1 **Title: Basic and effective reproduction numbers of COVID-19 cases in South Korea**
 2 **excluding Sincheonji cases**

3
 4 **Authors:** Jeongeun Hwang¹, Hyunjung Park², Jiwon Jung^{3,4}, Sung-Han Kim^{3,4}, Namkug Kim⁵,
 5 *

6 **Affiliations:**

7 ¹Department of Medicine, University of Ulsan College of Medicine, Asan Medical Center,
 8 Seoul, Republic of Korea

9 ²Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea

10 ³Department of Infectious Diseases, Asan Medical Center, University of Ulsan College of
 11 Medicine, Seoul, Republic of Korea

12 ⁴Office for Infection Control, Asan Medical Center, Seoul, Republic of Korea

13 ⁵Department of Convergence Medicine, Asan Medical Center, University of Ulsan College of
 14 Medicine, Asan Medical Center, Seoul, Republic of Korea, 05505

15 * Corresponding author

16

17 **Address for Correspondence:**

18 Namkug Kim, PhD

19 Department of Convergence Medicine, Asan Medical Center, 88, Olympic-ro 43-gil, Sonpa-
 20 gu, Seoul, 05505, Republic of Korea

21 E-mail: namkugkim@gmail.com

22

23 **Funding:**

24 Not applicable

25

26 **Abstract**

27 In February and March 2020, COVID-19 epidemic in South Korea met a large black swan
 28 effect by a Sincheonji cult mass infection in Daegu-Gyeongbuk area. The black swan made
 29 it difficult to evaluate that the current policies for infection prevention including social
 30 distancing, closing schools, hand washing, and wearing masks good enough or not.
 31 Therefore, in this study, we evaluated basic reproduction number (R_0) and time-dependent
 32 reproduction number (R_t) of confirmed cases based on various kinds of populations,
 33 including total, Daegu-Gyeongbuk, except-Daegu-Gyeongbuk, Sincheonji, and except-
 34 Sincheonji. In total, it seems the infection is going to be under control, but this is never true
 35 because in the except-Sincheonji and except-Daegu-Gyeongbuk cases, R_0 is still above 1.0,
 36 and R_t is drifting around 1.0. This study could be used to determine government policies in
 37 the near future.

38

39 **Keywords:** COVID-19

40

41 **Text**

42 South Korea currently has 9,241 cumulated confirmed COVID-19 cases as of March 26th,
 43 2020. There was a large “black swan” event called the Sincheonji cult mass infection
 44 incidence that lead to an exponential increase of confirmed cases. Since then, a great effort
 45 has been made to defend the public from virus, and to some extent, these efforts have
 46 suppressed the spread.

47 The basic and time-dependent effective reproduction number R_0 and R_t are two of the most
 48 representative characteristics for the dynamics of the infectious virus. To observe the basic
 49 features of the COVID-19 outbreak in South Korea without the black swan effect, we
 50 collected the daily counts of confirmed cases. The Korean Centers for Disease Control and
 51 Prevention (KCDC) has investigated, summarized, and released daily updates on the
 52 cumulated counts of confirmed cases, which are available online¹. The cumulated counts
 53 have been provided in total, according to the administrative provinces (Gwangyeok-Si and
 54 Do in the South Korean administrative system), and several mass infection cases have
 55 occurred. We collected counts of daily confirmed cases from the Sincheonji cult, which was
 56 named after a church with a COVID-19 outbreak, and subtracted them from the total counts
 57 to comprise the “except Sincheonji” cases. Additionally, we collected the case counts from
 58 Daegu-si and Gyeongsangbuk-do province, which surrounds Daegu-si, because the initial
 59 outbreak in the Sincheonji cult took place in those areas. We summed up the counts from
 60 the two provinces and subtracted them from the total counts to total the “except Daegu-
 61 Gyeonbuk” cases. The Sincheonji cult cases and Daegu-Gyeongbuk cases are not mutually
 62 inclusive nor exclusive. Because it took days to identify a confirmed case belonging to the
 63 Sincheonji cult, the time frames of the Sincheonji cases could have been confounded.

64

65 **Figure 1.** Cumulated counts of confirmed cases of the subgroups, including the total,
66 Daegu-Gyeoengbuk, except-Daegu-Gyeongbuk, Sincheonji, and except-Sincheonji cases.

67

68 The cumulated counts from January 21st, 2020 (the first confirmed case of COVID-19 in
69 South Korea) to March 26th, 2020 are shown in Figure 1. The first Sincheonji cult case
70 appeared on February 18th, 2020, and it was the first case in the Daegu-Gyeongbuk area.
71 Qualitatively, the cumulated counts curves for Daegu-Gyeongbuk and Sincheonji looks like
72 they are reaching for a plateau, but the curves for the except-Sincheonji and except-Daegu-
73 Gyeongbuk cases do not.

74 R statistics software version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria)
75 and the R0 package² was used. For the estimation of the subgroups' R_0 values, we fitted an
76 exponential growth model³. Sensitivity analysis of the exponential growth period was

performed. The generation time distribution should be obtained from the time lag between all infector–infectee pairs⁴, which is called a serial interval, but it is currently unavailable. Instead, we assumed that the serial interval had a gamma distribution with mean $\pm$ standard deviation of 2.0 ± 1.0 , 3.0 ± 1.5 , 4.0 ± 2.0 , 5.0 ± 2.5 , and 6.0 ± 3.0 days. The serial interval could be approximated by referring to the estimated incubation time of COVID-19. Recent research by Guan et al.⁵ suggested that the virus has an incubation period of four days with an interquartile range two to seven days. Some studies have estimated a wider range for the incubation period; data for human infection with other coronaviruses (e.g. MERS-CoV, SARS-CoV) suggested that the incubation period may range from two to 14 days⁶. Sensitivity tests for the generation time assumptions were performed (Figure S1).

Figure 2. R_0 with 95% confidence intervals in the subgroups, including the total, Daegu-Gyeongbuk, except Daegu-Gyeongbuk, Sincheonji, and except Sincheonji groups.

Figure 2 shows the estimated R_0 for each study subgroup and their 95% confidence intervals (by simulation) with an assumption of 4.0 ± 2.0 (mean $\pm$ standard deviation). In the sensitivity test, the R_0 values were sensitive to the generation time assumptions; therefore, the numbers themselves are not reliable in the current study. On the contrary, the relative scales among the R_0 values were robust to the generation time assumptions.

96 The Daegu-Gyeongbuk cases showed significantly higher R_0 than the except-Daegu-
 97 Gyeongbuk cases, and the Sincheonji cases showed significantly higher R_0 than the except-
 98 Sincheonji cases. In all subgroups, R_0 was larger than 1.0. R_0 is an initial-stage
 99 characteristic before the health authorities take action or social awareness rises. Since the
 100 Sincheonji cult mass infection incidence was noticed, health authority actions including
 101 COVID-19 tests, isolation, active surveillance, and school closures, were executed
 102 immediately. In addition, social awareness of COVID-19 soared, and South Korean residents
 103 began to more actively follow hygiene recommendations, including wearing masks, washing
 104 hands frequently, and practicing social distancing in everyday life, throughout the nation.

105 Adding to such efforts to fight the disease, we measured the time dependent reproduction
 106 rate R_t ⁷ for each subgroup. We again assumed the generation time to gamma distribution
 107 with a 4.0 ± 2.0 mean $\pm$ standard deviation.

Figure 3. Changes of R_t with 95% confidence intervals of subgroups, including the total, Sincheonji, except-Sincheonji, Daegu-Gyeongbuk, and except-Daegu-Gyeongbuk groups. A week after the first Sincheonji cult mass infection case (February 24th, 2020), R_t for all subgroups decreased for a couple of days. In the following weeks (around March 2nd, 2020 and after) Sincheonji and Daegu-Gyeongbuk cases fell under the $R_t = 1.0$ horizon, but R_t for the except-Sincheonji and except-Daegu-Gyeongbuk cases were drifting around the 1.0 horizon.

In Figure 3, R_t with 95% confidence intervals in the subgroups, including the total, Sincheonji, except-Sincheonji, Daegu-Gyeongbuk, and except-Daegu-Gyeongbuk subgroups, are shown. In all subgroups, R_t gradually decreased until the end of February. R_t for the Sincheonji and Daegu-Gyeongbuk cases are staying under $R_t = 1.0$ horizon, but in the except-Sincheonji and except-Daegu-Gyeongbuk cases, R_t is drifting around the $R_t = 1.0$ horizon currently (as of March 26th, 2020).

Although the analysis in Figure 3 is preliminary, and thus, we should be very cautious about its interpretations, we see a great uncertainty that South Korea is still in danger of COVID-19. Another instance of the mass infection case could happen if we lose the current momentum in social distancing and preventive action, even if no black swan comes. Assessing only the total counts may blind our sense of danger. Therefore, analyzing the except-Sincheonji and except-Daegu-Gyeongbuk cases separately from the total cases would be recommended. We are releasing the R_t estimates' daily updates for various regional subgroups of South Korea online; <http://covid19.mi2rl.co/>

Our current analysis has substantial limitations. Because we collected the daily press releases from the KCDC, there could be re-classified cases not reflected in our analysis. There could also be misclassification, because a Sincheonji cult may deny his or her beliefs

in the KCDC investigations. Imported cases were not accounted for in the current analysis, though the cumulated imported cases until March 17th, 2020 was 54, according to KCDC press release. Another source of confounding is that we are not aware of the true number of cases. In a recent study in China, Li et al.⁸ estimated that only 14% of infected cases have been confirmed by the authorities. Nevertheless, qualitative interpretations may stay robust to under-reporting if they are constant. Delays in confirmation and in classification are another source of error. The uncertainty in the generation time is also an issue, but it could be overcome by performing sensitivity analysis. If more detailed data on infector–infectee interactions becomes available, more reliable analysis could be achieved.

To conclude, in total, it seems that the virus is going to be under control, but this is never true because if we analyze the except-Sincheonji and except-Daegu-Gyeongbuk cases, the basic reproduction rate is still above 1.0, and the effective reproduction rate R_t is still higher than 1.0. The current efforts to suppress the COVID-19 outbreak should at least be maintained.

Disclosure

All authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Hwang J, Kim S, Jung J, Kim N. Methodology: Hwang J, Kim N. Formal analysis: Hwang J, Kim N. Data curation: Par H, Software: Hwang J, Kim N. Investigation: Jung J, Kim S, Kim N. Writing – original draft preparation: Hwang J, Park H. Writing – review and editing: Jung J, Kim S, Kim N. Approval of final manuscript: all authors.

ORCID

Jeongeun Hwang: <http://orcid.org/0000-0002-5821-5405>

Hyunjung Park: <https://orcid.org/0000-0002-7177-9862>

Jiwon Jung: <http://orcid.org/0000-0003-4333-3270>

Sung-Han Kim: <http://orcid.org/0000-0002-6596-8253>

Namkug Kim: <http://orcid.org/0000-0002-3438-2217>

REFERENCES

1. KCDC. Korean Centers for Disease Control and Prevention <http://cdc.go.kr>2020.
2. Obadia T, Haneef R, Boelle PY. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *Bmc Medical Informatics and Decision Making* 2012;12.
3. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B-Biological Sciences* 2007;274(1609):599-604.
4. Svensson A. A note on generation times in epidemic models. *Mathematical Biosciences* 2007;208(1):300-11.
5. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020.
6. CDC. Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html>. Updated March 7, 20202020.
7. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology*

181 2004;160(6):509-16.

182 8. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented
183 infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science
184 2020.

185

Cumulated confirmed cases

Except Sincheonji

Sincheonji

Except Daegu-Gyeongbuk

Daegu-Gyeongbuk

Total

1

2

3

4

5