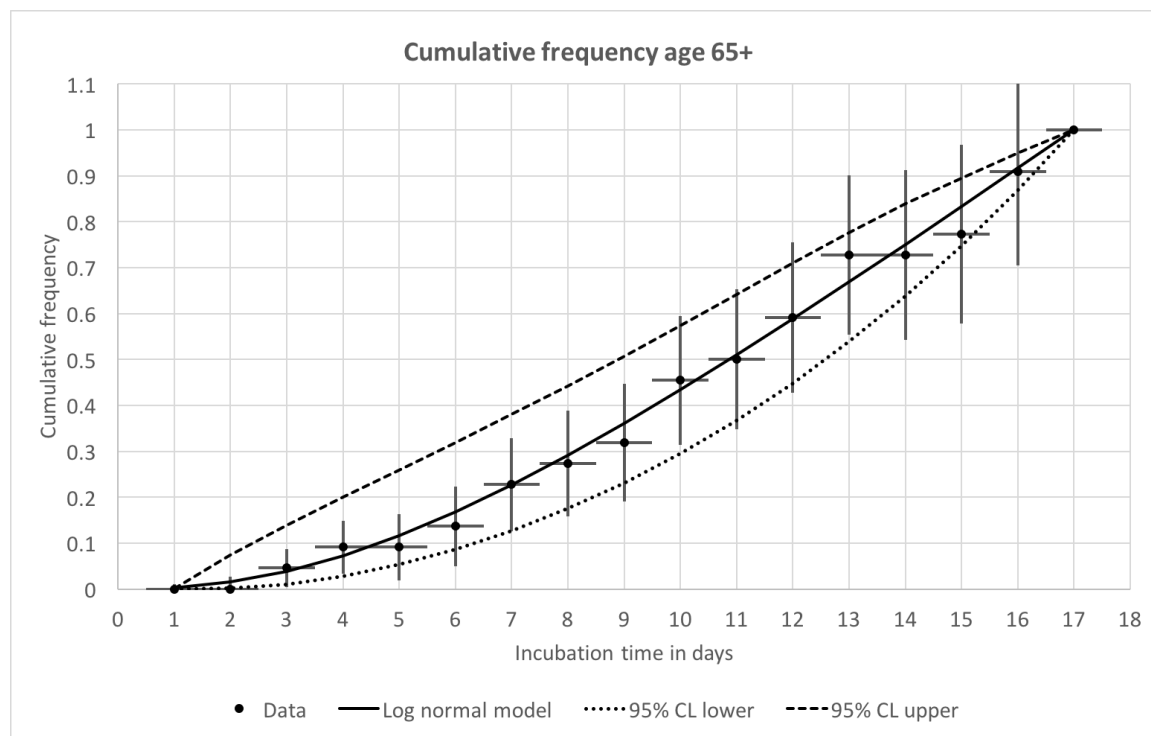


Figure 4: Cumulative frequency distribution for the 15-64 age group. For more information see the caption of the previous graph.

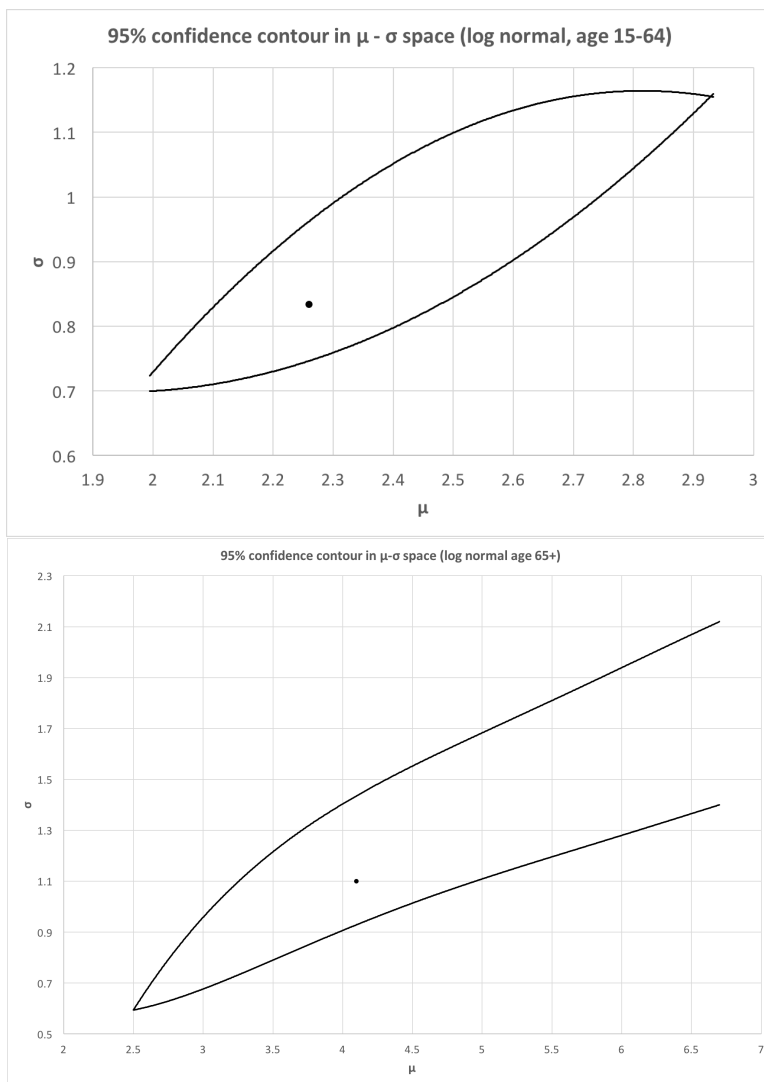


Figure 5: 95 % (*i.e.*  $\chi^2 + 3.8$ ) confidence contours identifying the likelihood of pairs of  $\mu$  and  $\sigma$  values. The central black dot locates the best fit model. Note the upper limits of both model parameters are unconstrained by the data in the case of the 65-86 age group.

- [1] World Health Organization, 2020, Global research on Coronavirus diseases (Covid-19) <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/>
- [2] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with  
150 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395:497- 506. [PMID: 31986264] doi:10.1016/S0140-6736(20)30183-5
- [3] Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20  
28 January 2020. *Euro Surveill*. 2020;25. [PMID:32046819] doi:10.2807/1560-  
155 7917.ES.2020.25.5.2000062
- [4] Linton NM, Kobayashi T, Yang Y, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right  
truncation: a statistical analysis of publicly available case data. *J Clin Med*.  
2020;9. [PMID: 32079150] doi:10.3390/jcm9020538
- [5] Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of  
160 novel coronavirus-infected pneumonia. *N Engl J Med*. 2020. [PMID: 31995857]  
doi:10.1056/NEJMoa2001316
- [6] Lauer, S.A., Grantz, K.H., Bi, Qifang, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases:  
165 Estimation and Application. *Annals of Internal Medicine*, doi:10.7326/M20-0504
- [7] Cash, W., Parameter Estimation in astronomy through application of the likelihood ratio. *ApJ*, 1978; 228, 939