LETTER

Letter to the editor: Generation interval for COVID-19 based on symptom onset data

Sergio Bacallado¹ , Qingyuan Zhao¹ , Nianqiao Ju²
1. Statistical Laboratory, University of Cambridge, Cambridge, United Kingdom
2. Department of Statistics, Harvard University, Cambridge, MA, United States

Correspondence: Sergio Bacallado (sb2116@cam.ac.uk)

To the editor: The purpose of this letter is to raise a few issues with the methodology of a recent paper by Ganyani et al. [1], which aimed to estimate the parameters of the generation interval of coronavirus disease (COVID-19) using records from Tianjin, China and Singapore at the start of the pandemic.

Dependence of serial intervals

The likelihood function (in the final displayed equation of the section Methods, Model) suggests that an underlying assumption is that the serial intervals Z_i, $i=2,...,n$ are independent and identically distributed. However, this assumption does not follow from the independence assumptions made in the paper.

Consider a transmission chain $i \rightarrow j \rightarrow k$. Following the authors' notation, the serial interval between cases i and j is $Z_j = X_j + Y_j$ where X_j is the generation interval for the transmission $i \rightarrow j$ and $Y_j = \delta_j - \delta_i$ is the difference of the incubation periods of i and j. Similarly, $Z_k = X_k + Y_k$. The paper assumes that the generation times X_j and X_k are independent, which may be reasonable. The paper further assumes that generation times and incubation periods are independent, which implies that X_j and Y_j are independent. We argue that the latter assumption is questionable (see below), but let us suppose that it holds. Given these two assumptions, it is clear that the serial intervals $Z_j = X_j + \delta_j - \delta_i$ and $Z_k = X_k + \delta_k - \delta_j$ are not independent because the incubation period of case j is shared, leading to a *negative* dependence. Similarly, there is a *positive* dependence between Z_j and Z_k in a transmission fork $i \rightarrow j$, $i \rightarrow k$. Hence the likelihood function used in Ganyani et al. [1] may be described as an approximate likelihood.

Metropolis–Hastings sampler for the infection tree

The Metropolis–Hastings sampler in the script MCMC_generation_interval.R shared with the original article has a proposal distribution which generates trees by sampling a random infector for each non-index case from a list of possible infectors. With the Singapore data, this almost always produces a graph with a cycle, and not a tree, which is inconsistent with the model. Sampling trees consistent with the data uniformly would not be straightforward unless constraints are put on negative generation intervals.

By using a Markov chain Monte Carlo algorithm which augments the state space with the infection time for each case, it is possible to sample the posterior distribution of the infection tree and generation interval parameters using the correct likelihood function. We provide a simple Gibbs sampler which implements this strategy in a Github repository [2].

Independence of generation interval and incubation period

Consider a transmission $v(i) \rightarrow i$. In order to express the density of Z_i as a convolution of X_i and Y_i , a key assumption is that the incubation period $\delta_{v(i)}$ is independent of the generation time $X_i = t_i - t_{v(i)}$. Note that the paper only stated the assumption as "assuming the incubation period is independent of the infection time", which we interpret as assuming X_i is independent of both δ_i (reasonable) and $\delta_{v(i)}$ (questionable). The latter independence would clearly not be true if the patient became infectious only after showing symptoms (i.e. $X_i \geq \delta_{v(i)}$). Even though the transmission of COVID-19 can be presymptomatic, it still seems difficult to defend the assumption that infectivity is independent of symptom onset biologically.

We believe it is important in further work to test this assumption, or study the sensitivity of the conclusions to deviations from it. Ganyani et al. report rates of pre-symptomatic transmission based on the theoretical model, assuming independence of the generation interval and incubation periods. It would be instructive to, in addition, impute the infection times in the datasets from the posterior distribution, and report the rate

of pre-symptomatic transmission among the cases in each dataset.

Conflict of interest

None declared.

Authors' contributions

All three authors contributed equally to producing the research. Sergio Bacallado and Qingyuan Zhao wrote the letter.

References

1. Ganyani T, Kremer TC, Chen D, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Euro Surveill. 2020;25(17):2000257. PMID: 32372755
2. Bacallado S. Covid_tree. San Francisco: GitHub; 6 Jun 2020. Available from: https://github.com/bacallado/covid_tree/

License, supplementary material and copyright