To the editor: I read with great interest the article by Jantien A Backer et al. entitled `Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020' [1]. Backer et al. fitted three parametric distributions for the incubation period: the Weibull distribution, the gamma distribution, and the lognormal distribution. As a result, Backer et al. showed that the Weibull distribution provided the best fit to the data, evaluated by the leave-one-out information criterion (loo-ic) [2]. However, I found that their model comparison method had a mistake.

Backer et al. estimated the moment of infection for each case using a uniform prior distribution over the exposure intervals. This means estimating one latent variable (the moment of infection) from one sample. In such cases, the interpretation of the loo-ic is very difficult. This is because the likelihood is evaluated using latent variables that are not observed. Unobserved latent variables should be marginalized (Supplementary Materials S1).

In addition, in this problem, the duration of exposure and the date of onset are known, but the date of infection is unknown. This problem is the same as observations called interval censoring in the field of survival analysis [3].

I conducted a simulation study for the comparison between Backer et al.'s method and the method using the marginalization.

Loo-ic is aimed at approximating the generalization error, but the Backer et al.'s method has a larger bias than the marginalized case (see Supplementary Materials S2). Here, bias indicates the difference between the generalization error and loo-ic.

This marginalization has another advantage. If the date of first exposure (i.e. for Wuhan residents) was not provided in the data table, it was arbitrarily selected by Backer et al. This method of selection may have affected the results of the analysis. When marginalizing latent variables, it is easy to integrate to infinity, and selection of the starting points for exposure is not necessary.

I analyzed the same dataset as that used for Backer et al.'s study using the marginalization. (Supplementary Materials S3). Table 1 shows the loo-ic of the three probability distributions. In contrast to the results of Backer et al., the gamma distribution gave the best fit. However, The difference between the gamma and lognormal distributions is small, so their comparison is not significant.

Table 1. Loo-ic of the three probability distributions.

Distribution	Loo-ic
Weibull	73.889
gamma	73.317
log-normal	73.323

Table 2 shows the 95% prediction intervals of the incubation period.

Table 2. 95% (2.5th to 97.5th percentile) prediction intervals of the incubation period evaluated by the three probability distributions.

	2.5%	50%	97.5%
Weibull	2.52	6.86	12.11
gamma	2.97	6.55	12.89
log-normal	2.66	6.78	18.99

Backer et al. reported an incubation period ranging from 2.1 to 11.1 days. However, my analysis suggests that the incubation period may be longer. The true distribution of the incubation periods is unknown, and model selection is an eternal challenge in statistics. However, in my opinion, the incubation period of 2019-nCoV is likely to be longer than that which was established by Backer et al. I hope that these values will assist in determining the appropriate quarantine period for 2019-nCoV.

Authors' contributions

KA wrote the manuscript.

Conflict of interest

None declared

References

1. 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(5).
2. Vehtari A., Gelman, A., & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput. 2017; 27(5): 1413-32.
3. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Stat Soc Series B Stat Methodol. 1976;38(3): 290-5.