Using symptom onsets instead of incidences results in time-lagged estimate of $R(t)$

Authors: Phyllis Ju, Sergio Bacallado, Qingyuan Zhao

The instantaneous reproduction number $R(t)$ is defined as the average number of secondary cases that would be generated by a primary case infected on day t if conditions remained the same after that day [1]. $R(t)$ can be estimated by the number of new infections on day t divided by the effective infectivity on day t of the individuals already infected [1, 2].

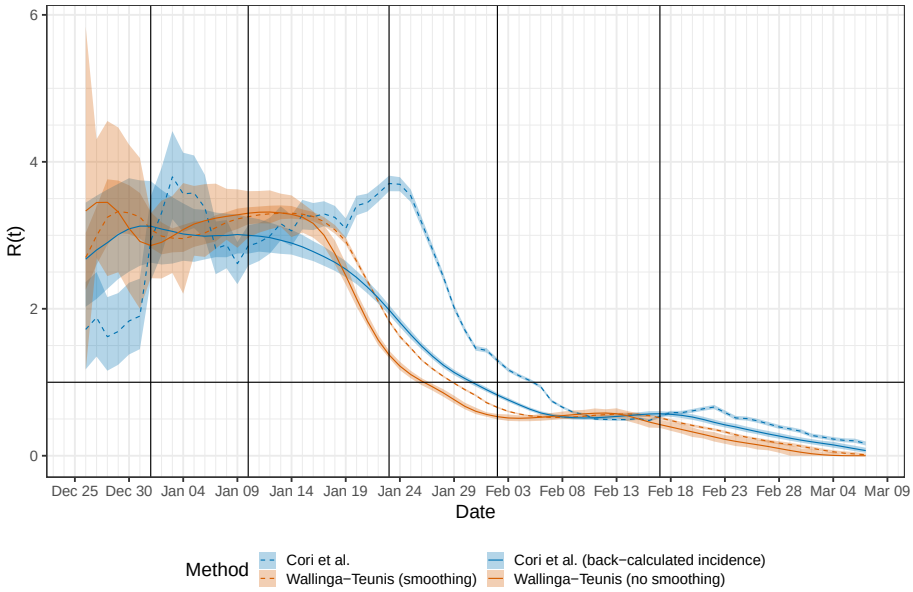
Although $R(t)$ is defined in terms of and ideally should be estimated using new infections, times of infection are rarely observed in practice. Instead, Cori et al. [2] suggested that their method can also be applied with the infection incidence curve replaced by the more commonly observed symptom onset curve. But this comes with an important caveat: the estimated $R(t)$ would have a time lag equal to the incubation period. They suggested that a possible strategy to correct the time lag is to use the incubation period distribution to back-calculate the curve of incidences from the observed curve of symptom onsets.

We followed this suggestion and reanalyzed the (digitized) epidemic curve in Figure 1 of Pan et al. To capture the uncertainty of the imputed incidences, we used 1000 Monte Carlo samples of the incidence curve instead of the point estimator suggested by Fraser [1]. To simulate the incidence, we used a Gamma-distributed incubation period with a median of 4.5 days and 95% quantile of 13.4 days, estimated in a previous study of ours [4]. For each simulated incidence curve, we obtained one posterior sample of $R(t)$ using the method of Cori et al. [2]. We then pooled the posterior samples across the 1000 simulations to assess the uncertainty of the estimated $R(t)$. To ease comparison, we replicated the analyses in Pan et al. and the comment by Lipsitch et al. which used the method of Wallinga and Teunis [3]. Code for the analysis and results can be found in https://github.com/phyllisju/rt.

Not surprisingly, back-calculating the incidence leads to an $R(t)$ curve quite different from the one in Pan et al. The new curve is visually smoother before Period 2 and close to the curve obtained by Lipsitch et al. The new $R(t)$ curve starts to decrease around January 15 and dips below 1 on January 31, before the implementation of some more aggressive public health measures on February 2. The earlier comment by Lipsitch et al. and response by Pan et al. attributed their difference in the estimated curves of $R(t)$ to the "forward-looking" and "backward-looking" nature of the statistical methods they used. Our re-analysis shows that another contributing factor is the time lag of the estimated $R(t)$, occurred when the method of Cori et al. [2] is applied with symptom onsets instead of (back-calculated) incidences.

We would like to stress three further points. First, instead of the instantaneous reproduction number $R(t)$, the method of Wallinga and Teunis [3] in fact targets the case reproduction number which can be viewed as a weighted average of $R(t)$ after t [1]. Second, the method of Wallinga and Teunis [3] is also commonly applied with symptom onsets instead of incidences, resulting in

a similar time lag with the actual case reproduction number. Finally, Cori et al. [2] did not adopt a fully Bayesian approach and instead used a sliding window method to smoothen $R(t)$. Caution is thus needed to interpret its credible intervals, especially in time periods where $R(t)$ changes sharply.



References

[1] C. Fraser. Estimating individual and household reproduction numbers in an emerging epidemic. *PLOS One*, 2(8), 2007.

[2] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez. A New framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, 2013.

[3] J. Wallinga and P. Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*, 160(6):509–516, 2004.

[4] Q. Zhao, N. Ju, S. Bacallado, and R. D. Shah. BETS: The dangers of selection bias in early analyses of the coronavirus disease (COVID-19) pandemic, 2020. arXxiv:2004.07743.