Potential roles of social distancing in mitigating the spread of coronavirus disease 2019 (COVID-19) in South Korea

Sang Woo Park[1,*] Kaiyuan Sun[2] Cécile Viboud[2] Bryan T. Grenfell[1,2,3] Jonathan Dushoff[4,5,6]

1 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA
2 Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
3 Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA
4 Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada
5 M. G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, ON, Canada
6 Department of Biology, McMaster University, Hamilton, ON, Canada

*Corresponding author: swp2@princeton.edu
This article does not necessarily represent the views of the NIH or the US government.

Abstract

On January 20, 2020, the first COVID-19 case was confirmed in South Korea. After a rapid outbreak, the number of incident cases has been consistently decreasing since early March; this decrease has been widely attributed to its intensive testing. We report here on the likely role of social distancing in reducing transmission in South Korea. Our analysis suggests that transmission may still be persisting in some regions.

Keywords

coronavirus disease, COVID-19, SARS-CoV-2, South Korea, social distancing

Contribution

Data collection: SWP; conceptualization: SWP, CV; analysis: SWP; first draft: SWP, JD. All authors contributed to the writing and approval of the final report.

Funding statement

Conflict of interest

None declared.

1 Background

Since its first appearance in Wuhan, China, in December 2019 [1], coronavirus disease (COVID-19) has spread internationally, including to South Korea. The first COVID-19 case in South Korea was confirmed on January 20, 2020, from a traveling resident of Wuhan, China [2]. In February, the disease spread rapidly within a church community in the city of Daegu [2]. The chains of transmission that began from this cluster distinguish the epidemic in South Korea from that in any other countries: As of March 24, 2020, 9,037 cases were confirmed, of which 56% were related to the church and 27% were in their 20s [2]. South Korea's intensive testing using novel contact tracing techniques allowed rapid identification and isolation cases and reduction of onward transmission [3, 4, 5]. Here, we describe potential roles of social distancing in mitigating the spread of COVID-19 in South Korea by using metro traffic data to compare epidemics in two major cities.

2 Data description

We analyzed epidemiological data describing the COVID-19 outbreak in South Korea between January 20–March 16, 2020. Daily number of reported cases in each geographic region was transcribed from press releases by the Korea Centers for Disease Control and Prevention (KCDC) [2]. Partial line lists were translated and transcribed from press releases by the KCDC and various local and provincial governments [6, 7, 8, 9, 10, 11, 12]. All data and code are stored in a publicly available GitHub repository: `https://github.com/parksw3/Korea-analysis`.

We compared epidemiological dynamics of COVID-19 from two cities in which the largest number of COVID-19 cases have been reported: Daegu and Seoul. Between January 20–March 16, 2020, 6,083 cases from Daegu and 248 from Seoul were reported by the KCDC. The epidemic in Daegu is characterized by a single, large peak followed by a gradual decrease, whereas the epidemic in Seoul consists of several small outbreaks (Fig. 1).

Daily metro traffic in Daegu and Seoul between 2017–2020 was obtained from `data.go.kr` and `data.seoul.go.kr`, respectively. We tabulated the total number of individuals who accessed the subway or monorail (using Seoul lines 1–9, and Daegu lines 1–3; Fig. 1). Soon after the first church-related case was confirmed in Daegu on Feb 18, 2020, the daily traffic volume decreased by about 80% and 50% in Daegu and Seoul, respectively.

3 Trends in time-dependent reproduction number and traffic volume

To estimate the time-dependent reproduction number R_t (the average number of secondary cases caused by an average individual, given conditions at time t [13]), we first estimated daily incidence of infection from the daily number of reported cases by the KCDC [2]. We adjusted the number of reported cases to account for changes in testing criteria, which occurred 4 times between January 20–March 16, 2020. Then, we inferred onset-to-confirmation
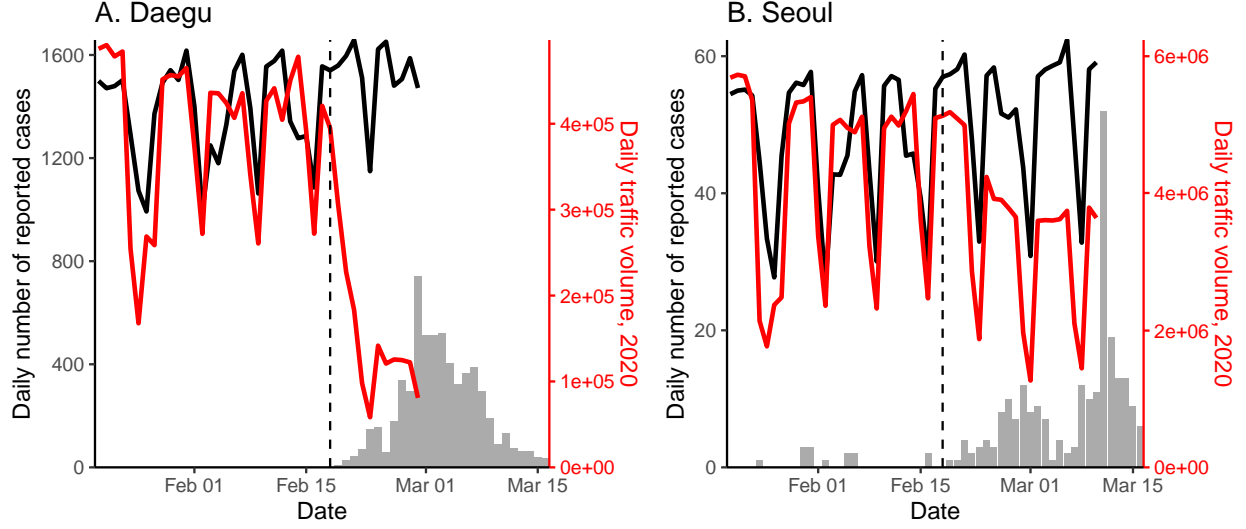
Figure 1: **Comparison of epidemiological and traffic data from Daegu and Seoul.**
Solid lines represent the daily metro traffic volume in 2020 (red) and mean daily metro traffic
volume between 2017–2019 (black). Daily traffic from previous years have been shifted by
1–3 days to align day of the weeks. Vertical lines indicate Feb 18, 2020, when the first case
was confirmed in Daegu. See Supplementary Materials for details.

delay distributions from the partial line list and combined them with previously estimated
incubation period distribution (Table 1) to obtain probability distributions for date of infec-
tion for each reported case. We accounted for right-censoring by dividing the daily incidence
by the probability that a case infected on a given day would have been reported before March
16, 2020. Implementation details are provided in the Supplementary Materials.

We estimated the time-dependent reproduction number using the renewal equation with
a 14-day sliding window [13]:

$$\mathcal{R}_t = \frac{I_t}{\sum_{k=1}^{14} I_{t-k} w_k}, \tag{1}$$

where I_t is the reconstructed incidence time series (i.e., the number of infected cases on day
t) and w_k is the generation-interval distribution randomly drawn from a prior distribution
(Table 1). We weighted each sample of $\mathcal{R}_t$ by a gamma probability distribution with a
mean of 2.6 and a standard deviation of 2 to reflect prior knowledge [17] and took weighted
quantiles to calculate the medians and associated 95% credible intervals. We estimated $\mathcal{R}_t$
between February 2, 2020 (14 days after the first confirmed case was imported) and March
10, 2020 (after this point the effects of censoring are too strong for reliable estimates).

Fig. 2 compares the reconstructed incidence (A,B) and estimates of $\mathcal{R}_t$ (C,D) in Daegu
and Seoul. In Daegu, incidence peaked shortly after the first case was confirmed and then
decreased (Fig. 2A). Likewise, the estimates of $\mathcal{R}_t$ gradually decrease and eventually drop
below 1 about a week after the reporting of its first case, coinciding with the decrease in
the metro traffic volume (Fig. 2C). The initial decrease in $\mathcal{R}_t$ may reflect behavior change

	Parameterization	Priors	Source
Incubation period distribution	$\text{Gamma}(\mu_I, \mu_I^2/\sigma^2)$	$\mu_I \sim \text{Gamma}(6.5\,\text{days}, 145)$ $\sigma \sim \text{Gamma}(2.6, 25)$	[14]
Generation-interval distribution	$\text{NegativeBinomial}(\mu_G, \theta)$	$\mu_G \sim \text{Gamma}(5\,\text{days}, 62)$ $\theta \sim \text{Gamma}(5, 20)$	[15, 16]

Table 1: **Assumed incubation and generation-interval distributions.** 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.

within the church; the first confirmed case in Daegu became symptomatic on February 7, 2020, and visited the church on February 9 and 16, 2020 [2]. Our estimates of $\mathcal{R}_t$ for Daegu are consistent with the estimates of $\mathcal{R}_t$ for South Korea by Abbott *et al.* [17] — their estimates drop below 1 slightly later because they rely on number of symptomatic cases instead.

In Seoul, estimates of $\mathcal{R}_t$ decrease slightly but remain around 1 (Fig. 2D). Our analysis suggests that social distancing in Seoul was less intense, and this could be why reduction in spread was less sharp. Stronger distancing or further intervention will be necessary to reduce $\mathcal{R}_t$ below 1.

While we find clear, positive correlations between the normalized traffic and the median estimates of $\mathcal{R}_t$ in both Daegu ($r = 0.90$; 95% CI: 0.79–0.95) and Seoul ($r = 0.76$; 95% CI: 0.59–0.87), these correlations are conflated by time trends, and are also likely conflated by other measures that could have affected $\mathcal{R}_t$. We do not find clear signatures of lagged correlation between $\mathcal{R}_t$ and traffic volume (Supplementary Materials). Similar patterns in the estimates of $\mathcal{R}_t$ are found in directly surrounding provinces (Gyeongsangbuk-do and Gyeonggi-do), providing support for the robustness of our analysis (Supplementary Materials).

4   Discussion

The South Korean experience with COVID-19 provides evidence that epidemics can be suppressed with less extreme measures than those taken by China [18]. It demonstrates the necessity of prompt identification and isolation of cases in preventing further spread [3, 4, 5]. Our analysis reveals potential roles of social distancing in mitigating the COVID-19 epidemic in South Korea. Even though social distancing alone may not be able to fully prevent the spread of the disease, its ability to flatten the epidemic curve (cf. Fig. 2B,D) reduces burden for healthcare system and provides time to plan for the future [19].

Our study is not without limitations. We did not account for differences in the delay distributions or changes in the number of tests among cities. The intensity of intervention is likely to vary across regions given that majority of COVID-19 cases in South Korea were reported from Daegu. We did not have sufficient data to account for these factors.
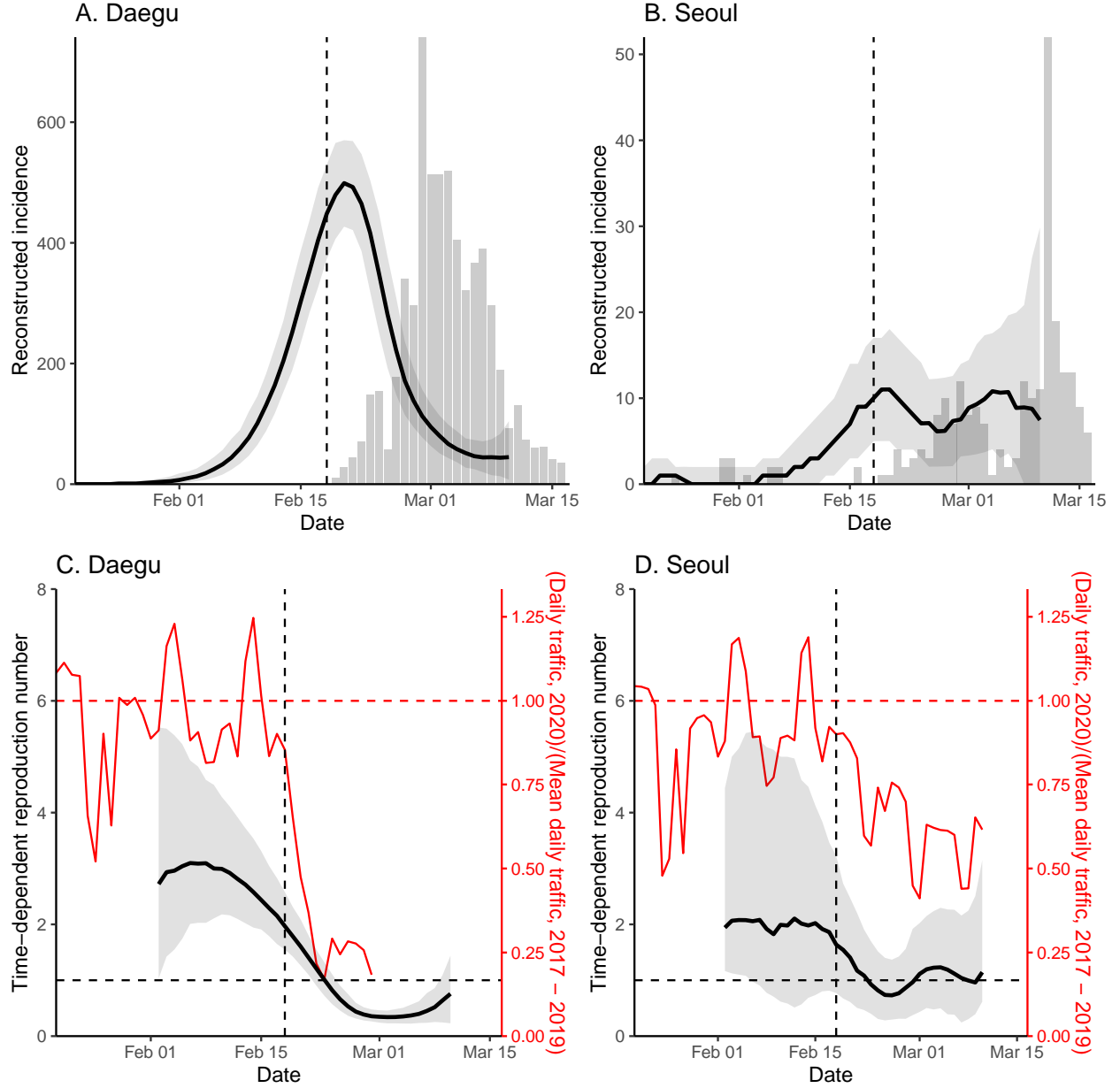
Figure 2: **Comparison of reconstructed incidence and time-dependent reproduction number in Daegu and Seoul.** Black lines and gray ribbons represent the median estimates of reconstructed incidence (A,B) and R_t (C,D) and their corresponding 95% credible intervals. Bar plots show the number of reported cases. Red lines represent the normalized traffic volume. Vertical lines indicate Feb 18, 2020, when the first case was confirmed in Daegu.

Nonetheless, the robustness of our findings is supported by the sensitivity analyses (Supplementary Materials). We were also unable to distinguish local and imported cases, which

may overestimate $\mathcal{R}_t$ [20]. We were able to perform a separate analysis for Seoul that accounts for imported cases using line list provided by the Seoul Metropolitan Government; our qualitative conclusions remained robust (Supplementary Materials).

Our analysis focused on comparing metro traffic, which serves as a proxy for the degree of social distancing, with epidemiological dynamics in two cities. The 80% decrease in traffic volume suggests that the strength of social distancing in Daegu may be comparable to that in Wuhan, China [21]. However, we are not able to directly estimate the effect of social distancing on epidemiological dynamics. Other measures, such as intensive testing of core transmission groups and school closure, are also likely to have affected the changes in $\mathcal{R}_t$ [2]. Future studies should consider quantifying contributions of different measures in preventing the spread.

Finally, our study highlights the importance of considering geographical heterogeneity in estimating epidemic potential. The recent decrease in the number of reported cases in South Korea is driven by the sharp decrease in Daegu. Our analysis reveals that the epidemic may still persist in other regions, including Seoul and Gyeonggi-do; recent reports from Seoul and Gyeonggi-do (around 10 new cases almost every day between March 11–24, 2020) provide further support for our conclusion [2]. Unless the reproduction number can be reduced below 1 in all regions, small outbreaks may continue to occur in South Korea.

References

[1] World Health Organization. Pneumonia of unknown cause – China. 2020. `https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/`. Accessed January 30, 2020.

[2] Korea Centers for Disease Control and Prevention (KCDC). Press release (in Korean). 2020. `https://www.cdc.go.kr/board/board.es?mid=a20501000000&bid=0015#`. Accessed January 20 – March 24, 2020.

[3] Neil M Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. 2020. `https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf`. Accessed March 22, 2020.

[4] COVID-19 National Emergency Response Center, Epidemiology & Case Management Team, Korea Centers for Disease Control & Prevention. Contact transmission of COVID-19 in South Korea: novel investigation techniques for tracing contacts. *Osong Public Health and Research Perspectives*, 11(1):60–63, 2020.

[5] Dennis Normile. Coronavirus cases have dropped sharply in South Korea. What's the secret to its success? *Science*, 2020. `https://www.sciencemag.org/news/2020/03/coronavirus-cases-have-dropped-sharply-south-korea-whats-secret-its-success`. Accessed March 21, 2020.

[6] Seoul Metropolitan City Government. Official website of the Seoul Metropolitan City Government. 2020. `http://www.seoul.go.kr/`. Accessed January 20 – March 24, 2020.

[7] Busan Metropolitan City Government. Official website of the Busan Metropolitan City Government. 2020. `http://www.busan.go.kr/`. Accessed January 20 – March 16, 2020.

[8] Gwangju Metropolitan City Government. Official website of the Gwangju Metropolitan City Government. 2020. `https://www.gwangju.go.kr/`. Accessed January 20 – March 16, 2020.

[9] Ulsan Metropolitan City Government. Official website of the Ulsan Metropolitan City Government. 2020. `https://www.ulsan.go.kr`. Accessed January 20 – March 16, 2020.

[10] Gyeonggi Provincial Government. Official website of Gyeonggi Provincial Government. 2020. `https://www.gg.go.kr/`. Accessed January 20 – March 16, 2020.

[11] Daegu Metropolitan City Government. Official website of the Daegu Metropolitan City Government. 2020. `http://www.daegu.go.kr/`. Accessed January 20 – March 16, 2020.

[12] Gyeongnam Provincial Government. Official website of the Gyeongnam Provincial Government. 2020. `http://www.gyeongnam.go.kr/`. Accessed January 20 – March 16, 2020.

[13] Christophe Fraser. Estimating individual and household reproduction numbers in an emerging epidemic. *PloS one*, 2(8), 2007.

[14] Jantien A Backer, Don Klinkenberg, and Jacco Wallinga. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance*, 25(5), 2020.

[15] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, David G Bonsall, and Christophe Fraser. Quantifying dynamics of SARS-CoV-2 transmission suggests that epidemic control and avoidance is feasible through instantaneous digital contact tracing. *medRxiv*, 2020. `https://www.medrxiv.org/content/10.1101/2020.03.08.20032946v1`. Accessed March 24, 2020.

[16] Tapiwa Ganyani, Cecile Kremer, Dongxuan Chen, Andrea Torneri, Christel Faes, Jacco Wallinga, and Niel Hens. Estimating the generation interval for COVID-19 based on symptom onset data. *medRxiv*, 2020. `https://www.medrxiv.org/content/10.1101/2020.03.05.20031815v1`. Accessed March 24, 2020.

[17] Sam Abbott, Joel Hellewell, James D Munday, June Young Chun, Robin N Thompson, Nikos I Bosse, Yung-Wai Desmond Chan, Timothy W Russell, Christopher I Jarvis, CMMID nCov working group, Stefan Flasche, Adam J Kucharski, Rosalind Eggo, and Sebastian Funk. Temporal variation in transmission during the COVID-19 outbreak. 2020. `https://cmmid.github.io/topics/covid19/current-patterns-transmission/global-time-varying-transmission.html`. Accessed March 21, 2020.

[18] Ilona Kickbusch and Gabriel Leung. Response to the emerging novel coronavirus outbreak. *BMJ*, 368:m406, 2020.

[19] Roy M Anderson, Hans Heesterbeek, Don Klinkenberg, and T Déirdre Hollingsworth. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, 395(10228):931–934, 2020.

[20] RN Thompson, JE Stockwin, RD van Gaalen, JA Polonsky, ZN Kamvar, PA Demarsh, E Dahlqwist, S Li, Eve Miguel, T Jombart, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356, 2019.

[21] Shengjie Lai, Nick W Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R Floyd, Amy Wesolowski, Mauricio Santillana, Chi Zhang, Xiangjun Du, et al. Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. *medRxiv*, 2020. `https://www.medrxiv.org/content/10.1101/2020.03.03.20029843v3`. Accessed March 24, 2020.

[22] Paul-Christian Bürkner et al. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, 80(1):1–28, 2017.

[23] Stan Development Team. RStan: the R interface to Stan, 2020. R package version 2.19.3.

Supplementary Materials

Epidemiological data

The daily number of reported cases from each region was translated and transcribed from the KCDC press release [2]. Following the KCDC's protocol, the daily number of reported cases prior to February 20, 2020, reflects the number of confirmed cases on each day. Between February 21 – March 1, 2020, the daily number of reported cases reflects the number of reported cases within the last 24 hours (9 AM to 9 AM). On March 2, 2020, the daily number of reported cases reflects the number of cases that were reported between 9 AM March 1, 2020, and 12 AM March 2, 2020. Since then, the daily number of reported cases reflects the number of reported cases within the last 24 hours (12 AM to 12 AM). The number of negative cases was not reported on January 25 and 31, 2020; we took the average of cumulative negative cases from one 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 city's 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

Testing criteria expanded 4 times between 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 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.

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 are 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 [22]. 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 [23]. 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:

$$\frac{f(x_i|\mu(t_i),\theta)}{F(t_{\max} - t_i|\mu(t_i),\theta)}, \tag{2}$$

where f is the negative binomial distribution with time-dependent mean $\mu(t_i)$ and dispersion parameter θ. 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 [23].

For each combination of date of infection and a posterior sample of the forward delay distribution, we drew 1000 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 incidence time series to estimate $\mathcal{R}_t$.
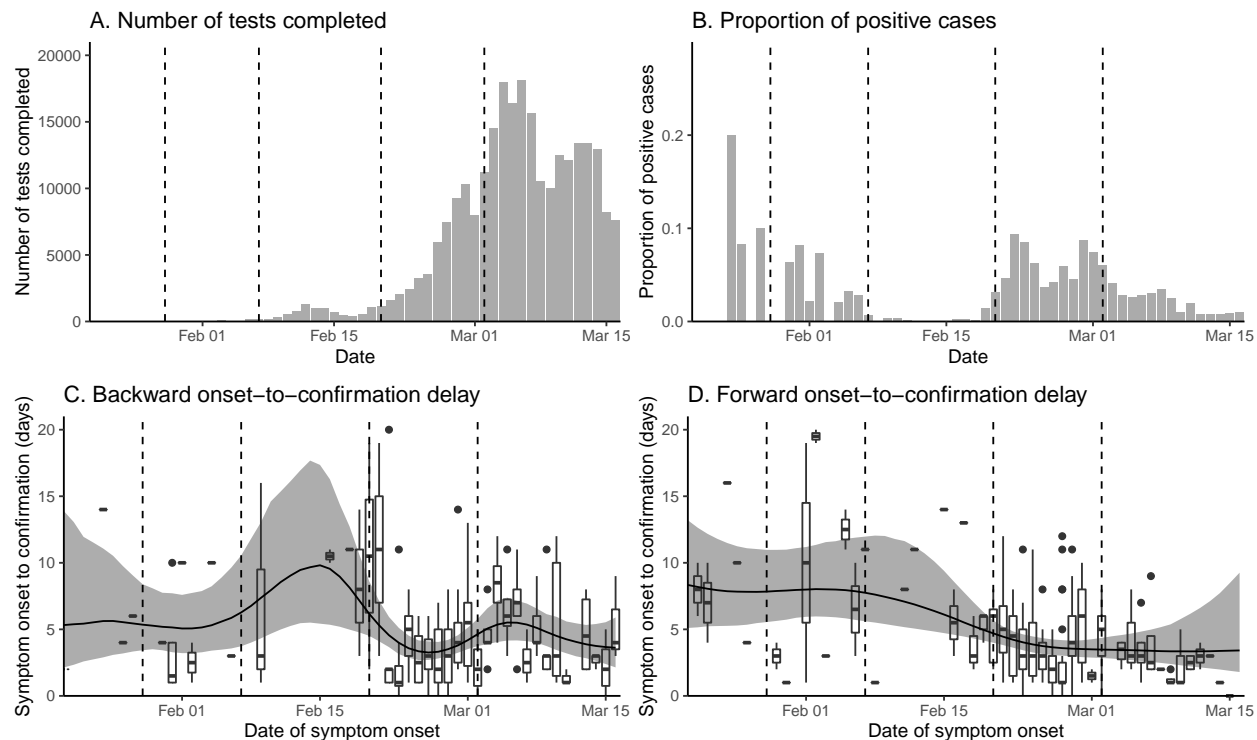
Figure S1: **Changes in the number of tests and delay distributions over time.**
Vertical lines indicate the date on which testing criteria expanded. Box plots (C–D) represent
the observed delays. Black lines and gray ribbons represent the median estimates of the mean
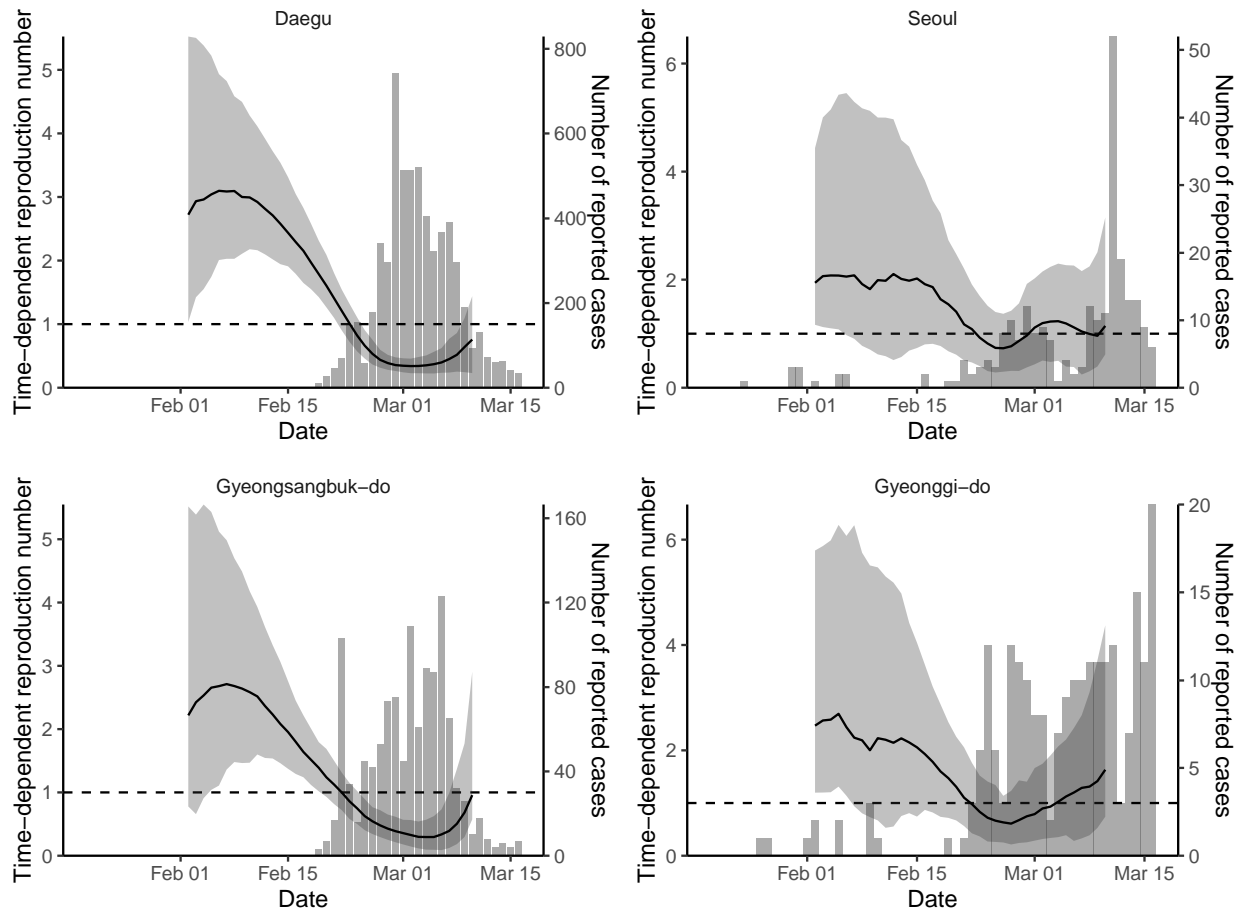delays and their associated 95% credible intervals.

Figure S2: **Comparison of time-dependent reproduction number and the daily number of reported cases in Daegu, Seoul, Gyeongsangbuk-do, and Gyeonggi-do.**
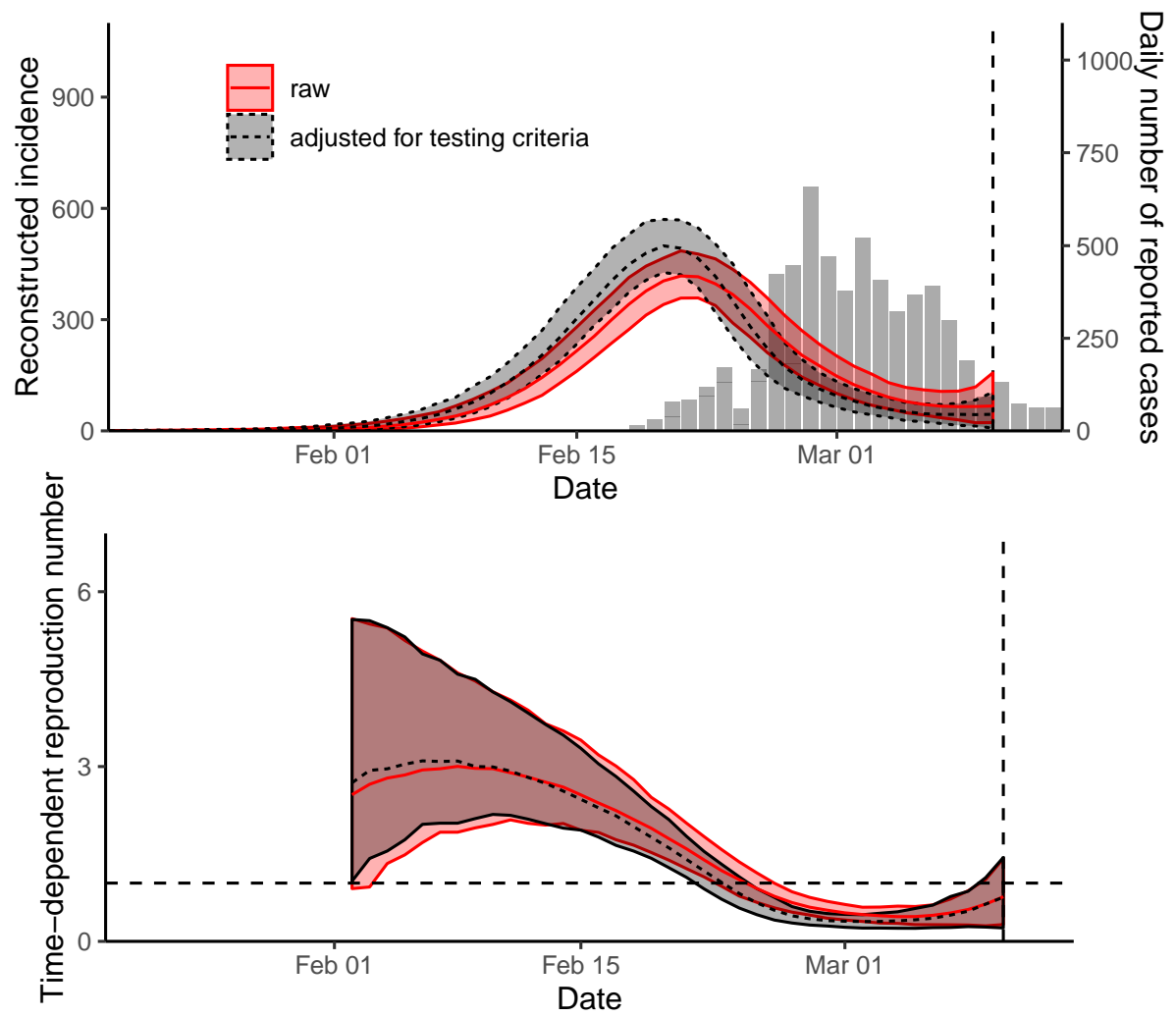
Figure S3: **Sensitivity analysis of estimates of $\mathcal{R}_t$ in Daegu.**
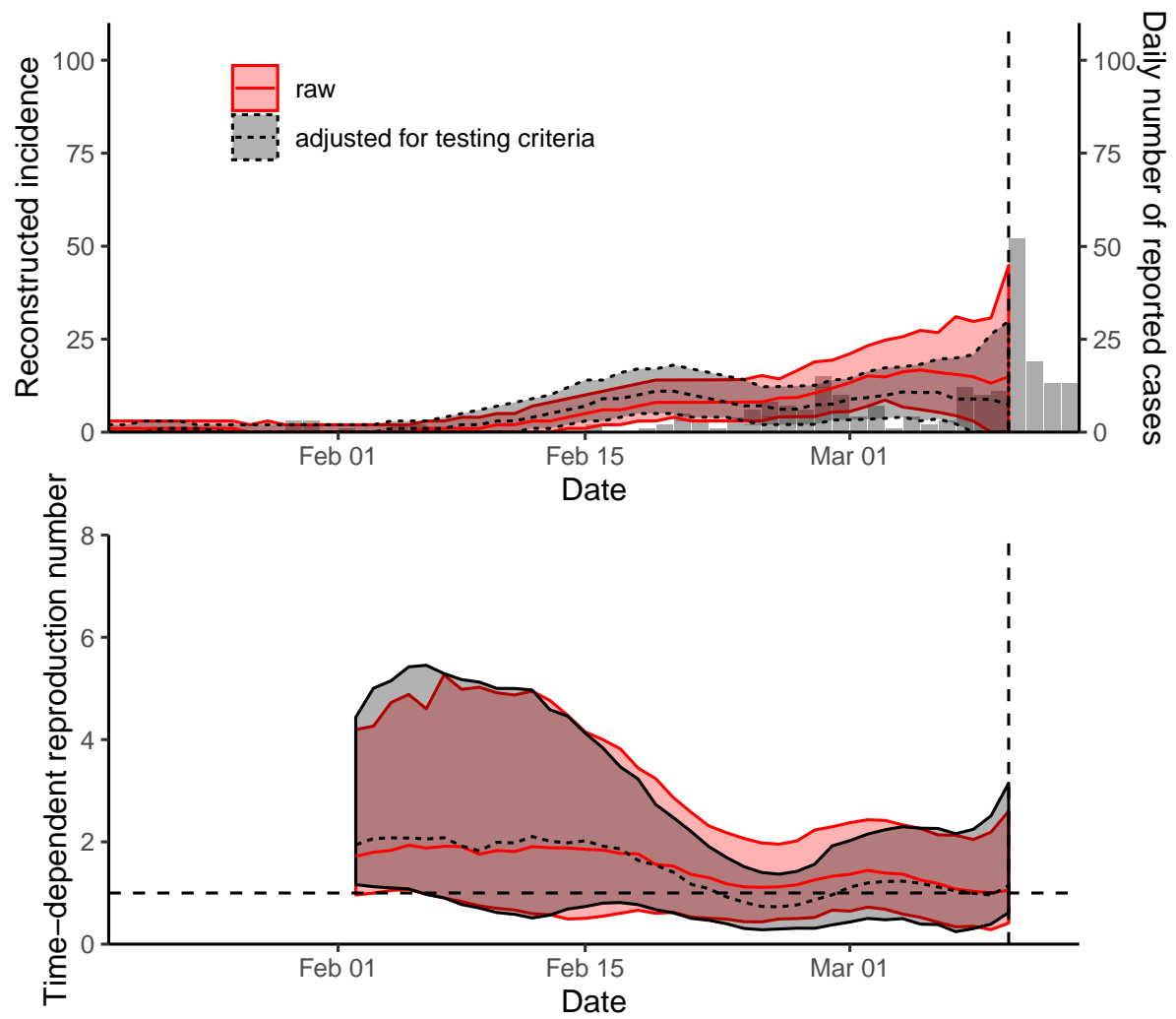
Figure S4: **Sensitivity analysis of estimates of $\mathcal{R}_t$ in Seoul.**
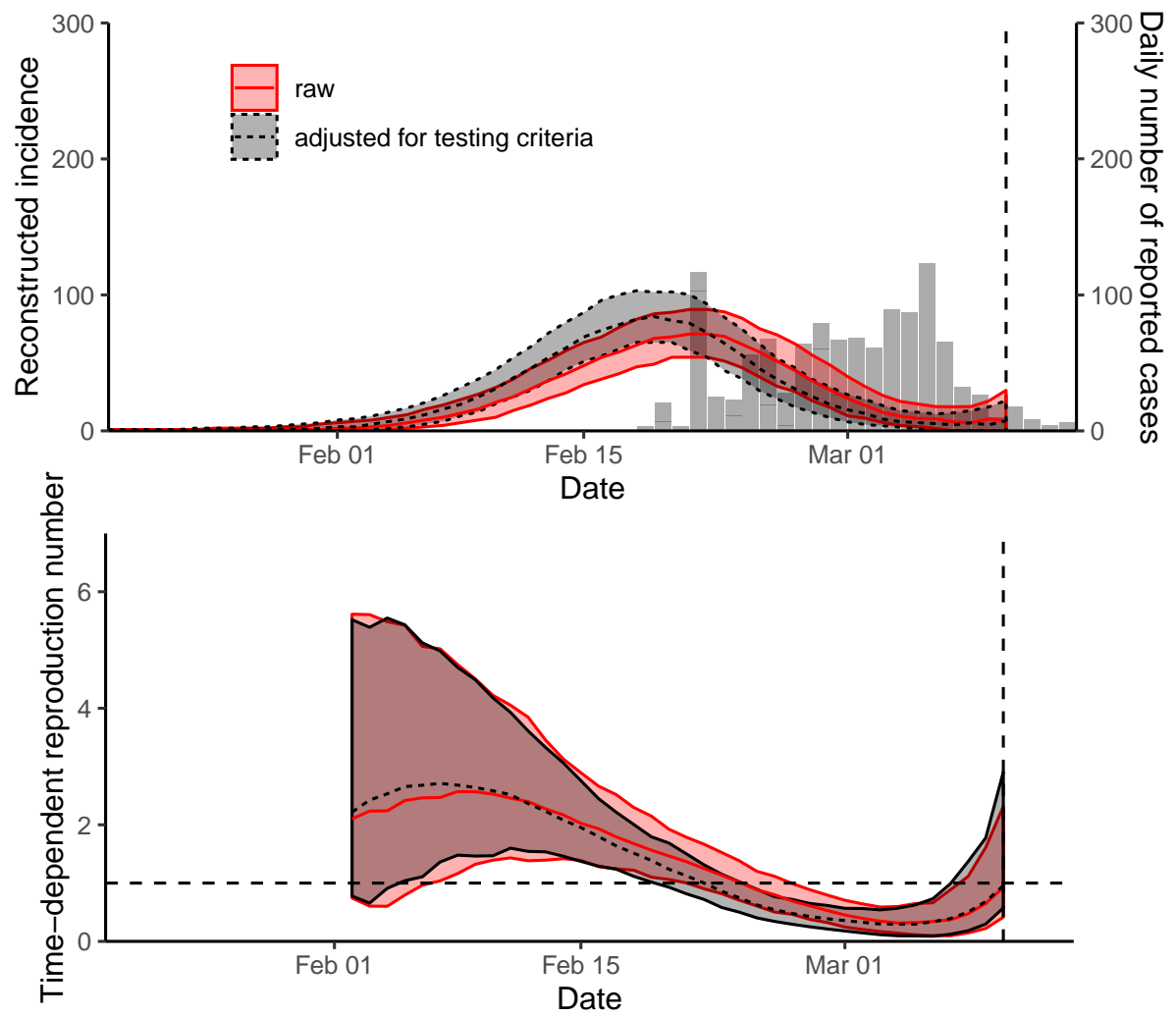
Figure S5: **Sensitivity analysis of estimates of $\mathcal{R}_t$ in Gyeongsangbuk-do.**
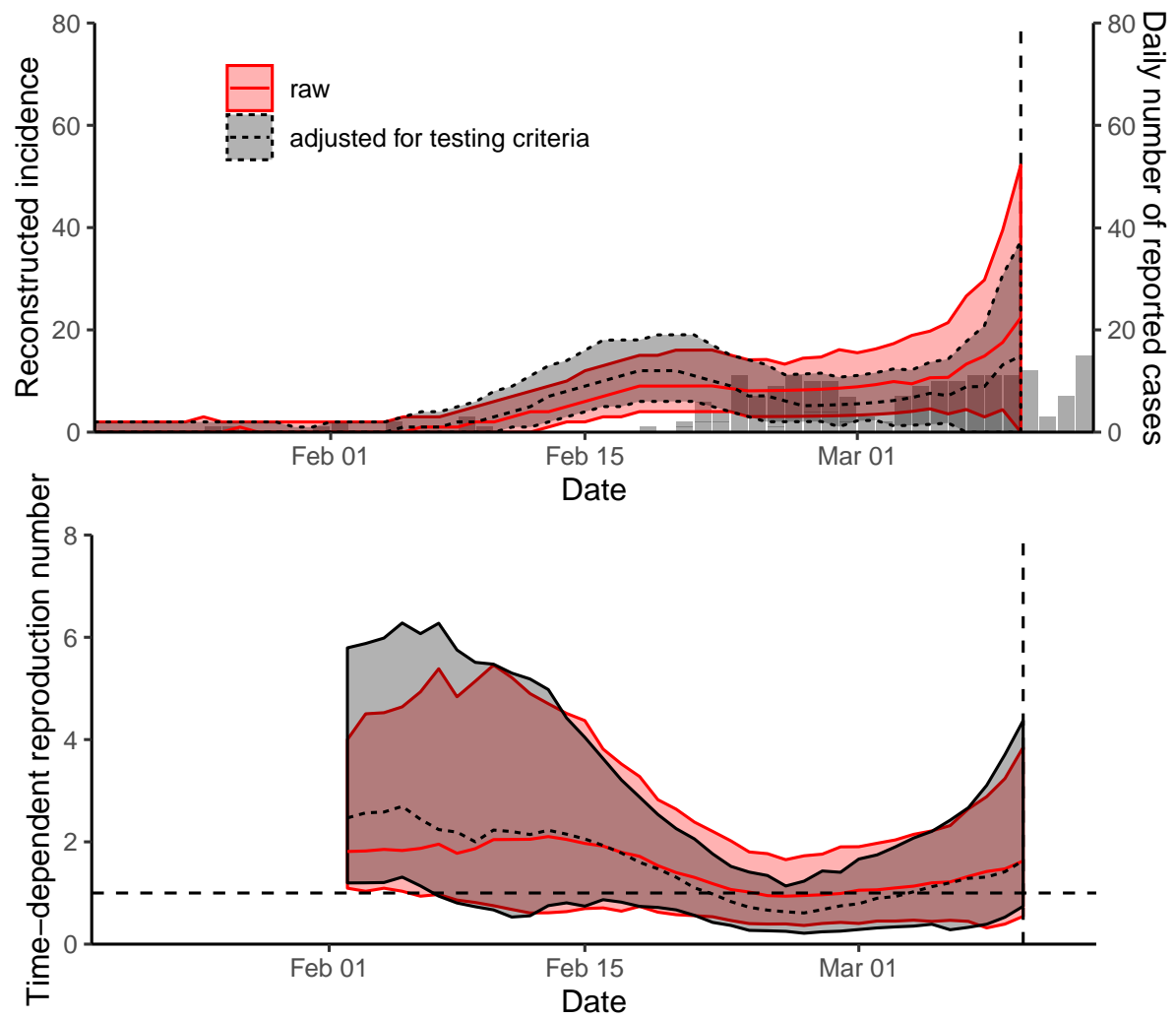
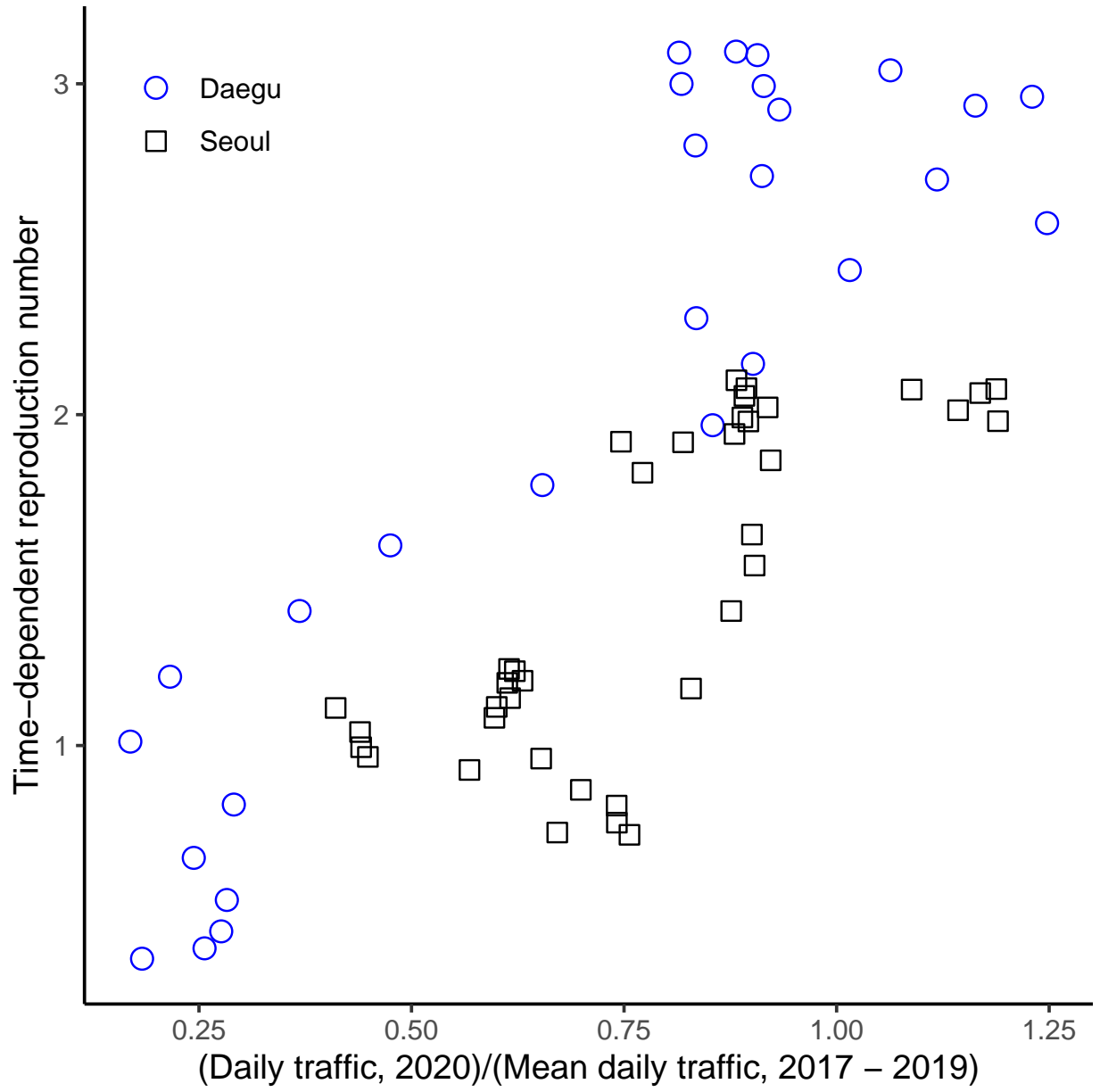Figure S6: **Sensitivity analysis of estimates of $\mathcal{R}_t$ in Gyeonggi-do.**

Figure S7: **Scatter plot of the normalized traffic volume and the median estimates of $\mathcal{R}_t$.**
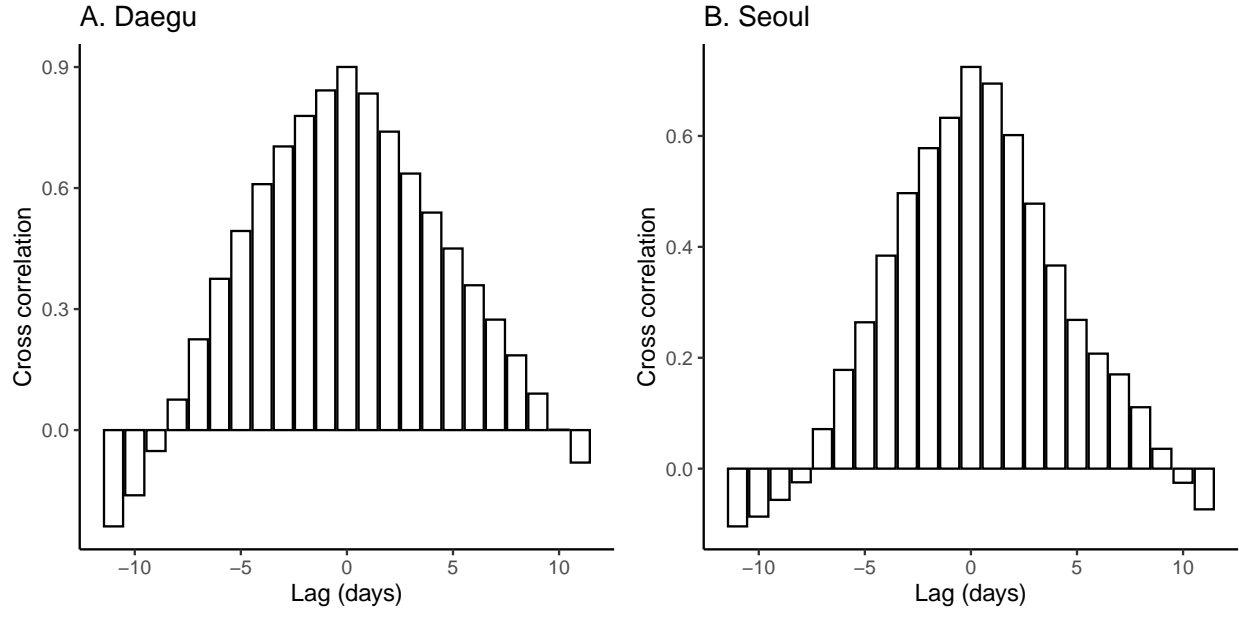
Figure S8:   **Cross correlation between the normalized traffic volume and the median estimates of $\mathcal{R}_t$.**
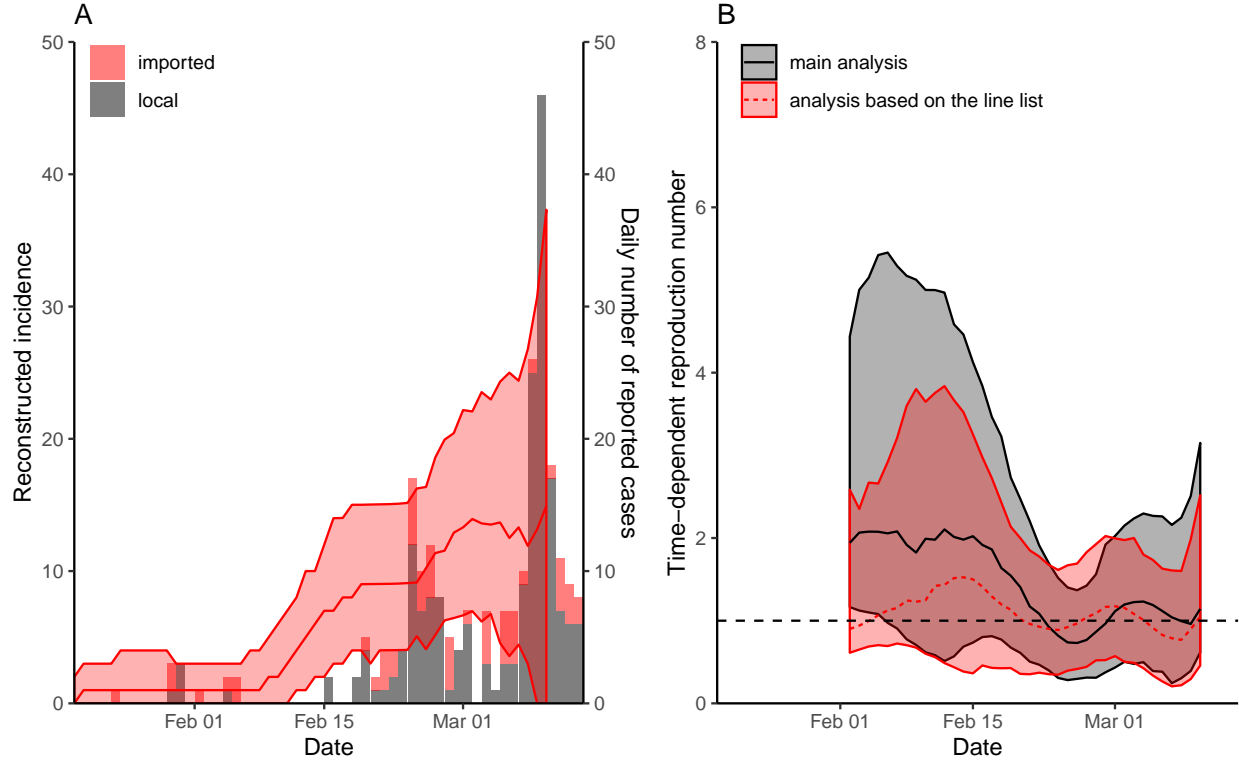
Figure S9: **Comparison of time-dependent reproduction number in Seoul using the number of reported cases by the KCDC and the line list provided by the Seoul Metropolitan Government.** Using line list, we reconstructed incidence for local I_t^{local} and imported I_t^{imported} cases separately based on the method described in the main text. Then, we estimated the time-dependent reproduction number via $\mathcal{R}_t = I_t^{\text{local}} / \sum_{k=1}^{14} I_{t-k} w_k$, where $I_t = I_t^{\text{local}} + I_t^{\text{imported}}$. We did not account for changes in testing criteria in this analysis. The line lists were obtained from `http://news.seoul.go.kr/welfare/archives/513105`.