Clinical onset serial interval and diagnostic serial interval of SARS-CoV-2/COVID-19 in South Korea

Sofia K. Mettler[1,2,*]

Marloes H. Maathuis[1]

[1] Seminar for Statistics, Department of Mathematics, ETH Zurich

[2] Faculty of Medicine, University of Zurich

[*] Correspondence: Sofia K. Mettler <sofiakyonhi.mettler@uzh.ch>

BACKGROUND The clinical onset serial interval, or the time between the onset of symptoms in successive cases in a chain of infection, is often used as a measurable proxy for the transmission serial interval of an infectious disease. Current estimates of the mean clinical onset serial interval of COVID-19 range from 3.96 to 7.5 days. In this article, we define the diagnostic serial interval as the time between the diagnosis dates of the infector and infectee. We study and compare the clinical onset and diagnostic serial intervals of SARS-CoV-2/COVID-19 in South Korea.

METHODS Analyzing the DS4C project data which summarize information on SARS-CoV-2/COVID-19 cases reported by regional governments in South Korea, we estimate the means of the clinical onset serial interval, the diagnostic serial interval and the difference between the two. We use the balanced cluster bootstrap method to construct 95% bootstrap confidence intervals.

RESULTS The mean clinical onset serial interval and mean diagnostic serial interval were estimated to be 3.58 days (95% CI: 2.62, 4.53) and 3.68 days (95% CI: 3.14, 4.22), respectively. A matched sample analysis showed that the diagnostic serial interval was significantly shorter than the clinical onset serial interval (estimated mean difference -1.17 days, 95% CI: -2.26, -0.09).

CONCLUSIONS The short diagnostic serial interval of SARS-CoV-2/COVID-19 in South Korea may explain why South Korea was able to contain the COVID-19 outbreak and avoid high mortality. We conjecture that the mean diagnostic serial interval may serve as a predictor for the success of a country's containment efforts.

Introduction

The serial interval of an infectious disease, also known as the generation time, is defined as the time between analogous phases in successive cases of a chain of infection [1]. The *transmission serial interval*, or the time between the infection events of the infector and infectee, is particularly important as it is one of the factors that determine how rapidly the disease can spread in the community, but it is difficult to measure [2]. The

clinical onset serial interval, a commonly used form of serial interval for its easier measurement, refers to the time between the onset of symptoms of the infector and the infectee [2]. Table 1 summarizes recent findings on the clinical onset serial interval of COVID-19.

Table 1: Estimated mean or median of the clinical onset serial interval of COVID-19 with the corresponding 95% confidence interval (CI) or credible interval (CrI).

	Li et al [3]	Nishiura et al [4]	Du et al [5]	Ki [6]
Parameter of Interest	Mean	Median	Mean	Mean
Estimate	7.5 days	4.0 days	3.96 days	6.6 days
95% CI or CrI	CI: (5.3, 19)	CrI: (3.1, 4.9)	CI: (3.53, 4.39)	-
Data	6 pairs	28 pairs	468 pairs	9 pairs
	Wuhan, China	meta-analysis[1]	93 Chinese cities	South Korea

In this article, we study the clinical onset serial interval of COVID-19 using a rich data set of infector-infectee pairs observed in South Korea between January 20th and April 7th, 2020. We also introduce and study the *diagnostic serial interval*, which we define as the time between the infector's diagnosis and the infectee's diagnosis. In the case of SARS-CoV-2/COVID-19, diagnosis is assumed to be laboratory confirmation of SARS-CoV-2 infection. While the clinical onset serial interval is largely a characteristic of the pathogen, the diagnostic serial interval contains additional information on public health policy and capacity.

Data and Methods

Data and inclusion criteria

A total of 10'384 individuals were confirmed to be infected with SARS-CoV-2 in South Korea between January 20th and April 7th [7]. The seventeen regional governments of South Korea have published information on infected individuals, including date of diagnosis, age, gender, infection route, infector (if known) and symptom onset date (if symptomatic and reported). Among the seventeen regional governments, fourteen have published information on over 85% of their confirmed cases, the exceptions being Daegu (publishing information on less than 1% of 6'803 patients as of April 7th), Jeju-do and Gangwon-do [7, 8]. The DS4C project summarizes the information published by the regional governments and makes these data available for public use [8]. Our analysis is based on the DS4C project data (as of April 8th), containing information on 3'127 individuals.

Among these 3'127 individuals, there were 728 known infector-infectee pairs. Date of diagnosis was recorded for all infectors and all infectees in these pairs. In order to reduce bias resulting from right truncation, we

[1] Infector-infectee information from China, Germany, South Korea, Singapore, Taiwan and Vietnam

excluded 29 pairs for which the infector's diagnosis date was later than March 31st. The remaining 699 infector-infectee pairs, containing 310 unique infectors and 699 unique infectees, were used to study the diagnostic serial interval.

Twelve of the seventeen regional governments have sporadically reported symptom-related information including the symptom onset dates. Among the 728 known infector-infectee pairs, there were 76 pairs for which the onset date of symptoms was known for both the infector and the infectee. These 76 pairs were composed of 38 unique infectors and 76 unique infectees. As no infector's symptom onset date was later than March 31st, all 76 pairs were kept for analysis. These data were used to study the clinical onset serial interval and the difference between the diagnostic and clinical onset serial intervals.

An inclusion flowchart is provided in the supplement.

Statistical Methods

We estimated the mean of the clinical onset serial interval based on the 76 pairs with known symptom onset dates. Similarly, we estimated the mean of the diagnostic serial interval based on the 699 pairs described above. We computed 95% confidence intervals for these parameters using the bootstrap method. To account for dependencies between observations caused by common infectors, we used the balanced cluster bootstrap method which regards each infector as a cluster [9].

For the 76 infector-infectee pairs for which we know both the clinical onset serial interval and the diagnostic serial interval, we also computed the difference Δ between these two intervals (Δ = diagnostic serial interval − clinical onset serial interval). We estimated the mean of the distribution of these differences, and used the balanced cluster bootstrap method to construct the corresponding 95% bootstrap confidence interval.

All computations were performed using the statistical software `R`. All `R` code is provided in the supplement.

Results

Figure 1 visualizes the relationship between infectors' and infectees' symptom onset dates (Figure 1(a)) and diagnosis dates (Figure 1(b)). The data points below the diagonal line in Figure 1(a) show that some infectees developed symptoms before their infectors, indicating the presence of asymptomatic transmission. A similar phenomenon occurs for the diagnosis dates in Figure 1(b), but to a lesser extent.

Histograms of the clinical onset serial intervals and diagnostic serial intervals are shown in Figure 2(a). The distributions appear to have different shapes. Also, as already indicated in Figure 1, we see some negative serial intervals, especially for the clinical onset serial interval. Precisely, among the 76 pairs for whom the clinical onset serial interval was calculated, 9.2 percent of the infectees developed symptoms before their infectors. Among the 699 pairs for whom the diagnostic serial interval was calculated, 2.1 percent of the

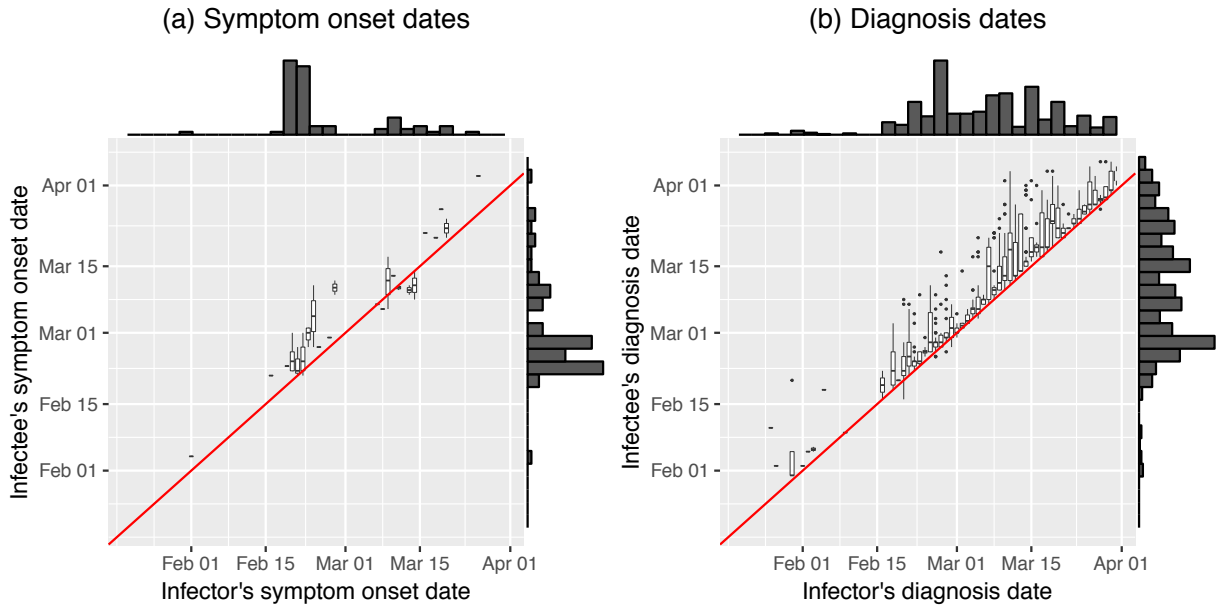

Figure 1: (a) Relationship between the symptom onset dates of infector-infectee pairs. Each boxplot displays the symptom onset dates of infectees corresponding to a given infector's symptom date. Histograms of the symptom onset dates of infectors and infectees are displayed at the horizontal and vertical margin, respectively. (b) Relationship between the diagnosis dates of infector-infectee pairs, analogous to (a).

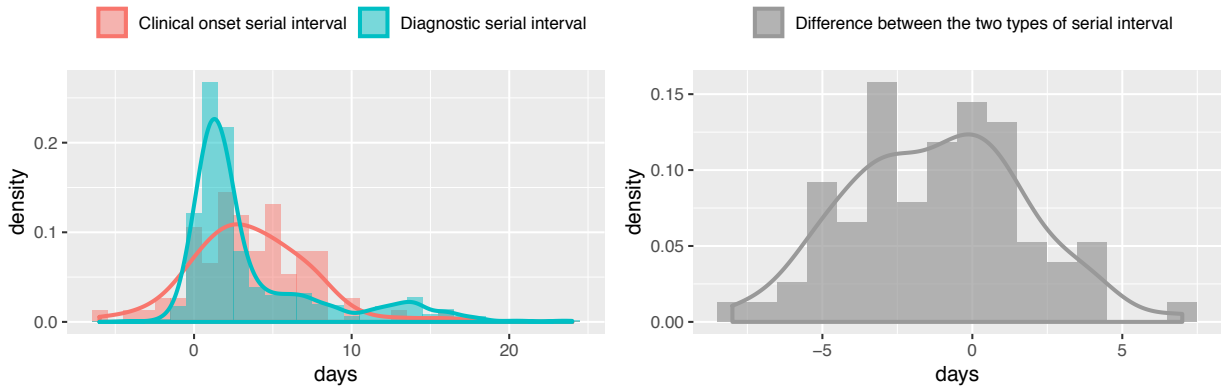

Figure 2: (a) Histograms of clinical onset serial intervals (76 infector-infectee pairs, red) and diagnostic serial intervals (699 infector-infectee pairs, blue) in days. Lines show the estimated density functions. (b) Histogram of the difference (Δ) between the clinical onset and the diagnostic serial intervals (76 infector-infectee pairs).

infectees were diagnosed before their infectors. Among the 76 pairs for whom both intervals were calculated, 56.6 percent of the pairs had a shorter diagnostic serial interval than clinical onset serial interval.

Table 2 shows the estimated means of the clinical onset serial interval, the diagnostic serial interval, and their difference Δ in days. The 95% confidence interval for the mean diagnostic serial interval is shorter than that for the mean clinical onset serial interval, due to the larger data set that can be used for the diagnostic serial intervals. The estimated means of the clinical onset serial interval and the diagnostic serial interval are similar (3.58 vs 3.68), but were estimated based on different data sets. The matched sample analysis based on the 76 pairs yielded an estimated mean difference Δ of -1.17 days (95% CI: -2.26, -0.09). This difference is significant at the $\alpha = 0.05$ level, indicating that the diagnostic serial intervals are significantly shorter than their corresponding clinical onset serial intervals for SARS-CoV-2/COVID-19 in South Korea.

Table 2: Estimated means of the clinical onset serial interval (first row, based on 76 pairs), the diagnostic serial interval (second row, based on 699 pairs) and their difference Δ (third row, based on 76 pairs) of SARS-CoV-2/COVID-19 in South Korea, including 95% bootstrap confidence interval (in days).

	Estimated mean (95% confidence interval)
Clinical onset serial interval	3.58 (2.62, 4.53)
Diagnostic serial interval	3.68 (3.14, 4.22)
Difference Δ	-1.17 (-2.26, -0.09)

Discussion

Estimating the transmission serial interval in the context of a pandemic is important because it provides the time window for the containment of the spreading infection. Nishiura et al [10] pointed out that the shorter the true generation time of SARS-CoV-2/COVID-19 is, the more undiagnosed infected individuals there will be. A particular challenge of SARS-CoV-2/COVID-19 is that these undiagnosed individuals may spread the virus resulting in an extraordinarily fast growing pandemic [4, 10, 11].

As infection dates are typically unknown, the transmission interval is very hard to obtain. We therefore studied the commonly used clinical onset serial interval of COVID-19 in South Korea. We also introduced and studied a new type of serial interval, the diagnostic serial interval. While the clinical onset serial interval is largely a characteristic of the pathogen, we argue that the diagnostic serial interval contains additional information on public health policy and capacity.

Our estimate of the mean clinical onset serial interval of 3.58 days (95% CI: 2.62, 4.53) is comparable to the findings of Nishiura et al [4] and Du et al [5] (see Table 1) and similar to the results for the diagnostic serial interval (estimated mean: 3.68 days, 95% CI: 3.14, 4.22). According to the matched sample analysis of 76

pairs for which we had both the clinical onset and the diagnostic serial intervals, the diagnostic serial interval was significantly shorter than the clinical onset serial interval (estimated mean difference: -1.17 days, 95% CI: -2.26, -0.09).

The relatively short diagnostic serial intervals in South Korea, compared to the clinical onset serial intervals, are likely due to the country's broad testing policy, laboratory capacity and its intensive efforts in contact tracing. In countries with more restrictive testing policies (e.g. symptoms such as fever or cough and pre-existing health conditions are prerequisites), we expect the diagnostic serial intervals to be longer.

Short diagnostic serial intervals imply faster identification of new infection events and therefore point to a better containment of SARS-CoV-2/COVID-19. We therefore pose that the diagnostic serial interval may be used as an early predictor for the success of a country's containment efforts. It will be interesting to study this in other countries and investigate the predictive power of this indicator.

Besides its relevance for public health policy, the diagnostic serial interval has the advantage that it is easier to estimate than the clinical onset serial interval, since dates of diagnosis are routinely recorded and less subjective than the onset of symptoms. Moreover, in contrast to the clinical onset serial interval, it is defined for asymptomatic carriers of SARS-CoV-2. This is especially favorable as asymptomatic transmission appears possible [11], and a significant portion of asymptomatic carriers remain asymptomatic [12, 13].

Our study has several limitations, the main one being that our data do not constitute a random sample of the population of infected individuals. As a result, we may suffer from selection bias in various ways. For example, certain regions are strongly underrepresented, one of which being the province of Daegu which was severely affected by COVID-19. Also, some selection bias for shorter intervals is likely to exist, as infector-infectee pairs with shorter serial intervals may be identified more easily than those with longer serial intervals.

Reference

1. Porta E. (2016) A Dictionary of Epidemiology, 6th ed [Online]. Oxford University Press.
2. Fine PE. (2003). The Interval between Successive Cases of an Infectious Disease. Am J Epidemiol 158:1039–1047.
3. Li Q et al. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med 382:1199-1207.
4. Nishiura H et al. (2020). Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases 93:284–286.
5. Du Z et al. (2020). Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. Emerg Infect Dis 26(6).
6. Ki M. (2020). Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Korea. Epidemiology and Health 42.
7. Korea Centers for Disease Control & Prevention. (April 8th 2020). Updates on COVID-19 in Republic of Korea.
8. Kim J. (2020). Data Science for COVID-19 (DS4C). Retrieved April 8, 2020 from https://www.kaggle.com/kimjihoo/coronavirusdataset.
9. Davison AC and Hinkley DV. (1997) Bootstrap Methods and Their Application. Cambridge University Press.
10. Nishiura H et al. (2020). The Rate of Underascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights. Journal of Clinical Medicine 9,419.
11. Rothe C et al. (2020). Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 382(10).
12. Nishiura H et al. (2020). Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). International Journal of Infectious Diseases [Journal Pre-proof].
13. Zhou X et al. (2020). Follow-up of the asymptomatic patients with SARS-CoV-2 infection. Clinical Microbiology and Infection: https://doi.org/10.1016/j.cmi.2020.03.024.