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] Jonathan Dushoff[3,4,5] Bryan T. Grenfell[1,2,6]

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

*Corresponding author: swp2@princeton.edu

Since its first appearance in Wuhan, China, in December 2019 [1], the novel 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]. The epidemic then took off in mid-February when COVID-19 began to spread within a church from a city of Daegu — as of March 18, 2020, 8,413 cases have been confirmed, of which 60% are related to the church [2]. While its success in suppressing the epidemic has been widely attributed to its intensive testing [3, 4], other factors, such as social distancing, are also likely to have played important roles. Here, we describe potential roles of social distancing in mitigating the spread of COVID-19 by using metro traffic data to compare epidemics in two geographically separated major cities in South Korea.

1 Data description

We analyzed epidemiological data, collected bewteen January 20–March 16, 2020, describing the COVID-19 outbreak in South Korea. Daily number of reported cases in each geographic region was transcribed from press releases by Korea Centers for Disease Control and Prevention (KCDC) [2]. Partial line lists were transcribed from press releases from KCDC and various local and provincial governments. All data and original reports are stored in a publicly available GitHub repository: `https://github.com/parksw3/COVID19-Korea`.

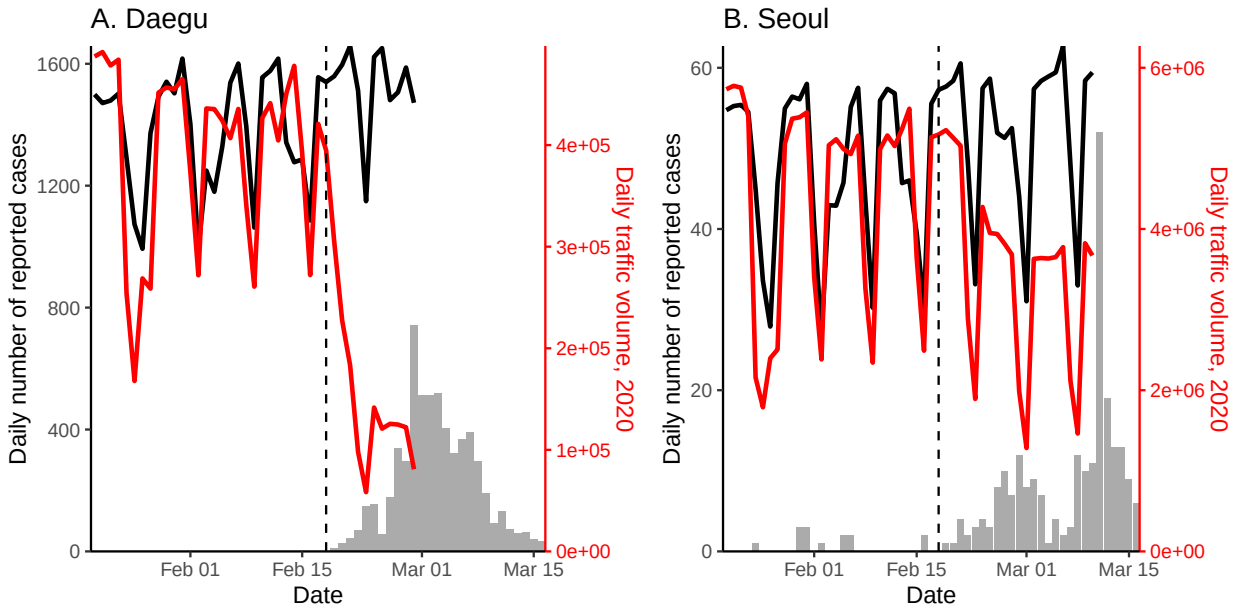

Figure 1: **Comparison of epidemiological and traffic data from Daegu and Seoul.** Solid lines represent the daily metro traffic volume in 2020 (red) and mean daily traffic volume between 2017–2019 (black). Daily traffic from previous years have been shifted by 1–3 days to align day of the weeks. Vertical lines indicate Feb 18, 2020, when the first case was confirmed in Daegu.

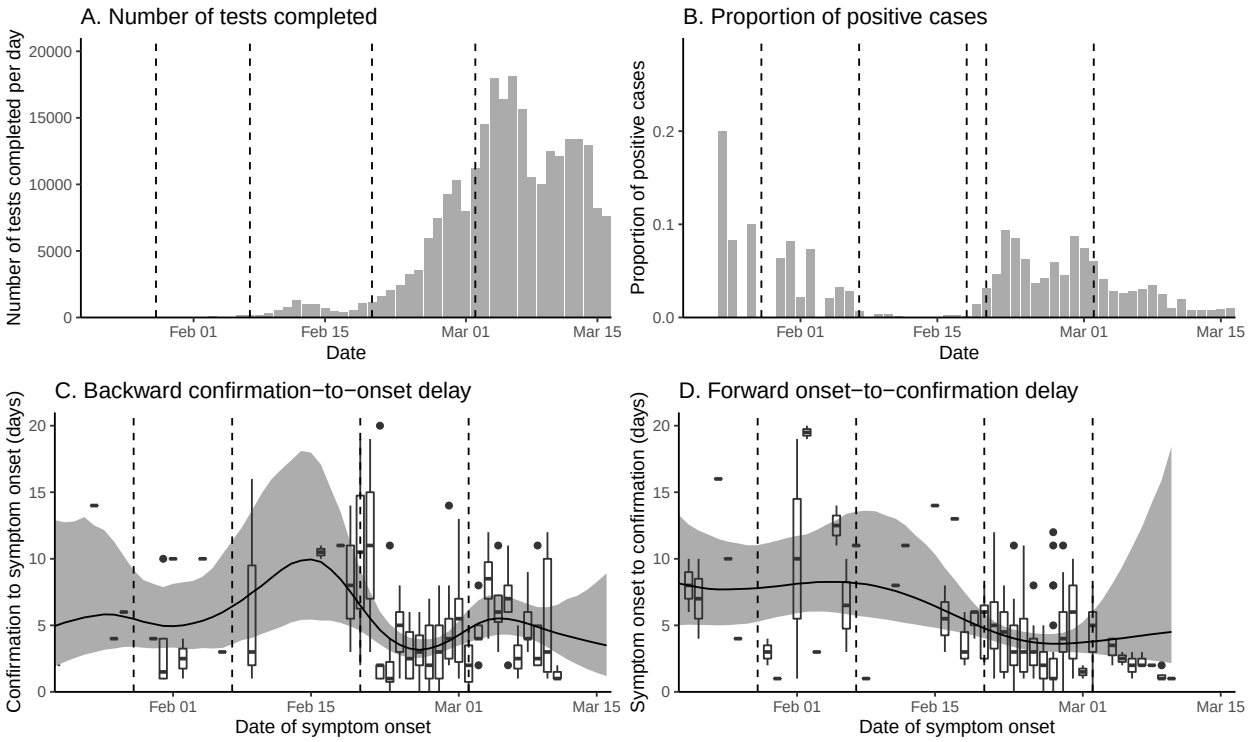

Figure 2: **Changes in the number of tests and delay distributions over time.** Vertical lines indicate the date on which testing criteria expanded. Boxplots (C–D) represent the observed delays. Black lines and gray ribbons represent the median estimates of the mean delays and their associated 95% credible intervals.

We compared epidemiological dynamics of COVID-19 from two cities in which the biggest number of COVID-19 cases have been reported: Daegu and Seoul. Between January 20–March 16, 2020, 6,083 cases were reported from Daegu and 248 from Seoul. Unlike the epidemic in Daegu, which is characterized by a single, large peak followed by a gradual decrease, 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 compared the the daily number of individuals who got on the subway — including a monorail in Daegu — across all stations (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% compared to previous years in Daegu and Seoul, respectively.

	Parameterization	Priors	Source
Incubation period distribution	$\text{Gamma}(\mu_I, \mu_I^2/\sigma^2)$	$\mu_I \sim \text{Gamma}(6.5\,\text{days}, 145)$ $\sigma \sim \text{Gamma}(2.6, 25)$	[7]
Generation-interval distribution	$\text{NegativeBinomial}(\mu_G, \theta)$	$\mu_G \sim \text{Gamma}(5\,\text{days}, 62)$ $\theta \sim \text{Gamma}(5, 20)$	[8, 9]

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 estiamtes.

2 Trends in time-dependent reproduction number and traffic volume

In order to estimate the time-dependent reproduction number R_t (i.e., the expected number of secondary cases caused by an individual infected at time t [5]), we first accounted for changes in testing criteria (Fig. 2B) by multiplying the daily number of reported cases by the relative detection rate for each criterion (i.e., a proportion of cases tested positive based on a criterion divided by the mean detection rate). Likewise, we accounted for changes in the number of tests completed on each day (Fig. 2A) by dividing the scaled number of reported cases by the relative number of tests completed on each day; this step was performed separately for each testing criterion because widening of criteria necessarily increases the number of tests. A sensitivity analysis showed that qualitative patterns of inferred R_t are robust to these adjustments (Supplementary Materials).

We then estimated time-dependent *backward* confirmation-to-onset delay distributions from the partial line list using a negative-binomial regression with `brms` package [6] (Fig. 2C). We combined the estimated posterior distribution of delay distributions with previously estimated incubation period distribution (Table 1) to obtain posterior samples for date of infection for each confirmed case. Likewise, we estimated time-dependent *forward* onset-to-confirmation delay distribution using the same negative-binomial regression model, while accounting for right-censoring (Fig. 2D), to calculate the degree of censoring (i.e., median probability that a case infected on a given day will be reported after March 16, 2020). Details are provided in the Supplementary Materials.

For each posterior sample of the reconstructed incidence time series (i.e., the number of infected individuals on day t), we divided incidence by $1-$degree of censoring and estimated the time-dependent reproduction number using the renewal equation with a 14-day sliding window [5]:

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

where I_t is the censoring-adjusted incidence time series and w_k is the generation-interval distribution randomly drawn from a prior distribution (Table 1). We restricted the calculation of $\mathcal{R}_t$ between February 2, 2020 (14 days after the first confirmed case was imported)

and March 10, 2020 (because the degree of censoring would be too large after this date to estimate $\mathcal{R}_t$).

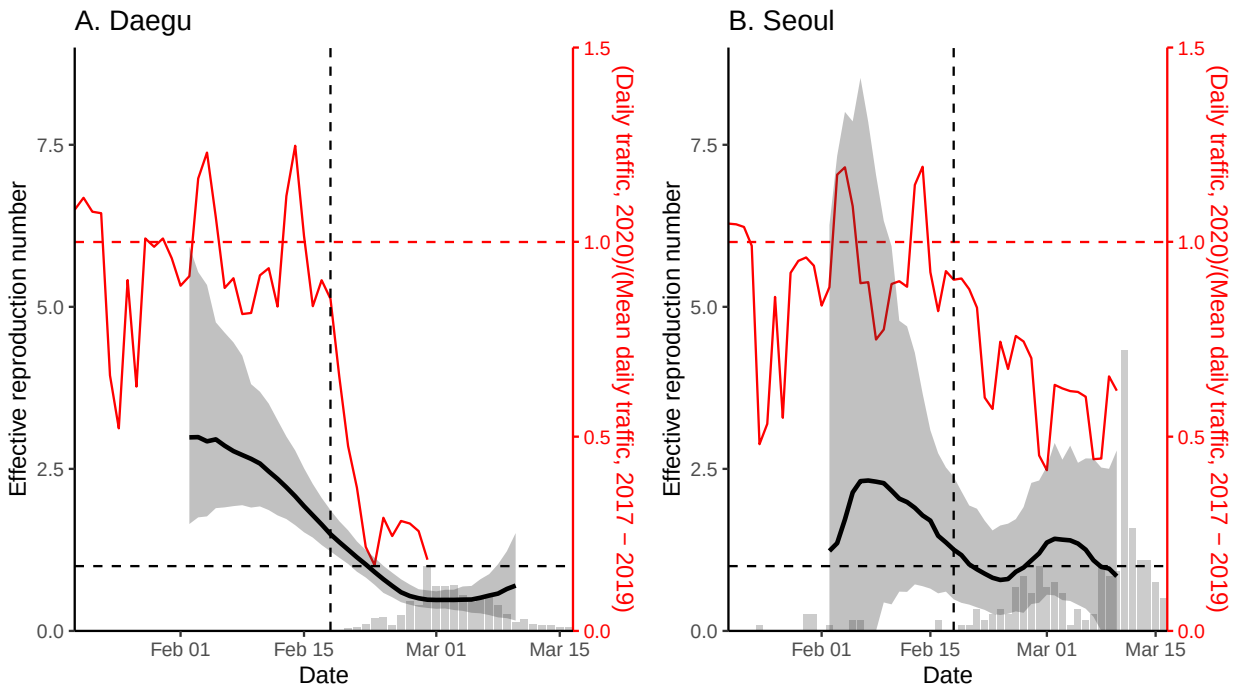

Figure 3: **Comparison of time-dependent reproduction number and normalized traffic in Daegu and Seoul.** Black lines and gray ribbons represent the median estimates of R_t and their corresponding 95% credible intervals. Red lines represent the normalized traffic volume. Vertical lines indicate Feb 18, 2020, when the first case was confirmed in Daegu.

Fig. 3 compares the estimates of $\mathcal{R}_t$ in Daegu and Seoul. In Daegu, 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. 3A). Our estimates of $\mathcal{R}_t$ for Daegu are consistent with the estimates of $\mathcal{R}_t$ for South Korea by Abbott *et al.* [10] — their estimates drop below 1 slightly later because they rely on number of symptomatic cases instead. On the other hand, estimates of $\mathcal{R}_t$ remain around 1 in Seoul (Fig. 3B); while social distancing is likely to have assisted in mitigating the spread in Seoul, our analysis suggests that 50% decrease in metro traffic was insufficient to suppress the spread. Stronger distancing or further intervention will be necessary to reduce $\mathcal{R}_t$ below 1. Similar patterns in the estimates of $\mathcal{R}_t$ are found in directly surrounding provinces, providing support for the robustness of our analysis (Supplementary Materials).

3 Discussion

The ongoing COVID-19 outbreak in South Korea provides a unique perspective to understanding and controlling the pandemic. Its experience provides evidence that the epidemic can be suppressed without draconian measures as in China [11]. 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. 3B) reduces burden for healthcare system and provides time to plan for the future [12].

Our study is not without limitations. In particular, we did not account for differences in the delay distributions or changes in the number of tests or detection rates 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 the data to account for these factors. Nonetheless, the robustness of our findings is supported by sensitivity analyses.

Our analysis focused on comparing metro traffic, which serves as a proxy for the degree of social distancing, with epidemiological dynamics in two cities. However, our analysis does not allow us to measure the direct "effect" of social distancing on epidemiological dynamics. Moreover, 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 studies highlights the importance of characterizing geographical heterogeneity in estimating epidemic potential. The recent decrease in the number of reported cases in South Korea is likely to be strongly correlated with the decrease in the number of cases in Daegu. Our analysis reveals that the epidemic may still persist in other regions, including Seoul and Gyeonggi-di. If the reproduction number cannot be reduced below 1 in all regions, small outbreaks may continue to occur in South Korea.

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.

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 Jan 20 – Mar 16, 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. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf. Accessed Mar 22, 2020.

[4] 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 Mar 21, 2020.

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

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

[7] 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.

[8] 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.

[9] 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.

[10] 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. *Science*, 2020. https://cmmid.github.io/topics/covid19/current-patterns-transmission/global-time-varying-transmission.html. Accessed Mar 21, 2020.

[11] Ilona Kickbusch and Gabriel Leung. Response to the emerging novel coronavirus outbreak, 2020.

[12] 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*, 2020.

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

[14] Andrew Gelman, Donald B Rubin, et al. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

Supplementary Materials

Estimating delay distributions between onset and reporting

We estimated both forward (onset-to-confirmation) and backward (confirmation-to-onset) delay distributions using Bayesian negative binomial regressions with log links. The backward delay distribution is inferred directly using the `brms` package [6]. The 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 devition of 2; note that these distributions are priors on link scale.

To estimate the forward delay distribution, we modified the stan code to account for right-censoring and ran the code using the `rstan` package [13]. In particular, we modified the likelihood such that given a delay of x_i days, symptom onset day t_i and the day of measurement of k, the likelihood of observing the delay is given by:

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

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 $k-t_i$ (otherwise, the case will be reported after the date of measurement). Convergence is assessed by ensuring that the Gelman-Rubin statistic is lower than 1.01 for all parameters [14].

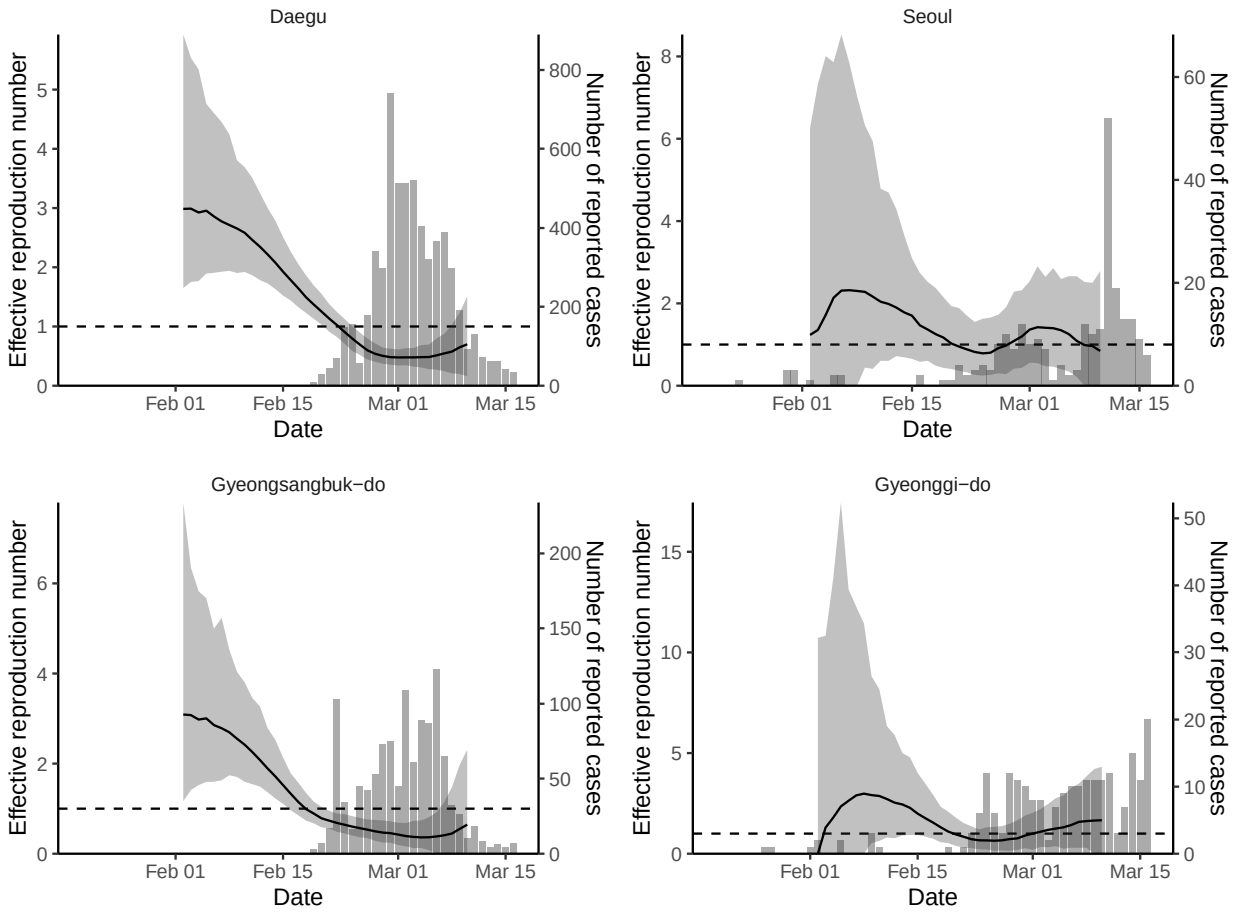

Figure S1: **Comparison of effective reproduction number and the daily number of reported cases in Daegu, Seoul, Gyeongsangbuk-do, and Gyeonggi-do.**

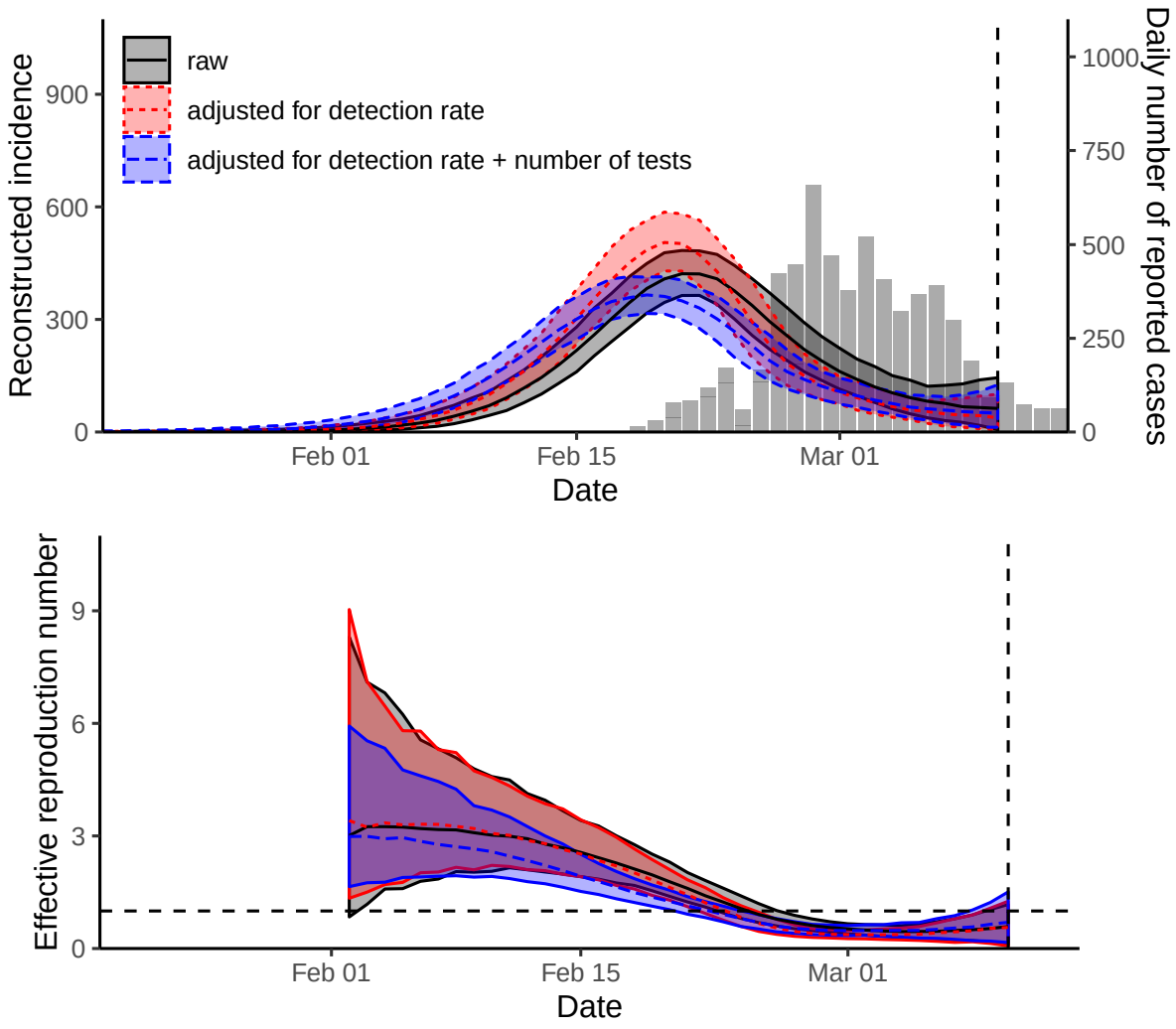

Figure S2: **Sensitivity analysis of estimates of $\mathcal{R}_t$ in Daegu.**

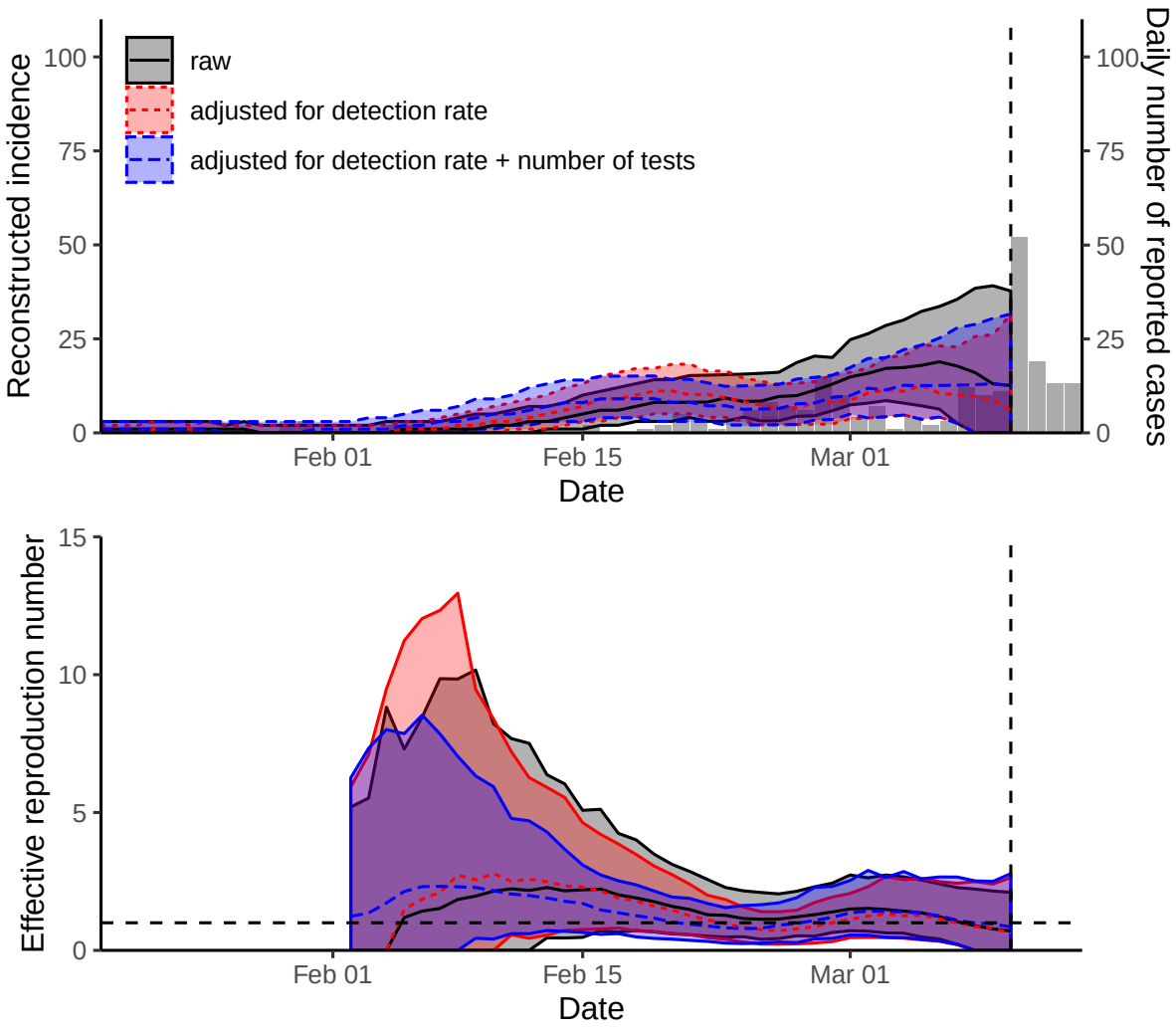

Figure S3: **Sensitivity analysis of estimates of $\mathcal{R}_t$ in Seoul.**

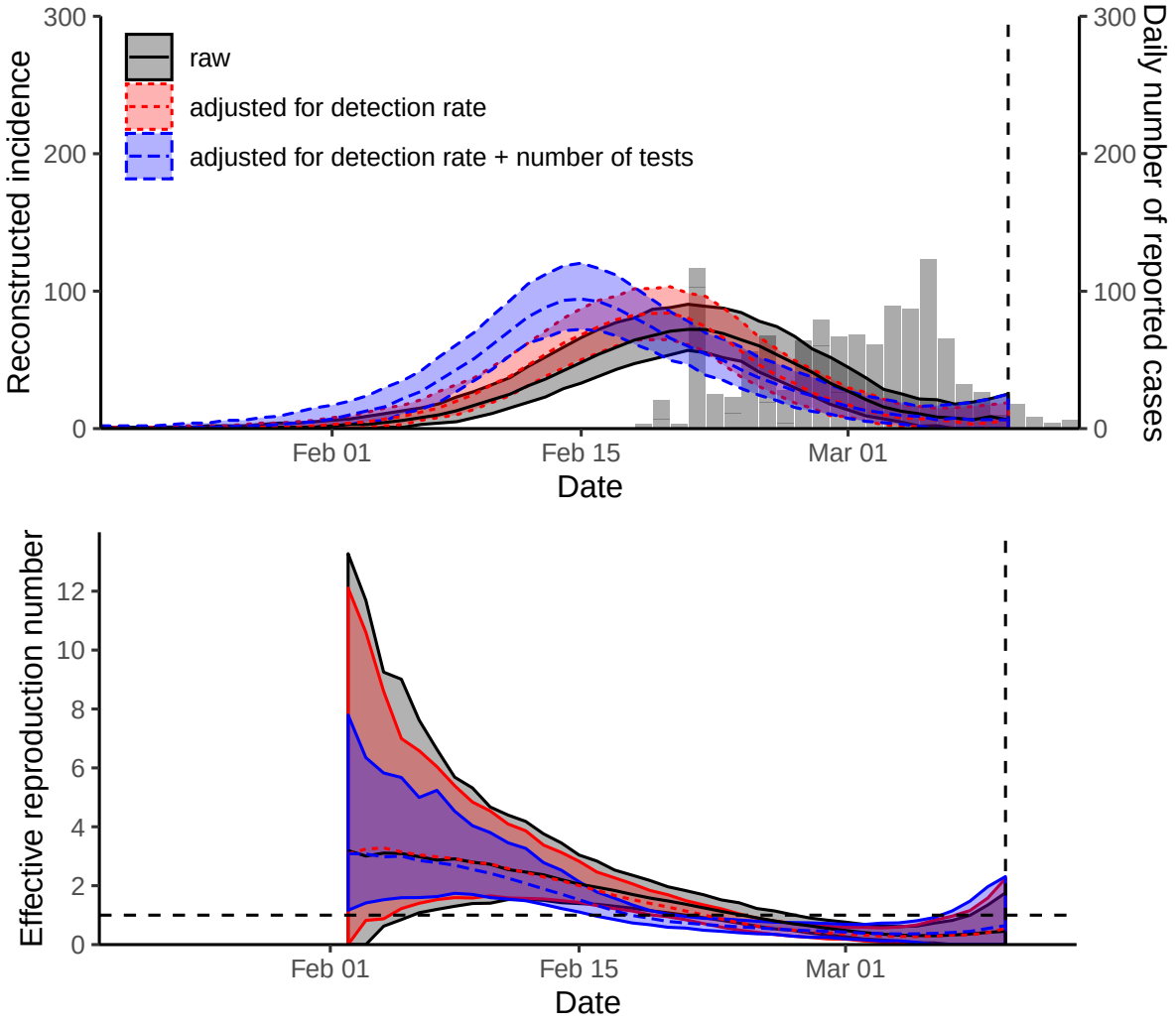

Figure S4: **Sensitivity analysis of estimates of $\mathcal{R}_t$ in Gyeongsangbuk-do.**

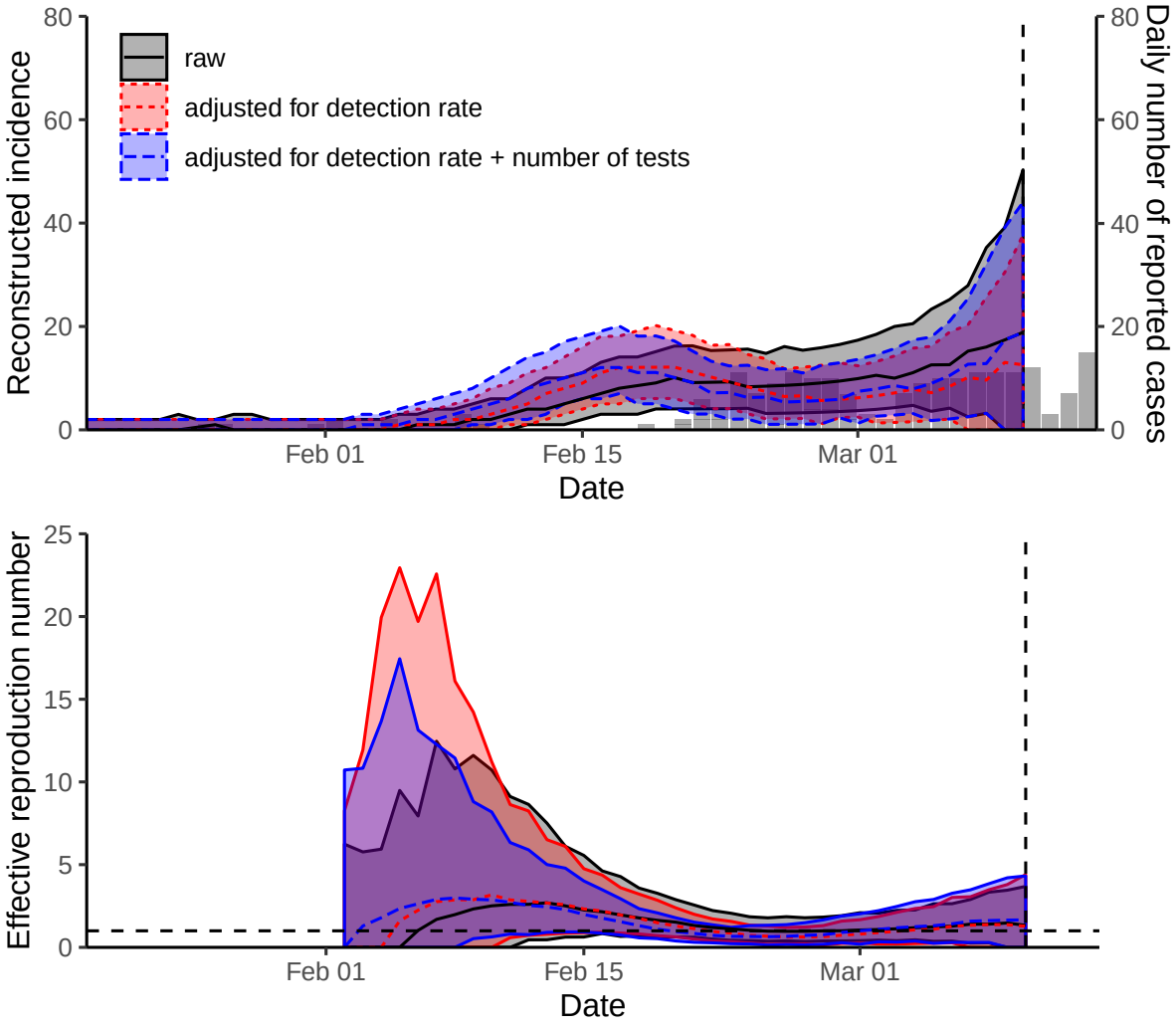

Figure S5: **Sensitivity analysis of estimates of $\mathcal{R}_t$ in Gyeonggi-do.**