

# Potential Role of Social Distancing in Mitigating Spread of Coronavirus Disease, South Korea

Sang Woo Park, Kaiyuan Sun, Cécile Viboud, Bryan T. Grenfell, Jonathan Dushoff

In South Korea, the coronavirus disease outbreak peaked at the end of February and subsided in mid-March. We analyzed the likely roles of social distancing in reducing transmission. Our analysis indicated that although transmission might persist in some regions, epidemics can be suppressed with less extreme measures than those taken by China.

The first coronavirus disease (COVID-19) case in South Korea was confirmed on January 20, 2020 (1). In the city of Daegu, the disease spread rapidly within a church community after the city's first case was reported on February 18 (1). Chains of transmission that began from this cluster distinguish the epidemic in South Korea from that in any other country. As of March 16, a total of 8,236 cases were confirmed, of which 61% were related to the church (1).

The Daegu Metropolitan Government implemented several measures to prevent the spread of COVID-19. On February 20, the Daegu Metropolitan Government recommended wearing masks in everyday life and staying indoors (2). On February 23, South Korea raised its national alert level to the highest level (1) and delayed the start of school semesters (3). Intensive testing and contact tracing enabled rapid identification and isolation of case-patients and reduction of onward transmission (4). We describe potential roles of social distancing in mitigating COVID-19 spread in South Korea by comparing metropolitan traffic data with transmission in 2 major cities.

## The Study

We analyzed epidemiologic data describing the COVID-19 outbreak in South Korea during January

20–March 16. We transcribed daily numbers of reported cases in each municipality from Korea Centers for Disease Control and Prevention (KCDC) press releases (1). We also transcribed partial line lists from press releases by KCDC and municipal governments. All data and code are stored in a publicly available GitHub repository (<https://github.com/parksw3/Korea-analysis>).

We compared epidemiologic dynamics of COVID-19 from 2 major cities: Daegu (2020 population: 2.4 million) and Seoul (2020 population: 9.7 million). During January 20–March 16, KCDC reported 6,083 cases from Daegu and 248 from Seoul. The Daegu epidemic was characterized by a single large peak followed by a decrease (Figure 1, panel A); the Seoul epidemic comprised several small outbreaks (Figure 1, panel B).

We obtained the daily number of persons who boarded the subway or monorail in Daegu and Seoul during 2017–2020. For Daegu, we used data from <https://data.go.kr> for lines 1–3; for Seoul, we used data from <https://data.seoul.go.kr> for lines 1–9 (Figure 1). Soon after the first church-related case was reported, traffic volume decreased by  $\approx 80\%$  in Daegu and  $\approx 50\%$  in Seoul. To our knowledge, KCDC first recommended social distancing on February 29 (1), and no official guidelines existed regarding public transportation, which suggests that distancing was, at least in part, voluntary.

We reconstructed the time series of a proxy for incidence of infection  $I_t$ , representing the number of persons who became infected at time  $t$  and reported later, and estimated the instantaneous reproduction number,  $R_t$ , defined as the average number of secondary infections caused by an infected person, given conditions at time  $t$  (5). We adjusted the daily number of reported cases to account for changes in testing criteria and censoring bias (Appendix, <https://wwwnc.cdc.gov/EID/article/26/11/20-1099-App1.pdf>) and then sampled infection dates using inferred onset-to-confirmation delay distributions from the partial line list (Appendix Figure 1) and previous

Author affiliations: Princeton University, Princeton, New Jersey, USA (S.W. Park, B.T. Grenfell); National Institutes of Health, Bethesda, Maryland, USA (K. Sun, C. Viboud, B.T. Grenfell); McMaster University, Hamilton, Ontario, Canada (J. Dushoff)

DOI: <https://doi.org/10.3201/eid2611.20201099>

estimated incubation period distribution (Table) to obtain our incidence proxy,  $I_t$ . Finally, we estimated  $R_t$  on the basis of the renewal equation (5):

$$R_t = \frac{I_t}{\sum_{k=1}^{14} I_{t-k} w_k}$$

where  $w_t$  is the generation-interval distribution randomly drawn from a prior distribution (Table). We weighted each sample of  $R_t$  using a gamma probability distribution with a mean of 2.6 and a SD  $\pm 2$  to reflect prior knowledge (S. Abbott, unpub. data, <https://doi.org/10.12688/wellcomeopenres.16006.1>) and took weighted quantiles to calculate medians and associated 95% credible intervals. We estimated  $R_t$  for February 2 (14 days after the first confirmed case) through March 10 (after which the effects of censoring were too strong for reliable estimates) (Appendix). All analyses were performed using R version 3.6.1 (<https://www.r-project.org>).

We reconstructed incidence proxy (Figure 2, panels A, B) and estimates of  $R_t$  (Figure 2, panels C, D) in Daegu and Seoul. In Daegu, incidence peaked shortly after the first case was confirmed (Figure 2, panel A). In Daegu, symptoms had developed in the first case-patient on February 7; this person had visited the church on February 9 and 16, indicating the disease probably was spreading within the church community earlier (1). Likewise, the estimates of  $R_t$  gradually decreased and eventually decreased to  $<1$  approximately 1 week after the first case was reported, coinciding with the decrease in the metropolitan traffic volume (Figure 2, panel C). The initial decrease in  $R_t$  was likely to have been caused by our resampling method for infection times for each reported case, which oversmooths the incidence curve and the  $R_t$  estimates (K. Gostic, unpub. data, <https://doi.org/10.1101/2020.06.18.20134>

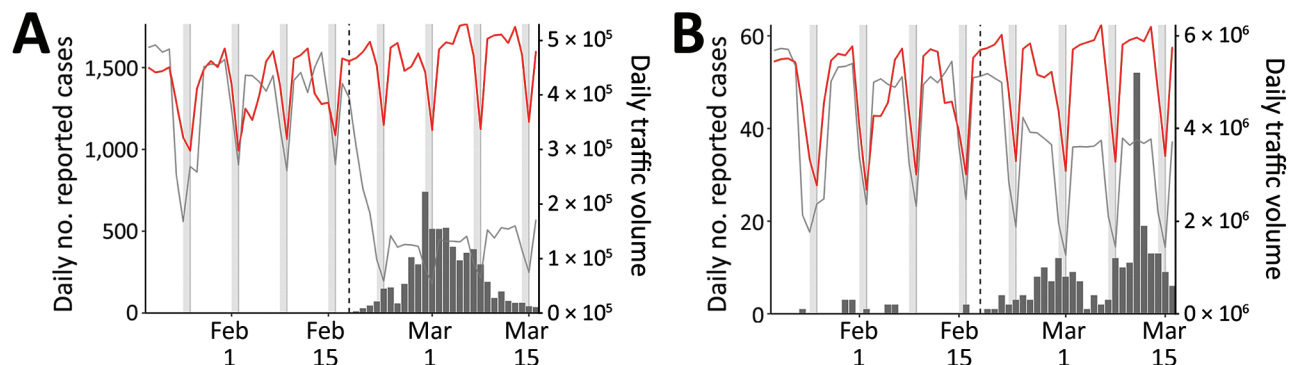
858). In Seoul, estimates of  $R_t$  decreased slightly but remained at  $\approx 1$  (Figure 2, panel D), which might be explained by less-intense social distancing. Stronger distancing or further intervention would have been necessary to reduce  $R_t$  to  $<1$  by March 10.

Although we found clear, positive correlations on a daily scale between normalized traffic and the median estimates of  $R_t$  in Daegu ( $r = 0.93$ ; 95% credible interval 0.86–0.96; Appendix Figure 2) and Seoul ( $r = 0.76$ ; 95% credible interval 0.60–0.87; Appendix Figure 2), these correlations are conflated by time trends and by other measures that could have affected  $R_t$ . We did not find clear signatures of lags in the correlation between  $R_t$  and traffic volume (Appendix Figure 3). Patterns in  $R_t$  were similar in directly adjacent provinces (Gyeongsangbuk-do and Gyeonggi-do), demonstrating the robustness of our analysis (Appendix Figure 4).

## Conclusions

The South Korea experience with COVID-19 provides evidence that epidemics can be suppressed with less extreme measures than those taken by China (9) and demonstrates the necessity of prompt identification and isolation of case-patients in preventing spread (4). Our analysis reveals the potential role of social distancing in assisting such efforts. Even though social distancing alone might not prevent spread, it can flatten the epidemic curve (compare Figure 2, panels B, D) and reduce the burden on the healthcare system (10).

Our study is not without limitations. Because of insufficient data, we could not account for differences in delay distributions or changes in testing capacity among cities; line list data were mostly derived from outside Daegu. Nonetheless, the sensitivity analyses support the robustness of our findings (Appendix



**Figure 1.** Comparison of daily epidemiologic and traffic data from Daegu (A) and Seoul (B) during the coronavirus disease (COVID-19) outbreak, South Korea. Black bars indicate no. COVID-19 cases; lines represent daily metropolitan traffic volume in 2020 (red) and mean daily metropolitan traffic volume during 2017–2019 (black). Daily traffic from previous years have been shifted by 1–3 days to align day of the weeks. Vertical dashed lines indicate February 18, 2020, when the first COVID-19 case was confirmed in Daegu. Gray bars indicate weekends.

**Table.** Assumed incubation and generation-interval distributions in an analysis of the potential role of social distancing in mitigating the spread of coronavirus disease, South Korea, 2020\*

Distribution	Parameterization	Priors	Source
Incubation period distribution	Gamma ( $\mu_i, \mu_i^2/\sigma^2$ )	$\mu_i \approx \text{gamma}$ (6.5 d, 145); $\sigma \approx \text{gamma}$ (2.6 d, 25)	(6)
Generation-interval distribution	Negative binomial ( $\mu_G, \theta$ )	$\mu_G \approx \text{gamma}$ (5 d, 62); $\theta \approx \text{gamma}$ (5, 20)	(7,8)

\*Gamma distributions are parameterized using its mean and shape. Negative binomial distributions are parameterized using its mean and dispersion. Priors are chosen such that the 95% quantiles of prior means and standard deviations are consistent with previous estimates.

Figures 5–8). We were unable to distinguish local and imported cases and thus might have overestimated  $R_t$  (11). Conducting a separate analysis for Seoul that accounts for imported cases did not affect our qualitative conclusions (Appendix Figure 9). Finally, although the method of resampling infection time can capture qualitative changes in  $R_t$ , estimates of  $R_t$  can be oversmoothed and should be interpreted with care (K. Gostic, unpub. data, <https://doi.org/10.1101/2020.06.18.20134858>). Nonetheless, our estimates of  $R_t$  are broadly consistent with previous estimates (12).

We used metropolitan traffic to quantify the degree of social distancing. The 80% decrease in traffic volume suggests that distancing measures in Daegu might be comparable to those in Wuhan, China (13). We were unable to directly estimate the effect of social distancing on population contacts or epidemiologic dynamics. Other measures, such as intensive testing and tracing of core transmission groups, are also likely to have affected transmission dynamics.

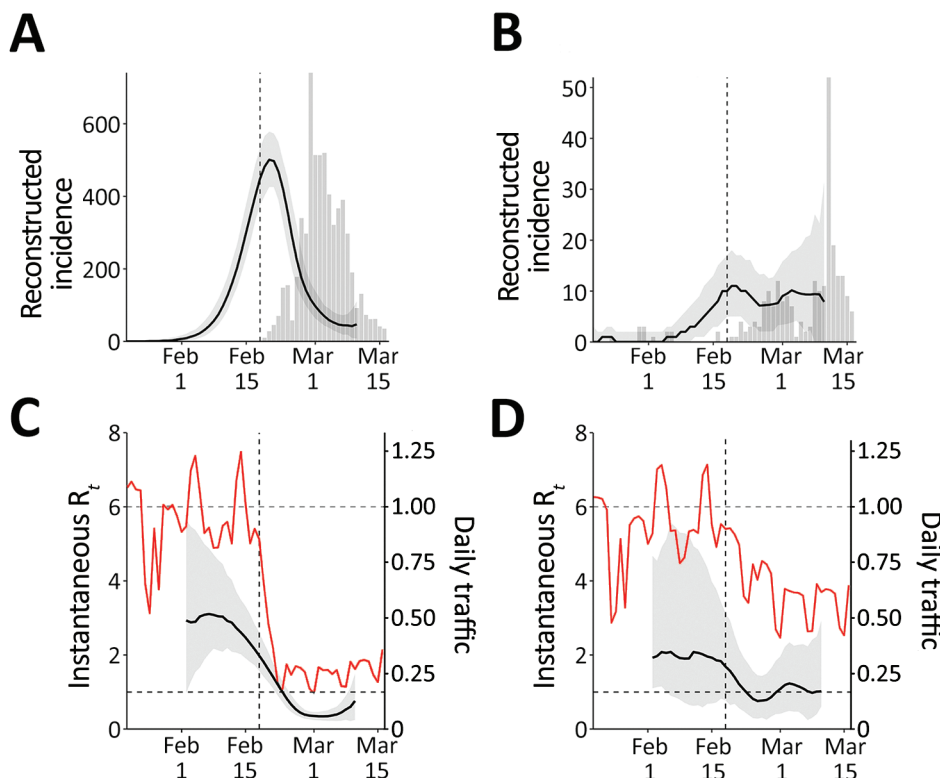
Our study highlights the importance of considering geographic heterogeneity in estimating epidemic potential. The sharp decrease in Daegu drove the number of reported cases in South Korea. Our analysis revealed that the epidemic remained close to the epidemic threshold in other regions, including Seoul and Gyeonggi-do. Relatively weak distancing might have assisted the recent resurgence of COVID-19 cases in Seoul (E. Shim, G. Chowell, unpub. data, <https://doi.org/10.1101/2020.07.21.20158923>).

This article was preprinted at <https://www.medrxiv.org/content/10.1101/2020.03.27.20045815v1>.

J.D. was supported from the Canadian Institutes of Health Research.

**About the Author**

Mr. Park is a PhD student in the Ecology and Evolutionary Biology Department at Princeton University. His research focuses on mathematical and statistical modeling of infectious diseases.



**Figure 2.** Comparison of reconstructed coronavirus disease incidence proxy and instantaneous reproduction number  $R_t$  in Daegu (A, C) and Seoul (B, D), South Korea. The instantaneous reproduction number  $R_t$  reflects transmission dynamics at time  $t$ . Black lines and gray shading represent the median estimates of reconstructed incidence (A, B) and  $R_t$  (C, D) and their corresponding 95% credible intervals. Gray bars show the number of reported cases. Red lines represent the normalized traffic volume (daily traffic, 2020, divided by the mean daily traffic, 2017–2019). Vertical dashed lines indicate February 18, 2020, when the first case was confirmed in Daegu.

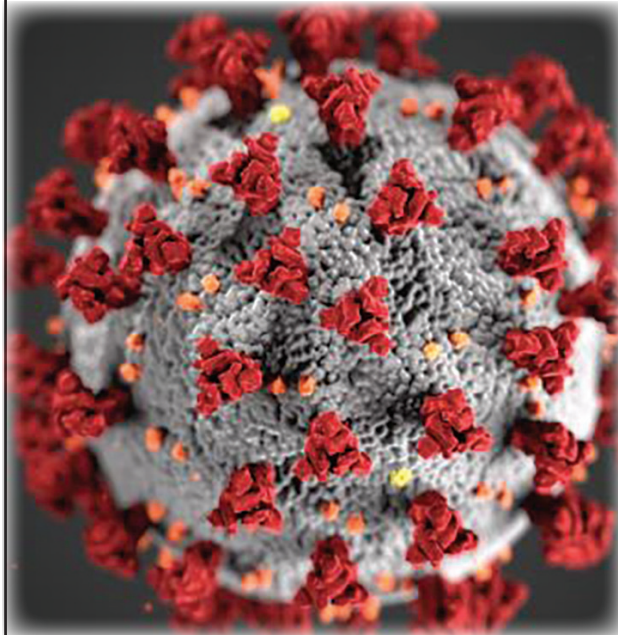
## References

1. Korea Centers for Disease Control and Prevention). Press release [in Korean]. Release dates 2020 Jan 20, Feb 18, Feb 19, Feb 24, Feb 29, Mar 16 [cited 2020 Jan 20–Jul 23]. <https://www.cdc.go.kr/board/board.es?mid=a20501000000&bid=0015#>
2. Daegu Metropolitan City Government. Press release [in Korean] [cited 2020 Jul 22]. <http://152.99.22.79:8080/SynapDocViewServer/viewer/doc.html?key=2c948aed7356835e017378b583741d07&contextPath=/SynapDocViewServer/>
3. Korean Ministry of Education. Press release [in Korean] [cited 2020 Jul 22]. <https://www.moe.go.kr/boardCnts/view.do?boardID=294&boardSeq=79829&lev=0&searchType=null&statusYN=W&page=21&s=moe&m=020402&opType=N>
4. Park YJ, Choe YJ, Park O, Park SY, Kim YM, Kim J, et al.; COVID-19 National Emergency Response Center, Epidemiology and Case Management Team. Contact tracing during coronavirus disease outbreak, South Korea, 2020. *Emerg Infect Dis.* 2020 Jul 16 [Epub ahead of print]. <https://doi.org/10.3201/eid2610.201315>
5. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS One.* 2007;2:e758. <https://doi.org/10.1371/journal.pone.0000758>
6. 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;25:25. <https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062>
7. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying dynamics of SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science.* 2020;368:eabb6936.
8. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveill.* 2020;25:2000257. <https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257>
9. Kickbusch I, Leung G. Response to the emerging novel coronavirus outbreak. *BMJ.* 2020;368:m406. <https://doi.org/10.1136/bmj.m406>
10. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet.* 2020;395:931–4. [https://doi.org/10.1016/S0140-6736\(20\)30567-5](https://doi.org/10.1016/S0140-6736(20)30567-5)
11. Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics.* 2019;29:100356. <https://doi.org/10.1016/j.epidem.2019.100356>
12. Ryu S, Ali ST, Jang C, Kim B, Cowling BJ. Effect of nonpharmaceutical interventions on transmission of severe acute respiratory syndrome coronavirus 2, South Korea, 2020. *Emerg Infect Dis.* 2020 Jun 2 [Epub ahead of print]. <https://doi.org/10.3201/eid2610.201886>
13. Kraemer MUG, Yang C-H, Gutierrez B, Wu C-H, Klein B, Pigott DM, et al.; Open COVID-19 Data Working Group. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science.* 2020;368:493–7. <https://doi.org/10.1126/science.abb4218>

Address for correspondence: Sang Woo Park, Department of Ecology and Evolutionary Biology, 106A Guyot Hall, Princeton University, Princeton, NJ 08544-2016, USA; email: [swp2@princeton.edu](mailto:swp2@princeton.edu)

## EID SPOTLIGHT TOPIC

# Coronavirus



This spotlight provides articles published in *Emerging Infectious Diseases* about human coronavirus diseases, including coronavirus disease 2019 (COVID-19), severe acute respiratory syndrome (SARS), and the common cold.

<https://wwwnc.cdc.gov/eid/spotlight/coronavirus>

**EMERGING  
INFECTIOUS DISEASES®**

# Potential Roles of Social Distancing in Mitigating Spread of Coronavirus Disease, South Korea

## Appendix

### Epidemiologic Data

The daily number of reported cases from each municipality was translated and transcribed from the KCDC press release (*1*). Following the KCDC's protocol, the daily number of reported cases before February 20, 2020, reflects the number of confirmed cases on each day. During February 21–March 1, 2020, the daily number of reported cases reflects the number of reported cases within the last 24 hours (9 a.m. to 9 a.m.). On March 2, 2020, the daily number of reported cases reflects the number of cases that were reported between 9 a.m. March 1, 2020, and 12 a.m. March 2, 2020. Since then, the daily number of reported cases reflects the number of reported cases within the last 24 hours (12 a.m. to 12 a.m.). The number of negative cases was not reported on January 25 and 31, 2020; we took the average of cumulative negative cases from 1 day before and after these dates instead to impute missing values. The daily number of reported cases by the KCDC may be slightly different from the reports by each municipal government as some cases may be transferred after they are confirmed. The sum of daily number of reported cases by the KCDC may be also slightly different from the cumulative number of cases reported the KCDC because it does not reflect possible location changes of the confirmed cases after reporting.

### Reconstruction of Incidence Time Series

According to the KCDC press release (*1*), testing criteria expanded 4 times during January 20–March 16, 2020: January 28, February 7, February 20, and March 2, 2020. We accounted for these changes by assuming that the proportion positive should remain roughly constant if we follow a consistent protocol of identifying and deciding whom to test. To do so,

we calculated the relative proportion of positive cases during each period (divided by the between-period mean) and multiplied the daily number of reported cases by the relative proportions of the corresponding criterion. Sensitivity analyses showed that results are robust to these adjustments (Appendix Figures 5–8).

We then estimated time-dependent *backward* onset-to-confirmation delay distributions from the partial line list: Given a cohort of infected individuals who were confirmed on the same day, what is the probability distribution of the onset-to-confirmation delay? The backward delay distribution depends on changes in the number of symptomatic cases—e.g., when the number of symptomatic cases is increasing, the backward delay distribution is likely to be shorter because individuals are more likely to have developed symptoms recently. The backward delay distribution was inferred using a negative-binomial regression with log-link using the `brms` package (2). Time-dependent mean of the negative binomial distribution is modeled using splines. We assumed weakly informative priors on the fixed effects: normal distributions with mean of 0 and standard deviation of 2; note that these distributions are priors on link scale.

For each posterior sample of the backward delay distribution, we drew a random sample of onset-to-confirmation delay and incubation period for each confirmed case. This allowed us to obtain posterior samples of possible infection dates for each case, which were then converted into posterior samples of incidence time series.

To account for right-censoring in the reported cases, we also estimated time-dependent *forward* onset-to-confirmation delay distribution using the same negative-binomial regression model: Given a cohort of infected individuals who became symptomatic on the same day, what is the probability distribution of the onset-to-confirmation delay? The forward delay distribution reflects the changes in the accuracy of case identification—e.g., a decrease in the delay reflects improvement in accuracy.

To estimate the forward delay distribution, we modified the `stan` code from the negative-binomial regression that we used to infer the backward delay distribution to account for right-censoring (in the observed delays) and ran the code using the `RStan` package (3). In particular, we modified the likelihood of the negative-binomial regression such that given a delay of  $x_i$  days, symptom onset day  $t_i$  and the day of measurement of  $t_{max}$ , the likelihood of observing the delay is given by:

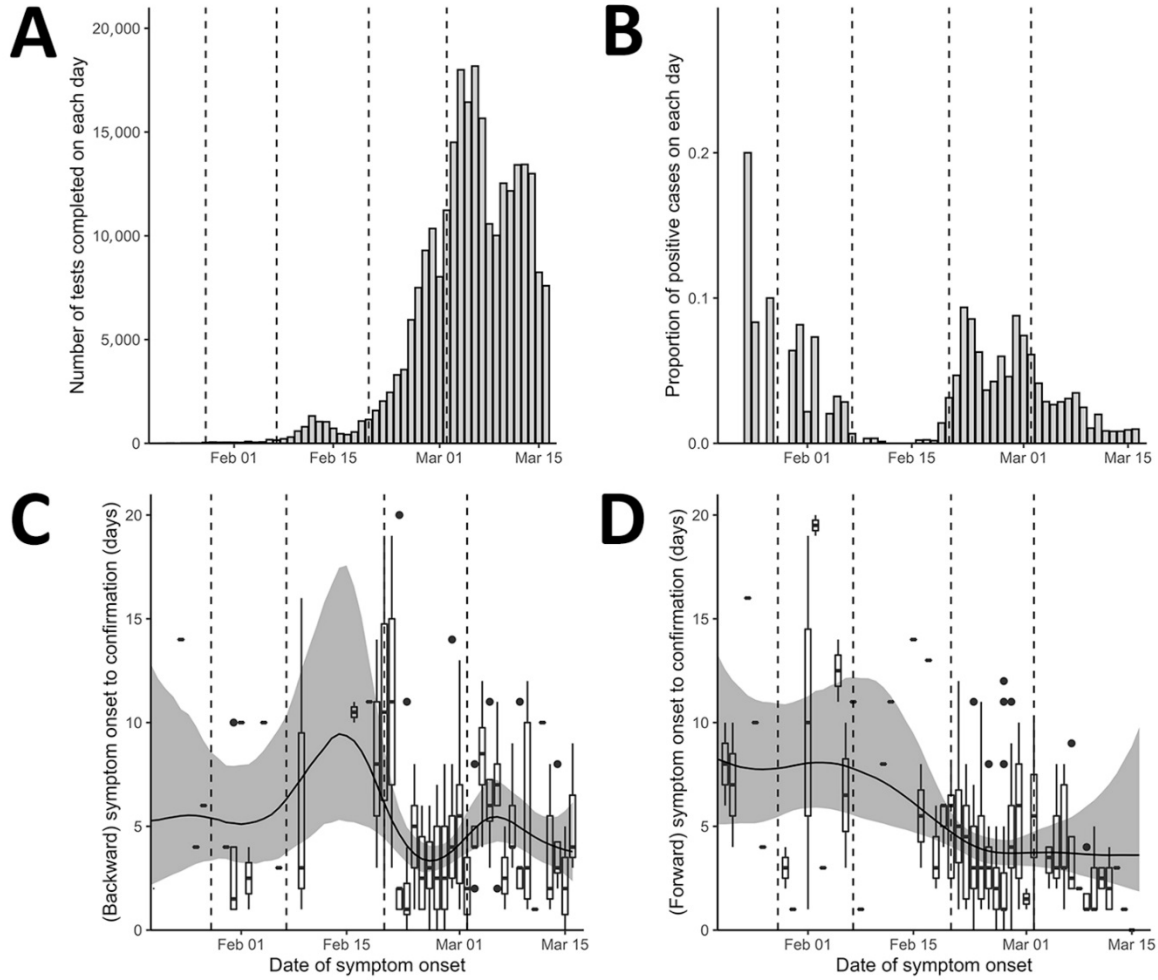
$$\frac{f(x_i|\mu(t_i), \theta)}{F(t_{max} - t_i|\mu(t_i), \theta)}$$

where  $f$  is the negative binomial distribution with time-dependent mean  $\mu(t_i)$  and dispersion parameter  $\theta$ . This likelihood accounts for the fact that the delay between symptom onset and confirmation cannot be longer than  $t_{max} - t_i$  (otherwise, the case will be reported after  $t_{max}$ ). Convergence is assessed by the lack of warning messages from the RStan package (3).

For each combination of date of infection and a posterior sample of the forward delay distribution, we drew 1,000 samples of incubation periods and onset-to-confirmation delays and calculated the median probability that an individual infected on a given day will be confirmed before March 16, 2020. Finally, we divided the daily number of infected cases by the median probability this probability. We used the reconstructed time series of incidence proxy to estimate  $\mathcal{R}_t$ .

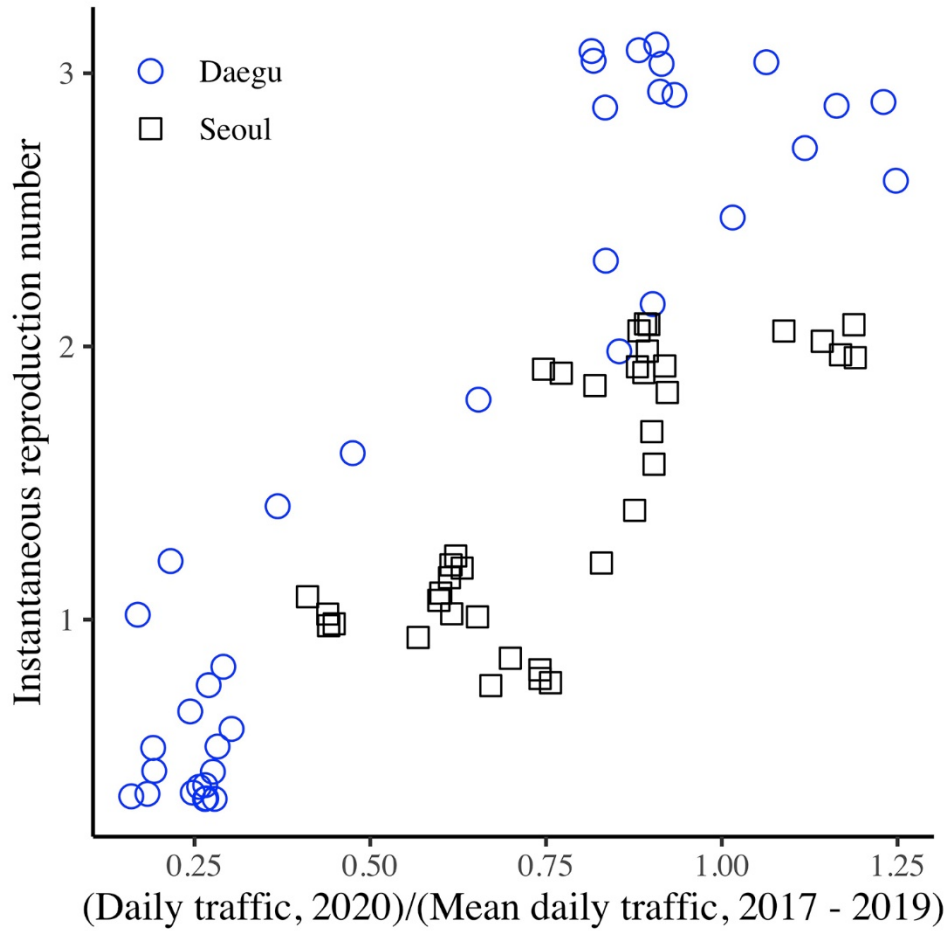
## References

1. Korea Centers for Disease Control and Prevention (KCDC). Press release [in Korean] [cited 2020 Jan 20–Jul 23]. Korea Centers for Disease Control and Prevention). Press release [in Korean]. Release dates 2020 Jan 20–Mar 16 [cited 2020 Jan 20–Jul 23]. <https://www.cdc.go.kr/board/board.es?mid=a20501000000&bid=0015#>
2. Bürkner P. brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw.* 2017;80:1–28. <https://doi.org/10.18637/jss.v080.i01>
3. Stan Development Team. RStan: the R interface to Stan. R package version 2.19.3. 2020 [cited 2020 Jul 28]. <https://cran.r-project.org/web/packages/rstan/vignettes/rstan.html>

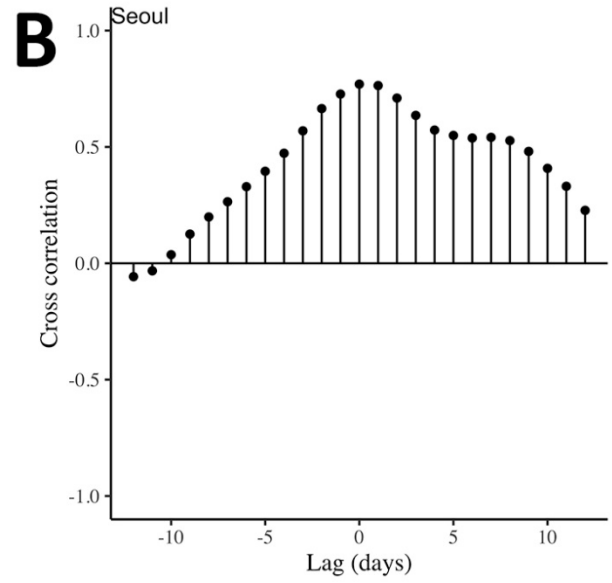
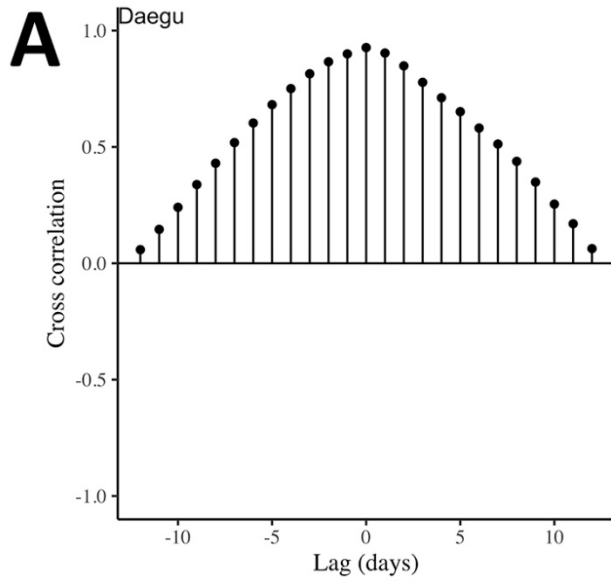


**Appendix Figure 1.** Changes in the number of tests and delay distributions over time. Vertical lines indicate the date on which testing criteria expanded. Box plots (C, D) represent the observed delays. Black lines and gray ribbons represent the median estimates of the mean delays and their associated 95% credible intervals.

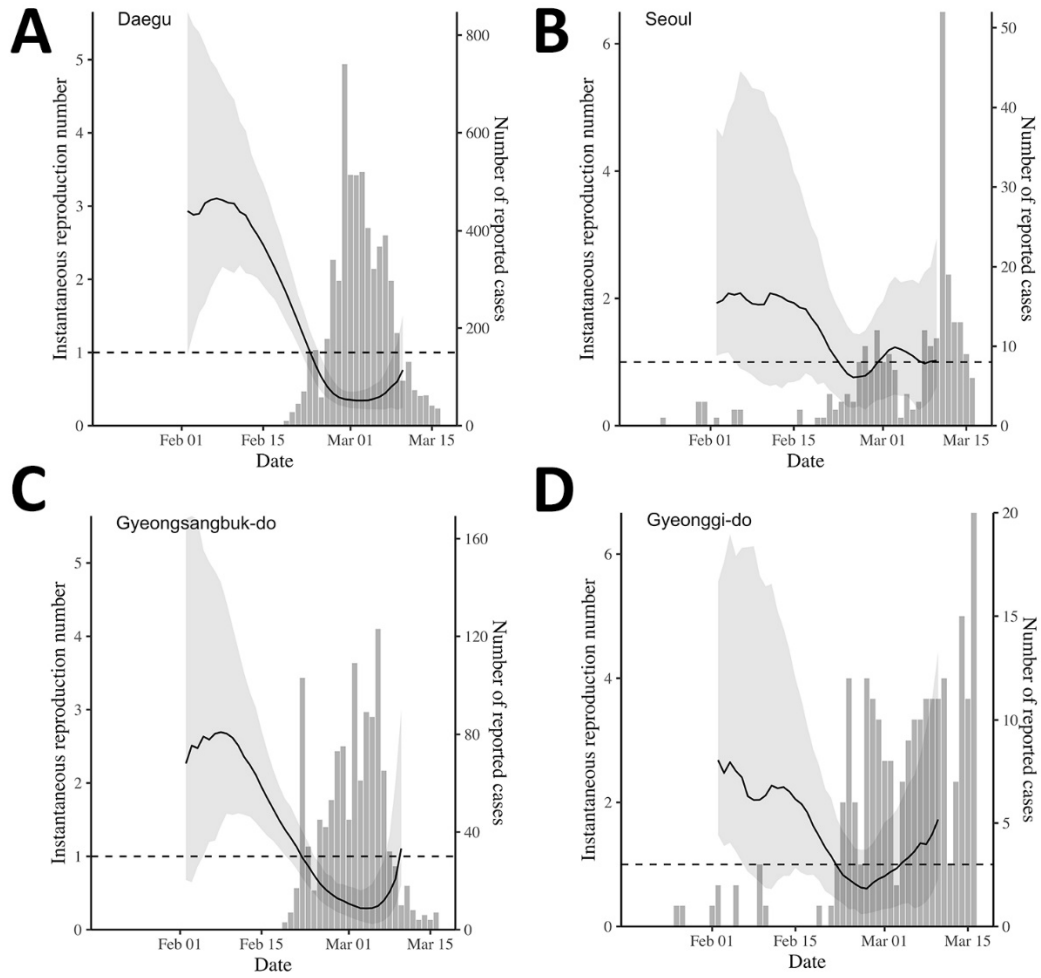




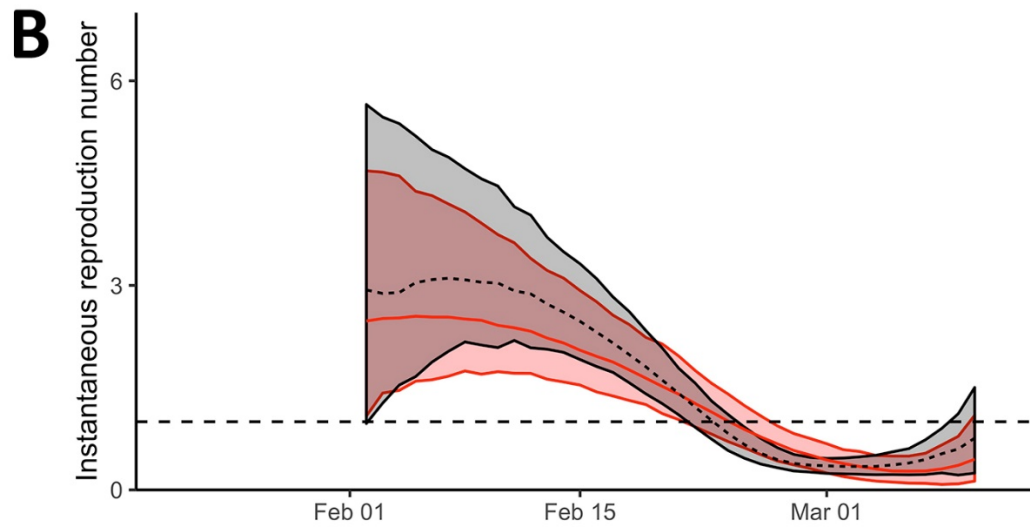
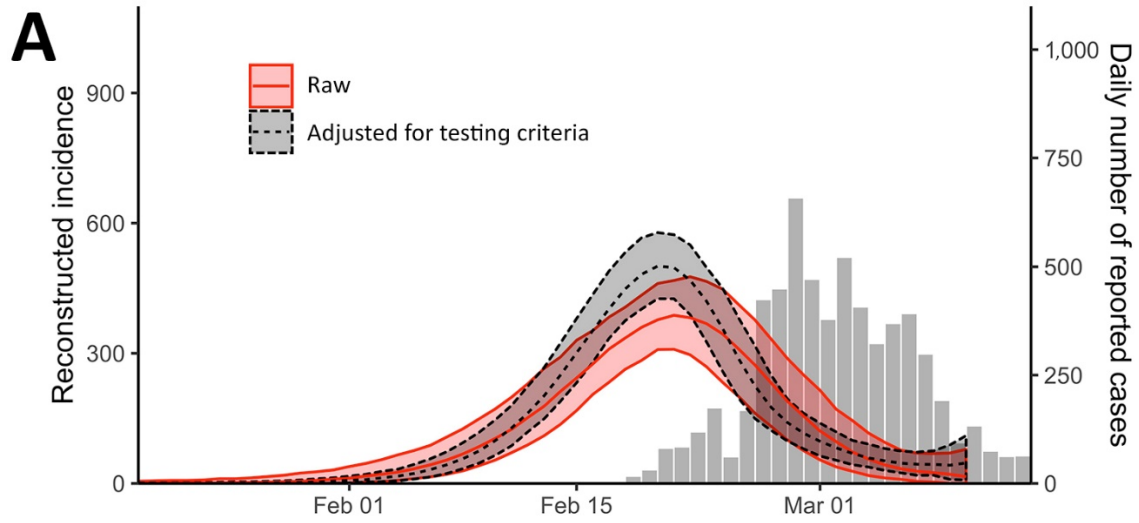
**Appendix Figure 2.** Scatter plot of the normalized traffic volume and the median estimates of  $\mathcal{R}_t$  on a daily scale.



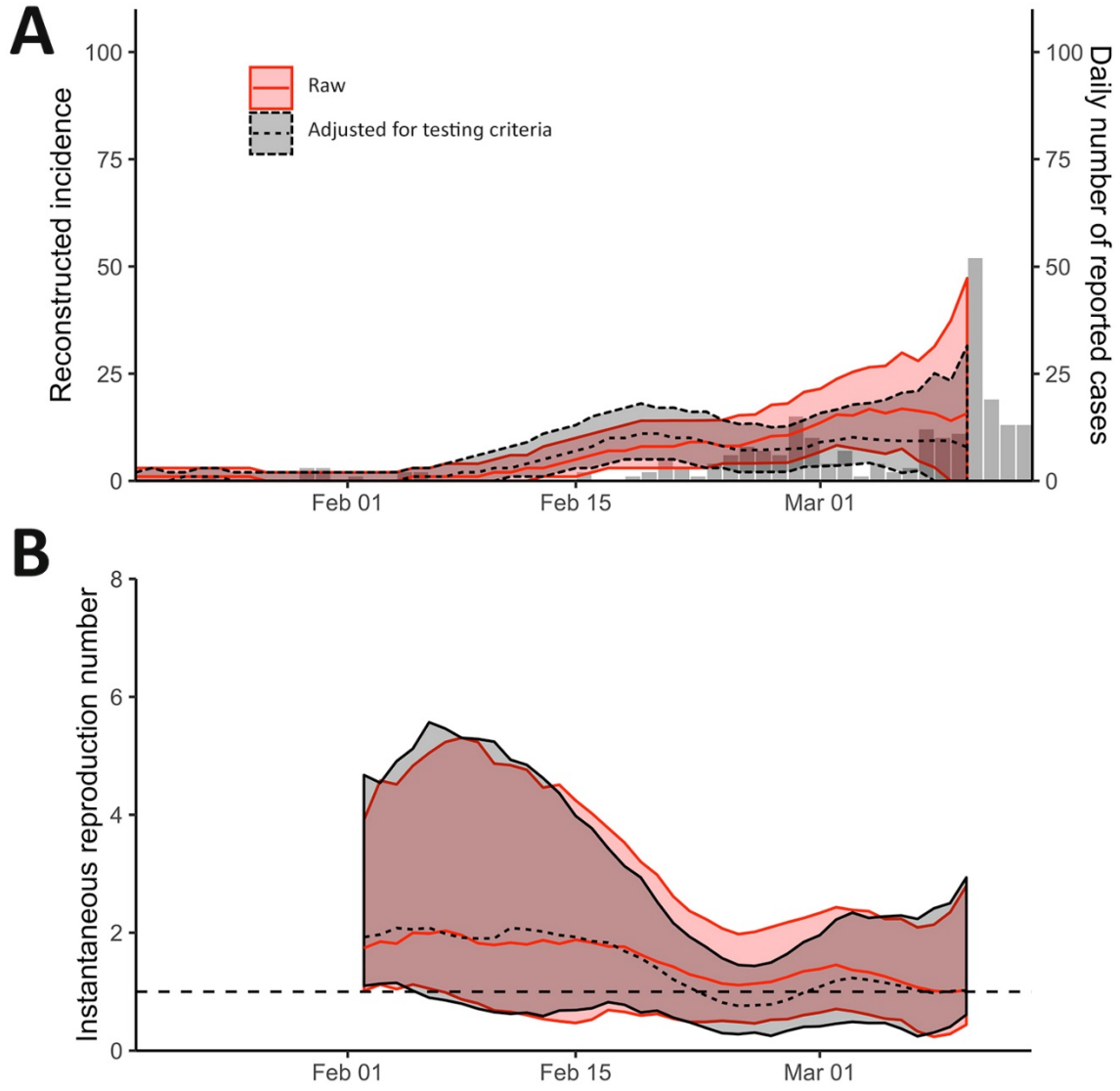
**Appendix Figure 3.** Cross correlation between the normalized traffic volume and the median estimates of  $\mathcal{R}_t$  in Daegu (A) and Seoul (B).



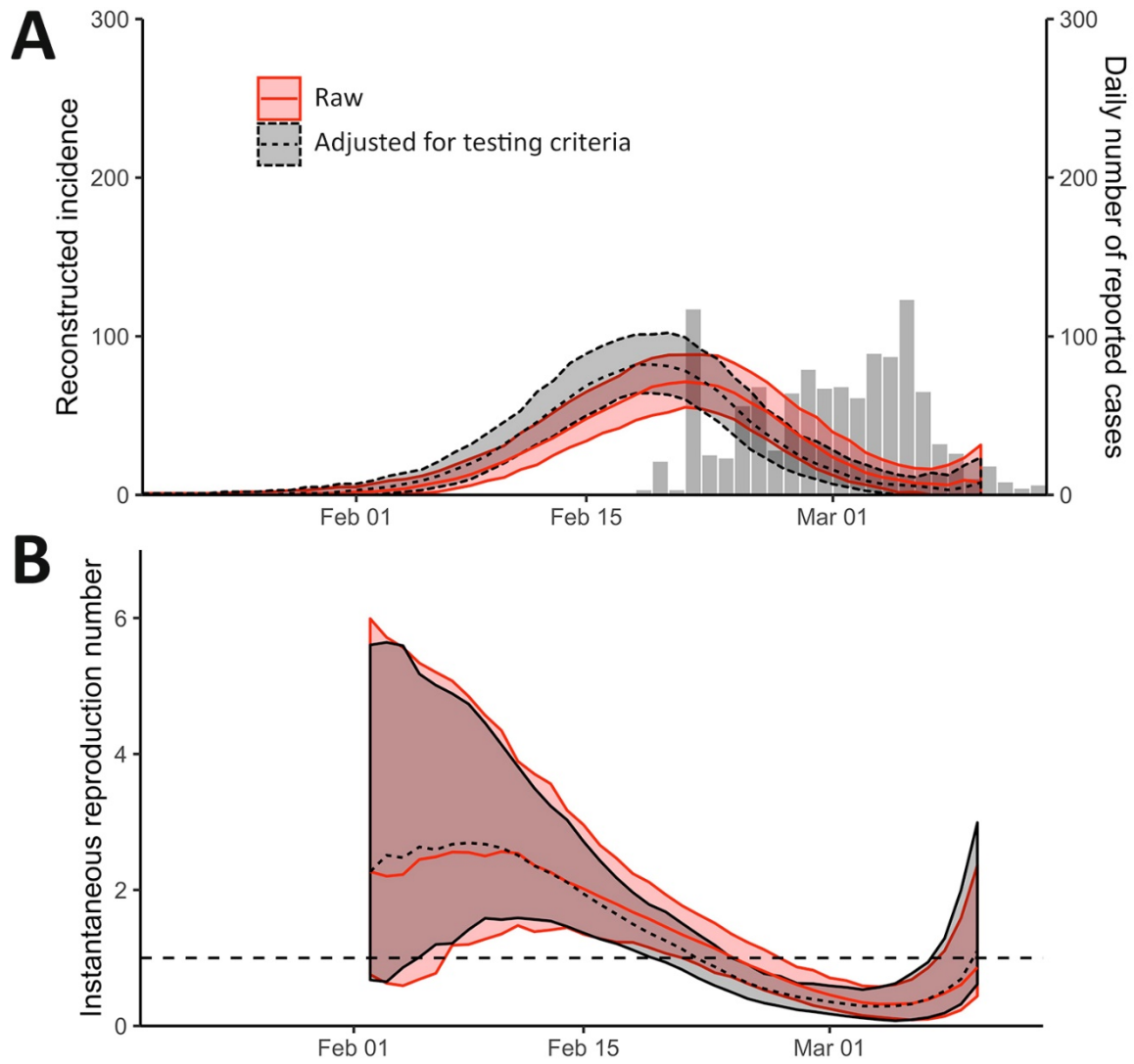
**Appendix Figure 4.** Comparison of  $\mathcal{R}_t$  estimates and the daily number of reported cases in Daegu (A), Seoul (B), Gyeongsangbuk-do (C), and Gyeonggi-do (D).



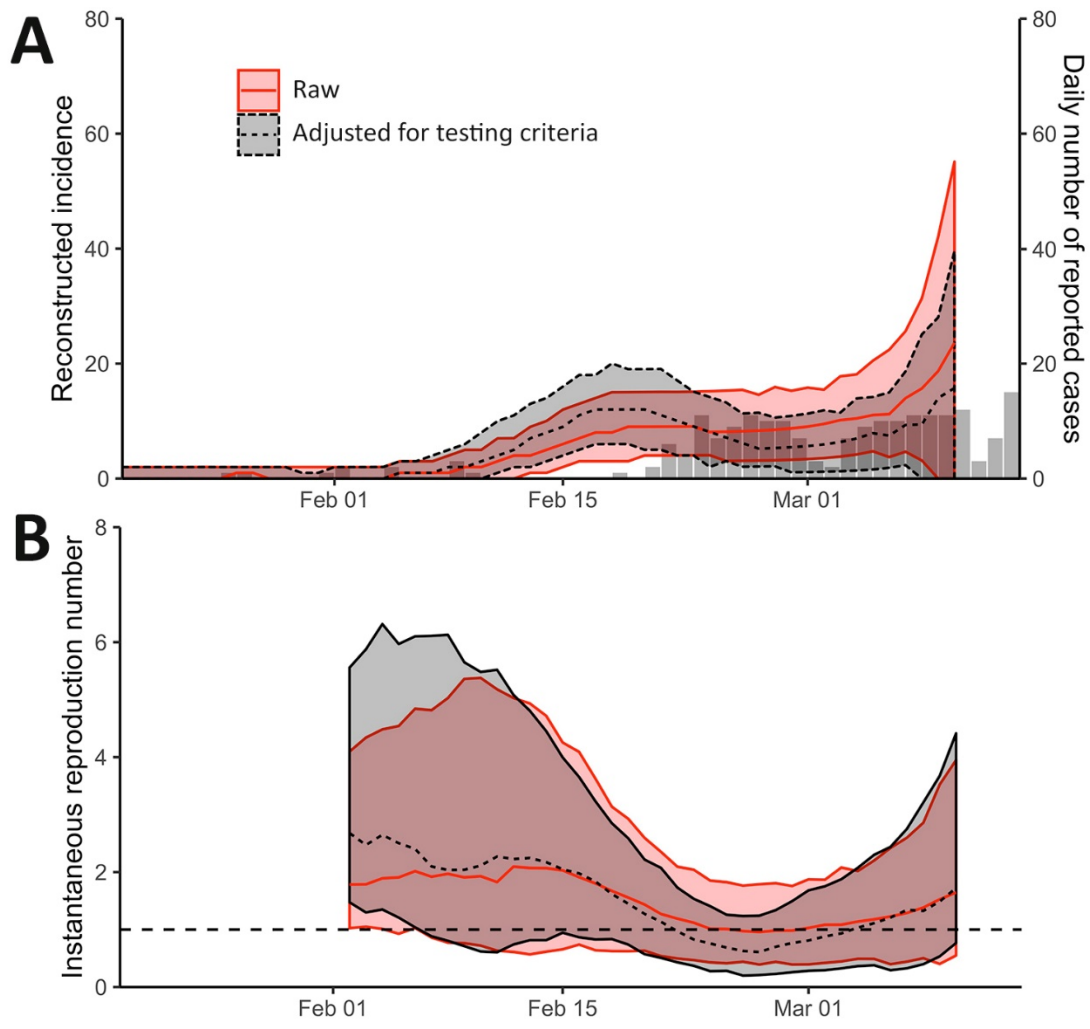
**Appendix Figure 5.** Sensitivity analysis of  $\mathcal{R}_t$  estimates in Daegu with respect to changes in testing criteria.



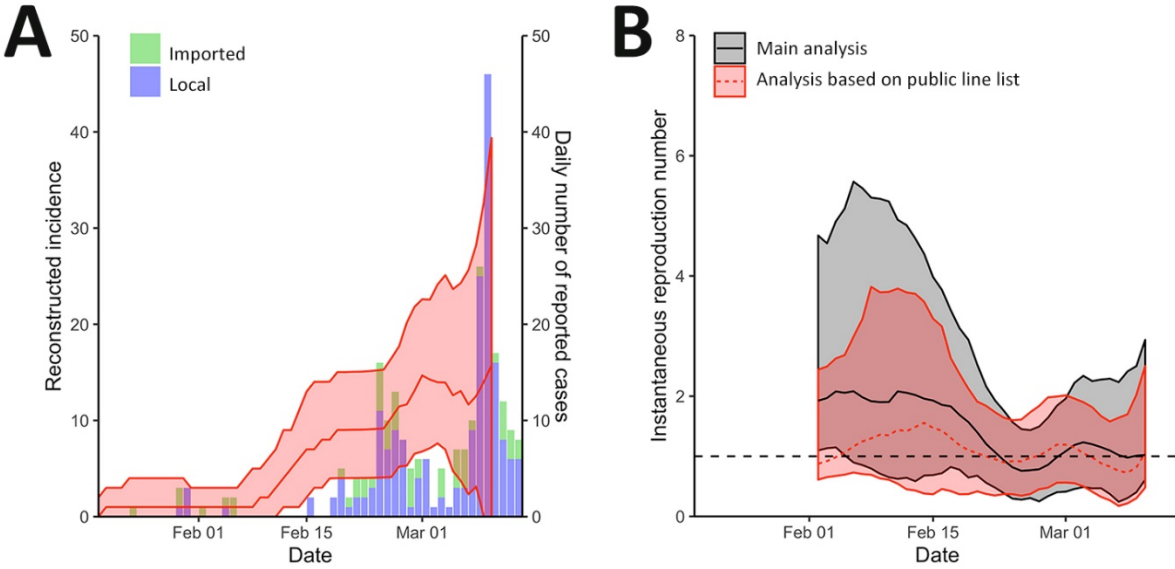
**Appendix Figure 6.** Sensitivity analysis of  $\mathcal{R}_t$  estimates in Seoul with respect to changes in testing criteria.



**Appendix Figure 7.** Sensitivity analysis of  $\mathcal{R}_t$  estimates in Gyeongsangbuk-do with respect to changes in testing criteria.



**Appendix Figure 8.** Sensitivity analysis of  $\mathcal{R}_t$  estimates in Gyeonggi-do with respect to changes in testing criteria.



**Appendix Figure 9.** Comparison  $\mathcal{R}_t$  in Seoul using the number of reported cases by the KCDC and public line list provided by the Seoul Metropolitan Government. Using public line list, we reconstructed incidence for local  $I_t^{local}$  and imported  $I_t^{imported}$  cases separately based on the method described in the main text. Then, we estimated the time-dependent reproduction number via  $\mathcal{R}_t = I_t^{local} / \sum_{k=1}^{14} I_{t-k} W_k$ , where  $I_t = I_t^{local} + I_t^{imported}$ . We did not account for changes in testing criteria in this analysis. The line list was obtained from <http://news.seoul.go.kr/welfare/archives/513105> and <https://www.seoul.go.kr/coronaV/coronaStatus.do>.