medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Title: Rapid transmission of coronavirus disease 2019 within a religious sect in South Korea: a mathematical modeling study

Authors: Jong-Hoon Kim[1*], Hyojung Lee[2,3+], Yong Sul Won[2+], Woo-Sik Son[2], and Justin Im[1]

[1]International Vaccine Institute, Seoul, South Korea

[2]National Institute for Mathematical Sciences, Daejeon, South Korea

[3]Department of Statistics, Kyungpook National University, Daegu 41566, South Korea

[+]Equal contribution

[*]Corresponding author:

Jong-Hoon Kim jonghoon.kim@ivi.int; kimfinale@gmail.com

Abstract: Rapid transmission of coronavirus disease 2019 (COVID-19) was observed in the Shincheonji Church of Jesus, a religious sect in South Korea. The index case was confirmed on February 18, 2020 in Daegu City, and within two weeks, 3,081 connected cases were identified. Doubling times during the initial stages of the outbreak were less than 2 days. A stochastic model fitted to the time series of confirmed cases suggests that the basic reproduction number (R_0) of COVID-19 was 8.5 [95% credible interval (CrI): 6.3, 10.9] among the church members, whereas (R_0 = 1.9 [95% CrI: 0.4, 4.4]) in the rest of the population of Daegu City. The model also suggests that there were 4 [95% CrI: 2, 11] undetected cases when the first case reported symptoms on February 7. The Shincheonji Church cluster is likely to be emblematic of other outbreak-prone populations where R_0 of COVID-19 is higher. Understanding and subsequently limiting the risk of transmission in such high-risk places is key to effective control.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

1. Introduction

Coronavirus disease 2019 (COVID-19) has become a global pandemic since it was first reported in Wuhan, China in December 2019 with the name of novel coronavirus disease [1]. The causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), transmits mainly through human-to-human contact [2], which can happen even during the infector is asymptomatic [3, 4]. Infection with the virus causes diseases with varying degree of symptoms including death [5, 6]. Infection mortality ratio is lowest among children aged between 5 and 9 years and increases loglinearly with age [7].

One key characteristic of COVID-19 pandemic is that transmission events in high-risk settings such as super-spreading events (SSEs) contribute to most transmissions [8, 9]. The risk of COVID-19 transmission is believed to high in places with high occupancy and poor ventilation [10]. One extreme example is the outbreak in the Diamond Princess cruise ship, where 17% (619/3711) of the passengers were infected from January 25 to February 20, 2020 [11]. Other examples include transmission events in bars and wedding [8] in Hong Kong, nursing homes in U.S. [12], telemarketers working in group in closed places [13] and fitness classes [14] in South Korea, and also religious gatherings, which we describe below.

Explosive spread of COVID-19 was observed in the Shincheonji Church of Jesus ("Shincheonji"), a religious sect in South Korea. The index case was confirmed on February 18, 2020 and within two weeks, 3,081 connected cases were identified [15]. A simple calculation reveals that the outbreak size doubled in less than every 2 days ($14/\log_2(3081) \approx 1.21$), which is smaller than doubling times reported in the early stages of the COVID-19 outbreak in China (2.5 and 3.1 days in Hubei Province and Hunan Province, respectively) [16], Spain (2.8 days) [17], the US (2.7 days) [18], and Korea (2.8 – 10.2 days) [19]. A total of 6,684 confirmed cases were reported in Daegu City as of March 31, 2020 of which 4,467 (66.8%) were Shincheonji members, representing close to half (47.9%, 4,467/9,334) of the city's total Shincheonji membership.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Previous studies highlighted that COVID-19 transmissions involve SSEs [20], which can play a key role in sustained community transmissions [8, 9]. However, there have been no attempts to model the dynamics of COVID-19 transmission within high-risk settings and their interaction with the general community. In this study, we modeled the outbreak in the Shincheonji community while accounting for its interaction with the rest of population. We used the stochastic model to account for the stochastic nature of the transmission events. We used the model to explore the differences in the basic reproduction number (R_0) between the high-risk setting and the general community, and quantify uncertainties related to the initial conditions and dynamics of transmission under the dynamic intervention programs.

2. Materials and Methods

2.1 Backgound on the Shincheonji Church of Jesus

Shincheonji was founded by Man-hee Lee in 1984 and has approximately 245,000 members including 30,000 foreigners [21]. At Shincheonji gatherings, worshipers used to sit close together on the floor and facial coverings, such as glasses and face masks, are forbidden. Members were expected to attend services despite illness [22]. The index case of the Daegu City outbreak was identified as a Shincheonji member and some 1,000 people were reported to have attended worship together [23]. Further tracing of church members identified clustering in apartment complexes. Of 142 residents in a particular Daegu apartment block, 94 (66%) were Shincheonji members of whom 46 (38.9%) tested positive for the virus [24].

2.2 Data

Time series of patients confirmed with COVID-19 in Shincheonji community and the overall Daegu City over the period of 19 February – 31 March 2020 was compiled based on the daily reports from Korea Disease Control and Prevention Agency (KDCA) [15] (Figure 1). The reports provide the number of cases confirmed with SARS-CoV 2 based on reverse transcription polymerase chain reaction (RT-PCR) by category (Shincheonji or non-Shincheonji). We made some adjustments to the

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

existing data before we fit the model. First, in the beginning of the outbreak, KDCA provided both daily and cumulative numbers of cases confirmed for SARS-CoV 2, which did not agree always. If there is a discrepancy between these numbers, we prioritized cumulative numbers as this figure was reported continuously throughout the outbreak. Second, data were missing for some days for the number cases for Shincheonji members. We imputed missing values using the cubic spline method (Figure S1 in the Supplementary Material).

2.3. Doubling time

The epidemic doubling time (T_d) represents the duration in which the cumulative incidence doubles. Assuming exponential growth with a constant epidemic growth rate (r), the epidemic doubling time can be calculated by the following equation [18, 25, 26]

$$T_d = \frac{\ln(2)}{r}. \quad (1)$$

Epidemic growth rate (r) may be estimated based on the data. For example, $r(t)$ can be estimated by the following equation:

$$r(t) = \frac{\ln\big(C(t)\big) - \ln\big(C(t-\Delta t)\big)}{\Delta t}, \quad (2)$$

where $C(t)$ indicates the cumulative number of infected people at time t and Δt is the duration over which $r(t)$ is assumed to be constant. $r(t)$ can be calculated over the fixed time interval (e.g., 1 day or 1 week) [27, 28] or variable time intervals (e.g., days on which the number of cases doubles, quadruples, etc.) [19, 25]. We calculated doubling based on prior 7 days or 1 day from 18 February

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

to 5 March 2020, when the epidemic peaked and no further doubling of cumulative number of cases occurred onward.

The basic reproduction number (R_0) is defined as the average number of secondary cases caused by a single infected case in an entirely susceptible population and it provides sufficient information to produce doubling times in the beginning of an outbreak. However, estimating R_0 requires additional information such as generation time or developing a mechanistic model, and its estimates come with higher degree of uncertainty [26]. Calculating doubling times requires fewer assumptions and also allows us to compare our results with estimates from different settings where doubling times, but not reproduction numbers, are available.

2.4. Mechanistic model of COVID-19 transmission

We developed a stochastic model of COVID-19 transmission within the Shincheonji community and the overall population of Daegu City. The model includes six disease states: susceptible (S), exposed but not infectious (E), pre-symptomatic but infectious (P), symptomatic and infectious (I), asymptomatic but infectious (A), confirmed and isolated (C), and recovered (R). The model includes two patches to model Shincheonji and non-Shincheonji people, separately. Transmission rates may differ for each patch and person from one patch may infect people from the other patch. (Figure 2).

This modeling framework of mixing between two distinct sub-populations has been adopted in previous works, ranging from sexually transmitted diseases [29] to vector-borne diseases such as dengue [30], where formulations for mixing between patches vary. We adopted the formulation used

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

in the work on modeling transmission of cholera between hotspot and non-hotspot areas [31]. Mixing between two sub-populations are defined by the 2 × 2 contact matrix,

$$C = \begin{pmatrix} c_{12} & c_{12} \\ c_{21} & c_{22} \end{pmatrix},$$

where c_{ij} indicates the fraction of time that individuals from patch i spends in patch j. Next, we impose two conditions on the matrix C:

(i) Individuals must reside in either of the two patches, i.e., $c_{i1} + c_{i2} = 1$ for i =1 (Shincheonji) and 2 (non-Shincheonji).

(ii) The population in each patch remains constant, i.e., $c_{12}N_1 = c_{21}N_2$, where N_1 and N_2 represent population size for patch 1 and 2, respectively.

The above conditions may transform the contact matrix C to the following form:

$$C = \begin{pmatrix} 1 - c_{12} & c_{12} \\ \dfrac{c_{12}N_1}{N_2} & 1 - \dfrac{c_{12}N_1}{N_2} \end{pmatrix}$$

containing only one unknown parameter c_{12}.

The force of infection for individuals from patch i at time t, $\lambda_i(t)$, is defined as follows:

$$\lambda_i(t) = \sum_j c_{ij}\,\beta_j \frac{\sum_k c_{kj}\,I_k(t)}{\sum_k c_{kj}\,N_k(t)}, \qquad (3)$$

where β_j indicates local transmission rate in patch j and $I_k(t)$ indicates number of infectious individuals from patch k.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

138 The transitions between states are modeled using an explicit tau-leap algorithm [32] to account for
139 stochasticity of the infection transmission process. The number of susceptible people in patch i at
140 time $t + \Delta t$, $S_i(t + \Delta t)$, is written as follows:
141

$$S_i(t + \Delta t) = S_i(t) - Q_i^{SE}(t, t + \Delta t). \quad (4)$$

142 $Q_i^{SE}(t, t + \Delta t)$ represents the number of people who transit from state S to state E from t to $t + \Delta t$
143 in patch i and is a random variable with binomial distribution:

$$\mathrm{Bin}\big(S_i(t), \Delta t \lambda_i(t)\big). \quad (5)$$

144 That is, it is represented as an integer varying between 0 and $S_i(t)$. For states from which more than
145 one potential transition exist (e.g., P to either A or I), multinomial distributions were applied. For
146 instance, the number of people transit from P to either I or A are given as follows:

$$\mathrm{Multi}(P_i(t), \Delta t \pi), \quad (6)$$

147 where π is a vector given as
148

$$\left(\frac{1-f}{1/\delta - 1/\epsilon}, \frac{f}{1/\delta - 1/\epsilon}\right). \quad (7)$$

149

150 The first element of π indicates a probability of transition from P to I and the second element
151 indicates the probability of transition from P to A. The number of people in other states (i.e.,
152 E, A, I, C, R) at time t can be described similarly. The model was implemented in a combination of R

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

and C++ languages, in which the core transmission model part is expensive and was written in C++.

All the computer codes that generate the results in this paper are available on GitHub [33].

2.5. Modeling intervention program

To account for intensification of the intervention such as case isolation and contact tracing with subsequent testing during the outbreak, we assumed case isolation rate (1 / mean time between symptom onset and case isolation) and transmission rate of the infectious people per unit time change over time. Specifically, we assumed that the case isolation rate, $\alpha(t)$, starts increasing on February 20 from the initial value of α^{init} when 4,474 out of 9,334 Shincheonji members were identified and were asked to self-isolate. During model fitting, we let data suggest the duration of intervention, d in day, which is the time required for the case isolation rate to reach its minimum, $\alpha(t) = \alpha^{\text{final}}$ for $t >$ February 20 + d. We assumed that the mean time between symptom onset and case isolation linearly decreases over the intervention period d. In other words, $\alpha(t)$ is formulated as follows:

$$\alpha(t) = \begin{cases} 0, & \text{if } t < \text{Feb } 17 \\ \alpha^{\text{init}}, & \text{if Feb } 17 \le t < \text{Feb } 20 \\ \alpha^{\text{init}} + (t - \text{Feb20})(\alpha^{\text{init}} - \alpha^{\text{final}})/d\,, & \text{if Feb } 20 \le t < \text{Feb } 20 + d \\ \alpha^{\text{final}}, & \text{if Feb } 20 + d \le t \end{cases} \quad (8)$$

where α^{final} is assumed to be 1 day based on the experiences in Busan City in Korea and $\alpha(t)$ is assumed to be zero before February 17 when the index case was detected.

Similarly, transmission rate per unit time at time t, $\beta_i(t)$ for $i = 1$ (Shincheonji members), 2 (non-Shincheonji people in Daegu City), is assumed to linearly decrease during the intervention period.

$$\beta_i(t) = \begin{cases} \beta_i^{\text{init}}, & \text{if } t < \text{Feb } 20 \\ \beta_i^{\text{init}} - (t - \text{Feb20})\left(\beta_i^{\text{init}} - \beta_i^{\text{final}}\right)/d\,, & \text{if Feb } 20 \le t < \text{Feb } 20 + d \\ \beta_i^{\text{final}}, & \text{if Feb } 20 + d \le t \end{cases} \quad (9)$$

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

174 Here, β_i^{init} and β_i^{final} indicate the transmission rate per unit time before the intervention and after the
175 intervention measures fully take effect, respectively. They can be derived once $R_{0,i}$ and R^{final} are
176 given as:

$$\beta_i^{\text{init}} = \frac{R_{0,i}}{\frac{1}{\delta} - \frac{1}{\epsilon} + \frac{f}{\gamma + \alpha\rho} + \frac{1-f}{\gamma + \alpha}} \tag{10}$$

and

$$\beta_i^{\text{final}} = \frac{R^{\text{final}}}{\frac{1}{\delta} - \frac{1}{\epsilon} + \frac{f}{\gamma + \alpha\rho} + \frac{1-f}{\gamma + \alpha}}. \tag{11}$$

177 *2.6. Parameter estimation*

178 Our model of COVID-19 transmission requires 15 parameters (Table 1). We divided the model
179 parameters into three classes depending on our belief on their relative certainty. The first class
180 includes parameters related to the natural history of infection and population size and we deemed that
181 available parameter estimates are reliable. For these parameters, we used their point estimates based
182 on analyses of data on COVID-19 transmissions in Korea or China. For the second class, which
183 includes parameters related to intervention programs, we used our best guesses based on supporting
184 evidence but still acknowledged their uncertainty. Therefore, we analyzed the models under various
185 assumptions on their values within some pre-specified ranges. Finally, we defined six parameters that
186 are critical for characterizing dynamics of COVID-19 transmission in Shincheonji members and non-
187 Shincheonji people. We estimated these parameters by fitting the model to daily confirmed COVID-
188 19 cases of Shincheonji members and non-Shincheonji people.

189

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Estimation of parameters $\theta = (R_{0,1}, R_{0,2}, I_0, c_{12}, d, R^{\text{final}})$ was based on Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) [34]. The ABC is a method for approximating posterior distributions given data D, $p(\theta|D)$, by accepting proposed parameter values when the difference between simulated data D^* and D, $d(D, D^*)$, is smaller than tolerance ϵ:

$$p(\theta|D) \approx p(\theta|d(D, D^*) \leq \epsilon).$$

For our model, $d(D, D^*)$ is defined as the sum of the squared differences in daily confirmed cases over the outbreak of duration T days, that is,

$$d_i(D, D^*) = \sum_{t=1}^{T} (D_t - D_t^*)^2,$$

for Shincheonji ($i = 1$) and non-Shincheonji ($i = 2$). Here, D_t and D_t^* represent observed daily confirmed cases and model predicted values at time day t, respectively. ABC-SMC was designed to increase efficiency of the ABC method and ABC is applied in a sequential manner by constructing intermediate distributions, which converge to the posterior distribution. Tolerance ϵ is gradually decreased and each intermediate distribution is obtained as a sample that is drawn with weights from the previous distribution and then perturbed through a kernel $K(\theta|\theta^*)$. The kernel helps keep the algorithm from being stuck in local optimum while maintaining the efficiency of the ABC-SMC method. Minimally informative uniform distributions were used as prior distributions and estimation procedure was repeated for ten different random seeds. The resulting distribution was summarized as median, 50% credible intervals (CrI; interval between 25% and 75% percentiles) and 95% CrI (interval between 2.5% and 97.5% percentiles). More details of the algorithm such as prior distribution for each parameter, the number of steps, the tolerance values for each step, perturbation kernel appear in the Supplementary Material.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

3. Results

3.1 Doubling time

Over the period of February 18 – March 5, during which doubling of confirmed cases occurred 12 times, doubling times were <1 day in the beginning and increased subsequently with daily doubling time presenting higher variability for both Shincheonji and non-Shincheonji values (Table 2). Doubling times calculated over sliding one-week intervals remained shorter than 3 days for the most part for both Sincheonji and non-Shincheonji population.

3.2 Comparison between observations and the mechanistic model

Our fitted model projects the trajectory of number of daily and cumulative confirmed cases in Shincheonji and in the rest of the population of Daegu City (Figure 3(a)-(d)). The model correctly projects the decreasing trends in both patches after reaching the peak on around March 3, 2020. However, for the non-Shincheonji, daily new cases are underestimated toward the end of the outbreak. R_0 were estimated to be quite different across two patches (Figure 3(e)). The local reproduction number in the patch representing Shincheonji members, $R_{0,1}$, was estimated to be 8.54 [95% credible interval (CrI): 6.30, 10.95] whereas the local reproduction number in the patch representing non-Shincheonji members, $R_{0,2}$, was estimated to be 1.87 [95% CrI: 0.38, 4.40]. The time taken for the intervention program to have exerted highest effect, d, is around 9.02 days [95% CrI: 7.85, 10.45], which leads to both reduced transmission rate per unit time and reproduction numbers (R^{final} = 0.34 [95% CrI: 0.18, 0.53]). The model also suggests that there were infectious people already when the first cases was symptomatic on February 7 (I_0= 4 [95% CrI: 2, 11]). The proportion of time that a person from Shincheonji members spends mixing with non-Shincheonji people, c_{12}, was estimate be around 0.14 [95% CrI: 0.05, 0.22]. Posterior distribution of parameters based on 2,000 samples obtained from 10 different random seeds and two-way correlations appear in Supplementary Material (Figure S2).

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

4. Discussion

Rapid transmission of COVID-19 within the Shincheonji community is likely to have been facilitated by high intensity contact between individuals gathering during services and in residential areas. Our mathematical modeling analyses quantify the rapid spread of COVID-19 in Daegu City driven by a community of Shincheonji members. The median R_0 among Shincheonji members ($R_{0,1}$) was 8.5, which is over 4-fold higher than what was estimated for the rest of the population in Daegu City ($R_{0,2}$ =1.9). While the R_0 in the Shincheonji community is higher than estimates from most transmission hotspots (e.g., in China [35-39] and Korea [40-43]), such high $R_0$is not unusual in particular considering that R_0 can be different depending on the local settings with varying contact rates [44]. Studies do report that R_0 estimates of COVID-19 that are comparable or even higher than our estimates for the Shincheonji community. During periods of intensive social contacts near the Chinese New Year in China, R_0 was estimated to be 6 [45, 46]. Also, R_0 estimates were around 5 among those traveled from Wuhan and were subsequently confirmed in other countries [47], and around 7 during the initial growth phase in the UK [48]. In an extreme setting such as the Diamond Princess ship, much higher estimates (R_0 = 14.8) were reported [49]. Although the previous studies that included data on the outbreak in Shincheonji community report smaller R_0 estimates [40, 42] than our estimates, the difference might stem from that prior studies did not model the Shincheonji community separately from the rest of the population and therefore measured the R_0 averaged across sub-population that are highly heterogeneous.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Although estimated daily doubling times show some variability (e.g., 14 days on February 28 and 69.6 days on March 1 for the non-Shincheonji population), they are short overall, which indicates rapid growth of the outbreak, and are compatible with estimates from other settings. Daily doubling times were lower than one day in the beginning of the outbreak and this is similar to the estimates from several regions in China [25]. The study by Shim *et al*. [19] used the dataset from Daegu City, Korea including Shincheonji population produced the doubling time of 2.8 days [95% CI: 2.5, 4.0]. Our daily doubling time estimates averaged over the period of February 18 – March 5 is 2.9 days and is consistent with the study. The period from February 18 to March 5 is likely to have been used in the study by Shim *et al*. because the authors calculated the doubling times on the days when the reported cases doubled and during the period of February 18 – March 5 the number of cases doubled 12 times and no further doubling occurred since then. The study by Lee *et al*. [50] used similar data, but reported seemingly inconsistent findings, doubling time of 2.9 days for the first week and 3.4 over the period around February 18 – March 4 considering that our estimate averaged over the first week is 0.9 day. One likely reason for this difference is that Lee *et al*. calculated the doubling time using the cumulative incidence estimated from a logistic model that used the initial value (i.e., number of infected people on February 18) as a free parameter. Figure 2D from their study indicates that the number of infected people on 18 February is much larger than 1 and this might have led to the higher doubling time than our estimates. This may also explain why Lee *et al*. estimates for a similar period (i.e., February 18 – March 5) is higher than our estimates and those by Shim *et al*. [19].

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

The relationship between the doubling time and the R_0 provides two insights on our inferences on R_0. For an $SEIR$ model, there exists an algebraic formula that describes the inverse relationship between initial epidemic growth rate and R_0 [51, 52]. This inverse relationship suggests that short doubling times during the early phase of the epidemic we calculated using the growth rates are consistent with high R_0 for Shincheonji we estimated using the stochastic dynamic transmission model (Table S2 in the Supplementary Material). On the other hand, while doubling times may be reduced and imply high R_0 for non-Shincheonji people as well, such short doubling times can arise through mixing (i.e., positive c_{12}) with Shincheonji of high R_0 even if the R_0 for the non-Shincheonji people are not as high.

Parameters around asymptomatic infections of COVID-19 are largely unknown [53] and we [54]tested the sensitivity of our inferences to our assumptions on two parameters related to asymptomatic infection, namely the proportion of asymptomatic infection, f, and relative rate of isolation of asymptomatic people, ρ (Figure S3 in Supplementary Material). R_0 for Shincheonji people, $R_{0,1}$, and the final reproduction number, R^{final}, showed a slight increase with increasing f or decreasing ρ while other parameter estimates remain relatively constant. We also tested the sensitivity of our parameter estimates to α^{final}, maximum rate of isolation near the end of the outbreak. α^{final} showed an inverse relationship with other intervention-related parameters such as duration of intervention d, and reproduction number at the end of the outbreak, R^{final}. Overall, while there are some quantitative differences in our parameter estimates in response to the change in our assumptions on fixed parameters, $R_{0,1}$ was always over 4-fold higher than $R_{0,2}$.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

While the first case was confirmed on February 18 for the Shincheonji outbreak, it was later revealed that the first case had symptoms on February 7 and even earlier transmission events were also suspected [54, 55]. This finding is consistent with our model analyses, which suggest there were 4 [95% CrI: 2, 11] infectious people on February 7. These undetected cases are likely to have contributed to the explosive outbreak in the Shincheonji community. Studies suggest that a substantial fraction of all SARS-CoV-2 infections were undetected. For Korea, it was suggested that the number of undetected cases may be larger than the number of detected cases [56]. A study suggests that > 80% of all infections were undocumented during the initial spread in China [57]. In France, over the period of 7 weeks since 28 June 2020 after the first lockdown, it was estimated that around 93% of all symptomatic cases were undetected initially and later around 69% of symptomatic cases were undetected by the time when case ascertainment improved [58].

We have shown that SARS-CoV 2 has disproportionately affected a religious community generating a large cluster of linked cases in Korea. Similar large clusters of cases in high-risk settings have been observed in Korea and elsewhere. In Korea, many similar outbreaks in high-risk settings have been reported in the news including the outbreak in a dance class [14] and a call center [13]. In Singapore, a total of 247 cases were confirmed as of March 17, 2020 and six clusters including the spread in a hotel and in a church accounted for 45.3% of the total cases [59]. In Hong Kong. 1,038 cases were confirmed from January 23 to April 28, 2020 and among them, 51.3% of cases were associated with

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

large clusters. Such social settings as bars, restaurants, weddings, and religious sites appeared at increased risk of large outbreaks [8].

One limitation of our analyses is that the model was fit to date of case confirmation because the date of symptom onset, which is more closely related with the date of infection, was not available. The daily number of confirmed cases can abruptly change depending on the intensity of intervention measures, of which the dynamics may not be consistent with disease transmission process. This means using the data on case confirmation under dynamics intervention measures is challenging. We tried to mitigate this difficulty by incorporating the dynamics of intervention programs by assuming that the start date and duration of enhanced case detection vary while the case detection rate increases over time and let the data suggest the values for those parameters.

5. Conclusions

The potential for large variations in R_0 for COVID-19 has important implications for the design and effectiveness of control strategies. The efficacy of components of intervention programs, such as contact tracing and physical distancing, is dependent on various environmental and societal factors (e.g., large gatherings, physical proximity, high risk behaviors such as singing, etc.) that influence the transmissibility of disease. Our analyses provide important insights that in order to minimize the risk of sudden outbreaks, efforts to identify and preempt high transmission scenarios will be key to controlling the spread of the COVID-19. Understanding and subsequently limiting the risk of

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

338 transmission in high-risk places such as the Shincheonji Church cluster in Korea is key to effective
339 control of COVID-19 transmission.

340

341 **Funding:** This research was partly supported by Government-wide R&D Fund project for infectious
342 disease research (GFID), Republic of Korea (grant number: HG18C0088) and National Institute for
343 Mathematical Sciences (NIMS) grant funded by the Korean Government (NIMS-B21910000).

344

345 **Acknowledgments:** All authors acknowledge discussions with the members of the Research and
346 Development on Integrated Surveillance System Development for Early Warning of Infectious
347 Diseases of Korea.
348

349

350

351

352

353

354

355

356

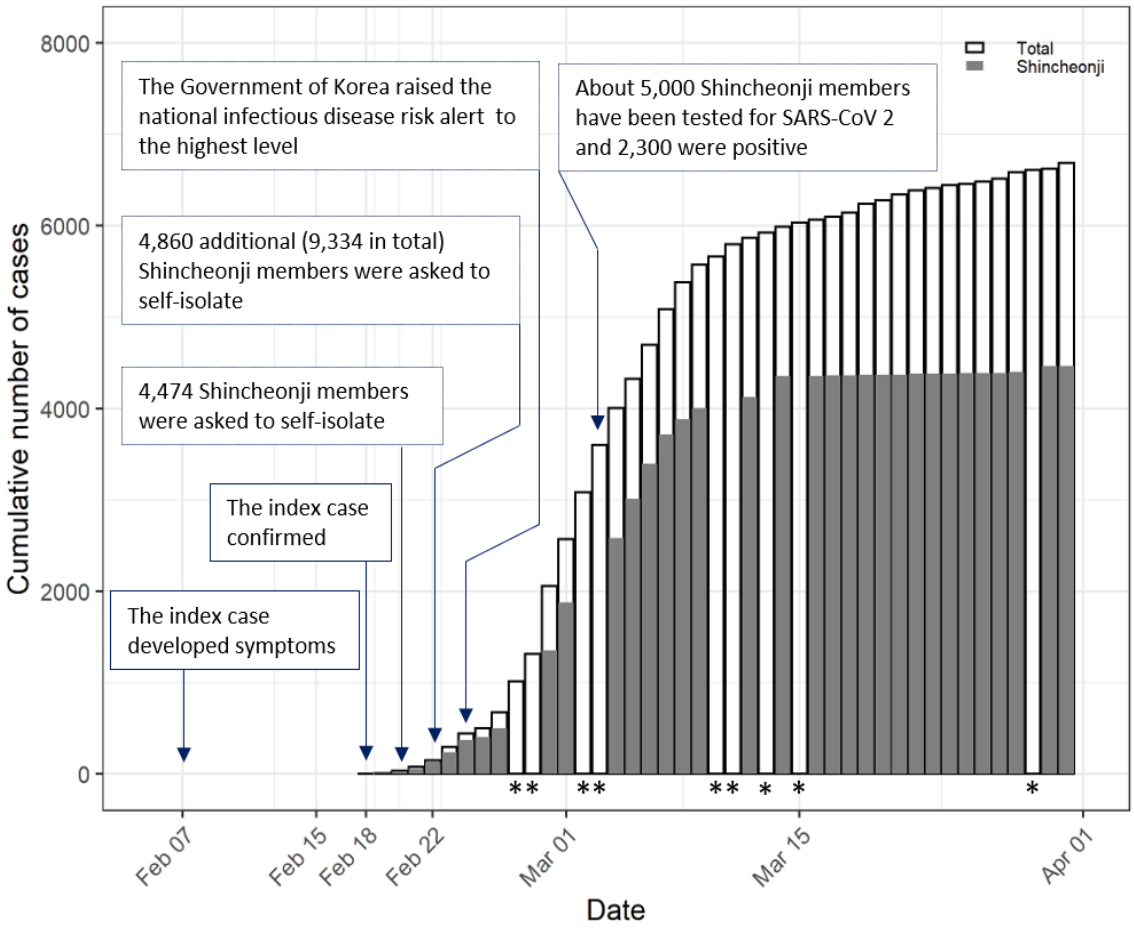
357

358

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

359 **Figure 1.** The daily number of confirmed COVID-19 cases from Daegu City, South Korea
360 between 18 February and 31 March 2020. White bars represent overall cases from Daegu City
361 including Shincheonji (Total) and grey bars represent cases associated with Shincheonji. Key
362 events during the outbreak are described in the boxes. *Data on Shincheonji-associated cases are
363 missing.

364



365

366

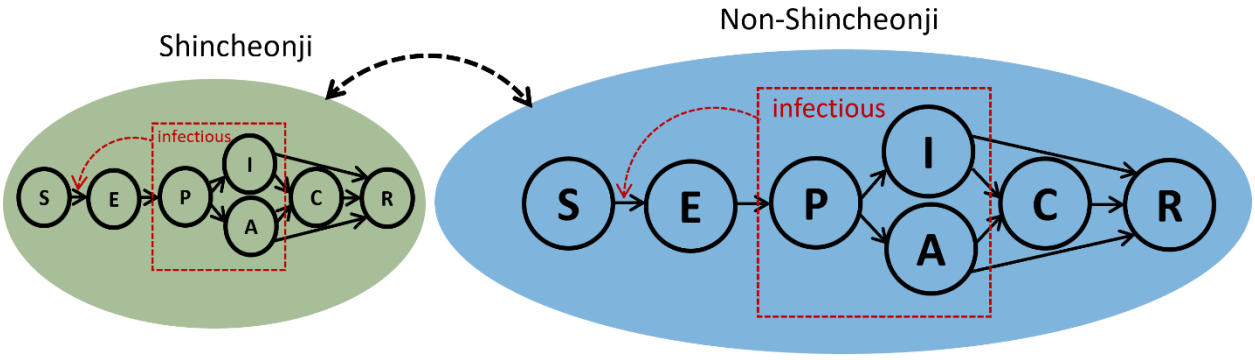
367

368

369

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

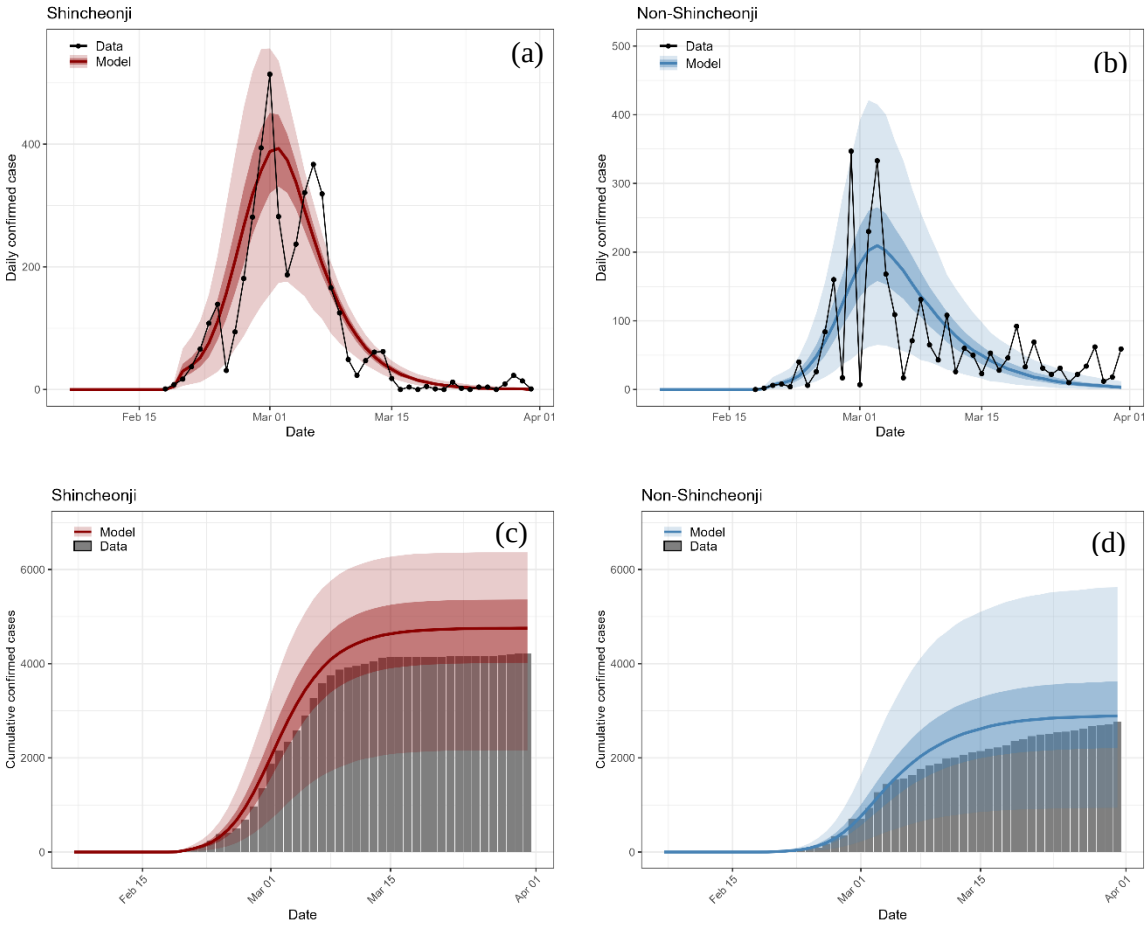
370 **Figure 2.** Schematic for the two-patch model. Circles with letters inside indicate compartments of
371 mutually exclusive states; susceptible (S), exposed but not infectious (E), pre-symptomatic but
372 infectious (P), symptomatic and infectious (I), asymptomatic but infectious (A), confirmed and
373 isolated (C), and recovered (R).
374



375
376
377
378
379
380
381
382
383
384
385
386
387

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

388 **Figure 3.** Parameter estimates and comparison between model outputs and data. Model outputs are
389 based on 2,000 simulation runs and presented with median (lines) with 50% (dark shading) and 95%
390 (light shading) percentiles. (a) and (b) represent daily number of confirmed COVID-19 cases from
391 the model and observations for Shincheonji and non-Shincheonji, respectively. (c) and (d) represent
392 cumulative confirmed cases from model outputs and data for Shincheonji and non-Shincheonji,
393 respectively. (e) shows posterior parameter estimates. Red dots represent the median, and thick and
394 thin bars represent 50% and 95% CrI of the posterior distribution for each parameter. Median for c_{12}
395 is 0.14 with its 50% CrI and 95% CrI being (0.11, 0.18) and (0.05, 0.22), respectively. Median for
396 R^{final} is 0.34 with its 50% CrI and 95% CrI being (0.28, 0.40) and (0.18, 0.53), respectively.
397



medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

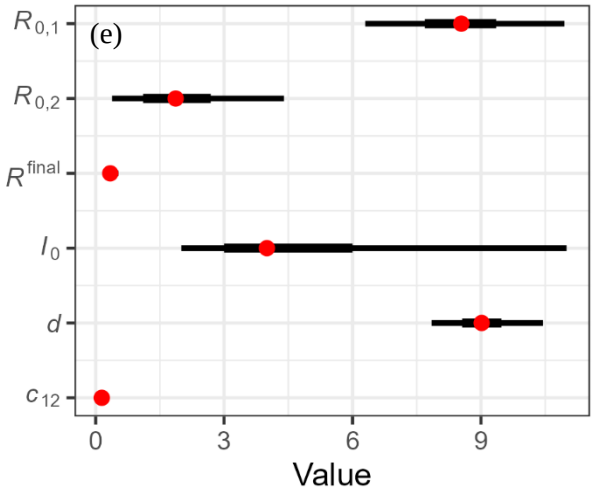


Table 1. Model parameters

Symbol	Parameter definition	Value	Source
	Parameters fixed throughout the simulation		
N_2	Population size of the Daegu city excluding Shincheonji members	~2.5 million	[60]
N_1	Population size of the Shincheonji members	9,334	[15]
$1/\delta$	Mean time between infection and onset of symptoms (i.e., mean incubation period)	5.2 days	[1]
$1/\epsilon$	Mean time between infection and infectiousness (i.e., mean latent period)	3 days	[61]
$1/\gamma$	Mean time between symptom onset and recovery	14 days	Assumption
τ	Time when the intense intervention measures begin to be implemented	February 20	[15]
$1/\alpha^{\text{init}}$	Mean time between symptom onset and isolation of a patient in the beginning of the outbreak (i.e., time $t < \tau$)	4.3 days	[41]
	Parameters for which a range of values explored		
ρ	Relate rate of isolation of asymptomatic cases to symptomatic cases. It is ρ before intervention measures, increases linearly and reaches 1 when the effect of intervention measures is highest	0.5 [0.1-0.9]	Assumption
f	Proportion of asymptomatic infections	0.306 [0.1 - 0.7]	[62]
$1/\alpha^{\text{final}}$	Mean time between symptom onset and isolation of a patient when the effect of intervention measures is highest	2 [1-3] days	Assumption[1]
	Estimated parameters		
c_{12}	Proportion of time that individuals from Shincheonji (N_1) spends in non-Shincheonji (N_2). We assume $c_{12}N_1 = c_{21}N_2$ to keep the population size constant for each patch. See the Method section for details.		
I_0	Number of infectious people in Shincheonji on February 7		

$R_{0,1}$	Basic reproduction number in the patch representing the Shincheonji members
$R_{0,2}$	Basic reproduction number in the patch representing the non-Shincheonji people in Daegu City
d	Time it takes for the intervention measures to be fully implemented
R^{final}	Reproduction number (Shincheonji members and non-Shincheonji population in Daegu City) after the intervention measures are implemented. This is used to reflect the reduced transmission rate through case isolation, quarantine, and social distancing

[1]The COVID-19 outbreak in Busan City, which occurred around the same time, indicate that delays from symptom onset to isolation were close to one day around the end of February when the outbreak started to flatten [63] (see data from Table S1 in the Supplementary Material).

Table 2. Doubling time during the Shincheonji outbreak during February 18 - March 5. On March 5, in total 4,443 cases (2,897 cases from Shincheonji and 1,546 cases from non-Shincheonji) were identified, which may indicate 12 times of doubling, and no subsequent doubling has happened.

Date	Cumulative incidence		Daily doubling time		Weekly rolling doubling time	
	Shincheonji	Non-Shincheonji	Shincheonji	Non-Shincheonji	Shincheonji	Non-Shincheonji
2020-02-18	1	0	-	-	-	-
2020-02-19	9	2	0.3	0.0	-	-
2020-02-20	26	8	0.7	0.5	-	-
2020-02-21	63	16	0.8	1.0	-	-
2020-02-22	129	20	1.0	3.1	-	-
2020-02-23	237	60	1.1	0.6	-	-
2020-02-24	376	66	1.5	7.3	-	-
2020-02-25	407	92	8.7	2.1	0.8	0.0
2020-02-26	501	176	3.3	1.1	1.2	1.1
2020-02-27	682*	335	2.2	1.1	1.5	1.3
2020-02-28	962*	352	2.0	14.0	1.8	1.6
2020-02-29	1356	699	2.0	1.0	2.1	1.4
2020-03-01	1870	706	2.2	69.6	2.3	2.0
2020-03-02	2152*	936	4.9	2.5	2.8	1.8
2020-03-03	2339*	1269	8.3	2.3	2.8	1.8
2020-03-04	2576	1437	7.2	5.6	3.0	2.3
2020-03-05	2897	1546	5.9	9.5	3.4	3.2

*Imputed numbers using a cubic spline method (see Methods and Supplementary Material)

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

1. Li, Q., et al., *Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.* N Engl J Med, 2020. **382**(13): p. 1199-1207.
2. Chan, J.F., et al., *A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster.* Lancet, 2020. **395**(10223): p. 514-523.
3. Rothe, C., et al., *Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany.* New England Journal of Medicine, 2020. **382**(10): p. 970-971.
4. Yu, P., et al., *A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period.* J Infect Dis, 2020. **221**(11): p. 1757-1761.
5. Guan, W.J., et al., *Clinical Characteristics of Coronavirus Disease 2019 in China.* N Engl J Med, 2020. **382**(18): p. 1708-1720.
6. Fu, L., et al., *Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis.* J Infect, 2020. **80**(6): p. 656-665.
7. O'Driscoll, M., et al., *Age-specific mortality and immunity patterns of SARS-CoV-2.* Nature, 2021. **590**(7844): p. 140-145.
8. Adam, D.C., et al., *Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong.* Nat Med, 2020. **26**(11): p. 1714-1719.
9. Lemieux, J.E., et al., *Phylogenetic analysis of SARS-CoV-2 in the Boston area highlights the role of recurrent importation and superspreading events.* medRxiv, 2020: p. 2020.08.23.20178236.
10. Jones, N.R., et al., *Two metres or one: what is the evidence for physical distancing in covid-19?* BMJ, 2020. **370**: p. m3223.
11. Russell, T.W., et al., *Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020.* Euro Surveill, 2020. **25**(12).
12. Chen, M.K., J.A. Chevalier, and E.F. Long, *Nursing home staff networks and COVID-19.* Proc Natl Acad Sci U S A, 2021. **118**(1).
13. Park, S.Y., et al., *Coronavirus Disease Outbreak in Call Center, South Korea.* Emerg Infect Dis, 2020. **26**(8): p. 1666-1670.
14. Jang, S., S.H. Han, and J.Y. Rhee, *Cluster of Coronavirus Disease Associated with Fitness Dance Classes, South Korea.* Emerg Infect Dis, 2020. **26**(8): p. 1917-1920.
15. Korea Disease Control and Prevention Agency. *Press release*. 2020; Available from: https://www.cdc.go.kr/board/board.es?mid=a20501000000&bid=0015.
16. Muniz-Rodriguez, K., et al., *Doubling Time of the COVID-19 Epidemic by Province, China.* Emerging infectious diseases, 2020. **26**(8): p. 1912-1914.
17. Guirao, A., *The Covid-19 outbreak in Spain. A simple dynamics model, some lessons, and a theoretical framework for control response.* Infectious Disease Modelling, 2020. **5**: p. 652-669.
18. Lurie, M.N., et al., *Coronavirus Disease 2019 Epidemic Doubling Time in the United States Before and During Stay-at-Home Restrictions.* The Journal of Infectious Diseases, 2020. **222**(10): p. 1601-1606.
19. Shim, E., A. Tariq, and G. Chowell, *Spatial variability in reproduction number and doubling time across two waves of the COVID-19 pandemic in South Korea, February to July, 2020.* Int J Infect Dis, 2020. **102**: p. 1-9.
20. Xu, X.K., et al., *Reconstruction of Transmission Pairs for Novel Coronavirus Disease 2019 (COVID-19) in Mainland China: Estimation of Superspreading Events, Serial Interval, and Hazard of Infection.* 2020(1537-6591 (Electronic)).
21. Chung, E. and A. Hill, *[DEBRIEFING] What is the Shincheonji Church of Jesus and who are its members? And more importantly, what are its links to the coronavirus?*, in *Korea Joongang Daily*. 2020.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

22. Choe, S.-H., *Shadowy Church Is at Center of Coronavirus Outbreak in South Korea*, in *The New York Times*. 2020.
23. Yonhap, *'Shincheonji' suspected as coronavirus hotbed*, in *The Korea Times*. 2020.
24. Myung, M.-j., *Apartment block enters cohort isolation for the first time*, in *THE DONG-A ILBO*. 2020.
25. Muniz-Rodriguez, K., et al., *Doubling Time of the COVID-19 Epidemic by Province, China.* Emerg Infect Dis, 2020. **26**(8): p. 1912.
26. Anderson, R.M., et al., *Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation.* The Royal Society, 2020.
27. Ebell, M.H. and G. Bagwell-Adams, *Mandatory Social Distancing Associated With Increased Doubling Time: An Example Using Hyperlocal Data.* American journal of preventive medicine, 2020. **59**(1): p. 140-142.
28. Patel, S.B. and P. Patel, *Doubling Time and its Interpretation for COVID 19 Cases.* Natl J Community Med, 2020. **11**(1): p. 141-3.
29. Koopman, J., et al., *Sexual partner selectiveness effects on homosexual HIV transmission dynamics.* J Acquir Immune Defic Syndr (1988), 1988. **1**(5): p. 486-504.
30. Lee, S. and C. Castillo-Chavez, *The role of residence times in two-patch dengue transmission dynamics and optimal strategies.* J Theor Biol, 2015. **374**: p. 152-64.
31. Azman, A.S. and J. Lessler, *Reactive vaccination in the presence of disease hotspots.* Proc Biol Sci, 2015. **282**(1798): p. 20141341.
32. Gillespie, D.T., *Approximate accelerated stochastic simulation of chemically reacting systems.* J Chem Phys, 2001. **115**(4): p. 1716-1733.
33. Model codes. Available from: https://github.com/kimfinale/COVID_Shincheonji.
34. Minter, A. and R. Retkute, *Approximate Bayesian Computation for infectious disease modelling.* Epidemics, 2019. **29**: p. 100368.
35. Alimohamadi, Y., M. Taghdir, and M. Sepandi, *The Estimate of the Basic Reproduction Number for Novel Coronavirus disease (COVID-19): A Systematic Review and Meta-Analysis.* Korean J Prev Med, 2020. **0**(0).
36. Wu, J.T., K. Leung, and G.M. Leung, *Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study.* Lancet, 2020.
37. Zhao, S., et al., *Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak.* Int J Infect Dis, 2020. **92**: p. 214-217.
38. Riou, J. and C.L. Althaus, *Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020.* Euro Surveill, 2020. **25**(4).
39. Imai, N., et al. *Report 3: Transmissibility of 2019-nCoV*. 2020.
40. Choi, S. and M. Ki, *Estimating the reproductive number and the outbreak size of COVID-19 in Korea.* Epidemiol Health, 2020. **42**: p. e2020011.
41. Ki, M., *Epidemiologic characteristics of early cases with 2019 novel coronavirus (2019-nCoV) disease in Korea.* 2020(2092-7193 (Electronic)).
42. Shim, E., et al., *Transmission potential and severity of COVID-19 in South Korea.* Int J Infect Dis, 2020. **93**: p. 339-344.
43. Bae, T.W., K.K. Kwon, and K.H. Kim, *Mass Infection Analysis of COVID-19 Using the SEIRD Model in Daegu-Gyeongbuk of Korea from April to May, 2020.* J Korean Med Sci, 2020. **35**(34): p. e317.

medRxiv preprint doi: https://doi.org/10.1101/2021.08.05.21261683; this version posted August 7, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

44. Temime, L., et al., *A Conceptual Discussion About the Basic Reproduction Number of Severe Acute Respiratory Syndrome Coronavirus 2 in Healthcare Settings.* Clinical Infectious Diseases, 2020.
45. Tang, B., et al., *Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions.* J Clin Med, 2020. **9**(2).
46. Sanche, S., et al., *The Novel Coronavirus, 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated.* medRxiv, 2020: p. 2020.02.07.20021154.
47. Zhao, Q., Y. Chen, and D.S. Small, *Analysis of the epidemic growth of the early 2019-nCoV outbreak using internationally confirmed cases.* medRxiv, 2020: p. 2020.02.06.20020941.
48. Dropkin, G., *COVID-19 UK Lockdown Forecasts and R0.* Frontiers in Public Health, 2020. **8**(256).
49. Rocklöv, J., H. Sjödin, and A. Wilder-Smith, *COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures.* J Travel Med, 2020. **27**(3).
50. Lee, W., et al., *COVID-19 in South Korea: epidemiological and spatiotemporal patterns of the spread and the role of aggressive diagnostic tests in the early phase.* International Journal of Epidemiology, 2020. **49**(4): p. 1106-1116.
51. Ma, J., *Estimating epidemic exponential growth rate and basic reproduction number.* Infect Dis Model, 2020. **5**: p. 129-141.
52. Ma, J., et al., *Estimating initial epidemic growth rates.* Bull Math Biol, 2014. **76**(1): p. 245-60.
53. Fox, S.J., et al., *The impact of asymptomatic COVID-19 infections on future pandemic waves.* medRxiv, 2020: p. 2020.06.22.20137489.
54. *Serial interval*. Available from: https://en.wikipedia.org/wiki/Serial_interval.
55. Korea Disease Control and Prevention Agency, *Press release (February 22).* 2020.
56. Lee, C., C. Apio, and T. Park, *Estimation of Undetected Asymptomatic COVID-19 Cases in South Korea Using a Probabilistic Model.* Int J Environ Res Public Health, 2021. **18**(9).
57. Li, R., et al., *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2).* Science, 2020. **368**(6490): p. 489-493.
58. Pullano, G., et al., *Underdetection of cases of COVID-19 in France threatens epidemic control.* Nature, 2021. **590**(7844): p. 134-139.
59. Tariq, A., et al., *Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020.* BMC Med, 2020. **18**(1): p. 166.
60. Korean Statistical Information Service. 2019; Available from: http://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=INH_1B04005N_22.
61. He, X., et al., *Temporal dynamics in viral shedding and transmissibility of COVID-19.* Nature Medicine, 2020. **26**(5): p. 672-675.
62. Kim, S.W., et al., *Clinical Characteristics and Outcomes of COVID-19 Cohort Patients in Daegu Metropolitan City Outbreak in 2020.* J Korean Med Sci, 2021. **36**(1): p. e12.
63. Son, H., et al., *Epidemiological characteristics of and containment measures for COVID-19 in Busan, Korea.* Epidemiol Health, 2020. **42**(0): p. e2020035-0.

Supplementary Material

Title: Rapid transmission of coronavirus disease 2019 within a religious sect in South Korea: a mathematical modeling study

Authors: Jong-Hoon Kim[1*], Hyojung Lee[2+], Yong Sul Won[2+], Woo-Sik Son[2], and Justin Im[1]

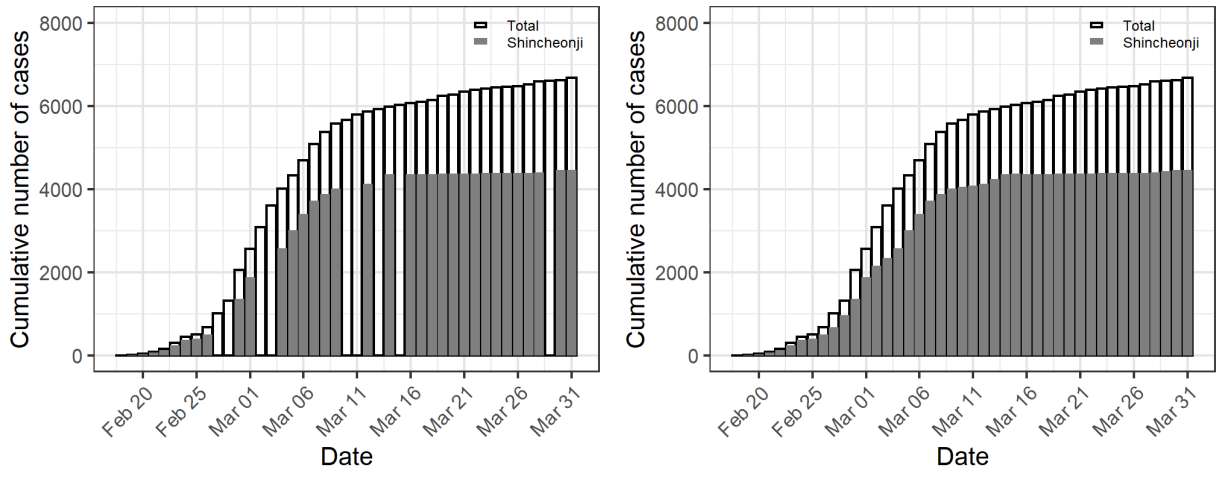
[1]International Vaccine Institute, Seoul, South Korea

[2]National Institute for Mathematical Sciences, Daejeon, South Korea

Data source

Figure S1 shows the cumulative cases before and after imputation. We used the cubic spline method provided in the imputeTS package of R.

Figure S1. Cumulative number of cases in Shincheonji community and the overall Daegu before (left panel) and after imputation (right panel).



Model equations

The model is implemented using a tau-leap algorithm. The number of people who transit from state *x* to state *y* over the time interval from *t* to *t* +Δ*t* in patch *i*, $Q_i^{xy}(t, t+\Delta t)$, is defined as follows:

$$Q_i^{SE}(t, t+\Delta t) = \text{Bin}(S_i(t), \lambda_i(t)\Delta t),$$

where Bin(*n*, *p*) represents binomial and with parameters *n* and *p*, respectively.

$$Q_i^{EP}(t, t+\Delta t) = \text{Bin}(E_i(t), \epsilon\Delta t),$$
$$Q_i^{Px}(t, t+\Delta t) = \text{Mult}(P_i(t), \pi),$$

where Multi(*n*, π ={ π_1, π_2}) indicates multinomial distribution with parameters *n* and π. π is given as follows:

$$\pi = \left\{\left(\frac{1}{1/\delta - 1/\epsilon}\right)(1-f), \left(\frac{1}{1/\delta - 1/\epsilon}\right)f\right\}. \quad (1)$$

The first element of π indicates a probability for *x* = *I* (i.e., transition from *P* to *I*) and the second element indicates the probability of transition from *P* to *A* (*x* = *A*).

$$Q_i^{Ix}(t, t+\Delta t) = \text{Mult}(I_i(t), \pi),$$
$$\pi = \{\Delta t\alpha(t), \Delta t\gamma\},$$

The first element of π indicates a probability for *x* = C (i.e., transition from *I* to *C*) and the second element indicates the probability of transition from *I* to *R* (*x* = *R*).

$$Q_i^{Ax}(t, t+\Delta t) = \text{Mult}(A_i(t), \pi),$$
$$\pi = \{\Delta t\alpha(t), \Delta t\gamma\}.$$

The first element of π indicates a probability for *x* = C (i.e., transition from *A* to *C*) and the second element indicates the probability of transition from *A* to *R* (*x* = *R*).

The number of people in each state at time *t* + Δ*t* can be described using the terms defined above:

$$S_i(t+\Delta t) = S_i(t) - Q_i^{SE}(t, t+\Delta t),$$
$$E_i(t+\Delta t) = E_i(t) + Q_i^{SE}(t, t+\Delta t) - Q_i^{EP}(t, t+\Delta t),$$
$$P_i(t+\Delta t) = P_i(t) + Q_i^{EP}(t, t+\Delta t) - Q_i^{PA}(t, t+\Delta t) - Q_i^{PI}(t, t+\Delta t),$$
$$A_i(t+\Delta t) = A_i(t) + Q_i^{PA}(t, t+\Delta t) - Q_i^{AC}(t, t+\Delta t) - Q_i^{AR}(t, t+\Delta t),$$
$$I_i(t+\Delta t) = I_i(t) + Q_i^{PI}(t, t+\Delta t) - Q_i^{IC}(t, t+\Delta t) - Q_i^{IR}(t, t+\Delta t),$$
$$C_i(t+\Delta t) = C_i(t) + Q_i^{IC}(t, t+\Delta t) + Q_i^{AC}(t, t+\Delta t) - Q_i^{CR}(t, t+\Delta t),$$
$$R_i(t+\Delta t) = R_i(t) + Q_i^{AR}(t, t+\Delta t) + Q_i^{IR}(t, t+\Delta t) + Q_i^{CR}(t, t+\Delta t).$$

The model comprises two sets of above equations that describe two patches (i.e., a community of Shincheonji members and the non-Shincheonji people in Daegu City) and these equations are linked through the force of infection function, $\lambda(t)$, which is defined in the main text.

Table S1. Delay from onset of symptoms to isolation during the COVID-19 outbreak in Busan City, Korea.

Date	Median	Mean	Number of cases
2/6/2020	17.0	17.0	1
2/16/2020	8.0	8.0	1
2/17/2020	5.5	5.5	2
2/19/2020	4.0	3.7	7
2/20/2020	2.5	2.6	8
2/21/2020	1.5	2.2	6
2/22/2020	1.0	2.0	9
2/23/2020	2.0	1.8	4
2/24/2020	0.0	0.7	3
2/25/2020	1.5	1.5	2
2/26/2020	1.5	1.8	4
2/27/2020	1.0	2.0	3
2/28/2020	3.0	2.3	3
2/29/2020	6.0	6.0	2
3/1/2020	2.0	2.0	1
3/2/2020	3.0	3.0	1
3/4/2020	3.0	3.0	1
3/6/2020	8.0	8.0	1
3/9/2020	3.0	3.0	2
3/12/2020	1.0	1.0	1
4/8/2020	10.0	10.0	1
5/6/2020	3.5	3.5	2
5/10/2020	2.0	2.0	1
5/27/2020	2.0	2.0	1

*Number of isolated cases

Model fitting

A pseudocode for Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) is presented below adopting what was presented in the previous study (Minter and Retkute, 2019):

1. Set the number of generations G and the number of particles N

2. Set the tolerance schedule $\epsilon_1 < \epsilon_2 < \epsilon_3 < \cdots < \epsilon_G$ and set the generation indicator $g = 1$
3. Set the particle indicator $i = 1$
4. If $g = 1$, sample θ^{**} from the prior distribution $p(\theta)$. If $g > 1$, sample θ^{*} from the previous generation $\{\theta_{g-1}\}$ with weights $\{w_{g-1}\}$ and perturb the particle to obtain $\theta^{**} \sim K(\theta|\theta^{*})$
5. If $p(\theta^{**}) = 0$, return to Step 4.
6. Generate n data sets D_j^{**} from the model using θ^{**} and calculate

   $\hat{p}(D|D^{**}) = \frac{1}{n}\sum_{j=1}^{n} 1(d(D, D_J^{**}) \leq \epsilon_g)$.
7. If $\hat{p}(D|D^{**}) = 0$, return to Step 4
8. Set $\theta_g^{(i)} = \theta^{**}$ and calculate the corresponding weight of the accepted particle i

$$w_g^{(i)} = \begin{cases} \hat{p}(D|D^{**})p(\theta^{**}), & \text{if } g = 1 \\ \dfrac{\hat{p}(D|D^{**})p(\theta^{**})}{\sum_{j=1}^{N} w_{g-1}^{(j)} K\left(\theta_g^{(i)} \middle| \theta_{g-1}^{(i)}\right)}, & \text{if } g > 1 \end{cases}$$

9. If $i < N$, increment $i = i + 1$ and go to step 4.
10. Normalize the weights so that $\sum_{i=1}^{N} w_g^{(i)} = 1$
11. If $g < G$, set $g = g + 1$, go to step 3

$K(\theta|\theta^{*})$ was assumed to follow a multivariate normal distribution that was truncated to give only positive values. Twenty generations (i.e., $G = 20$) were used with the following tolerance for Shincheonji ($\epsilon_{1..G}^{1}$) and non-Shincheonji ($\epsilon_{1..G}^{2}$):

Initial tolerance values ϵ_1^s were set as

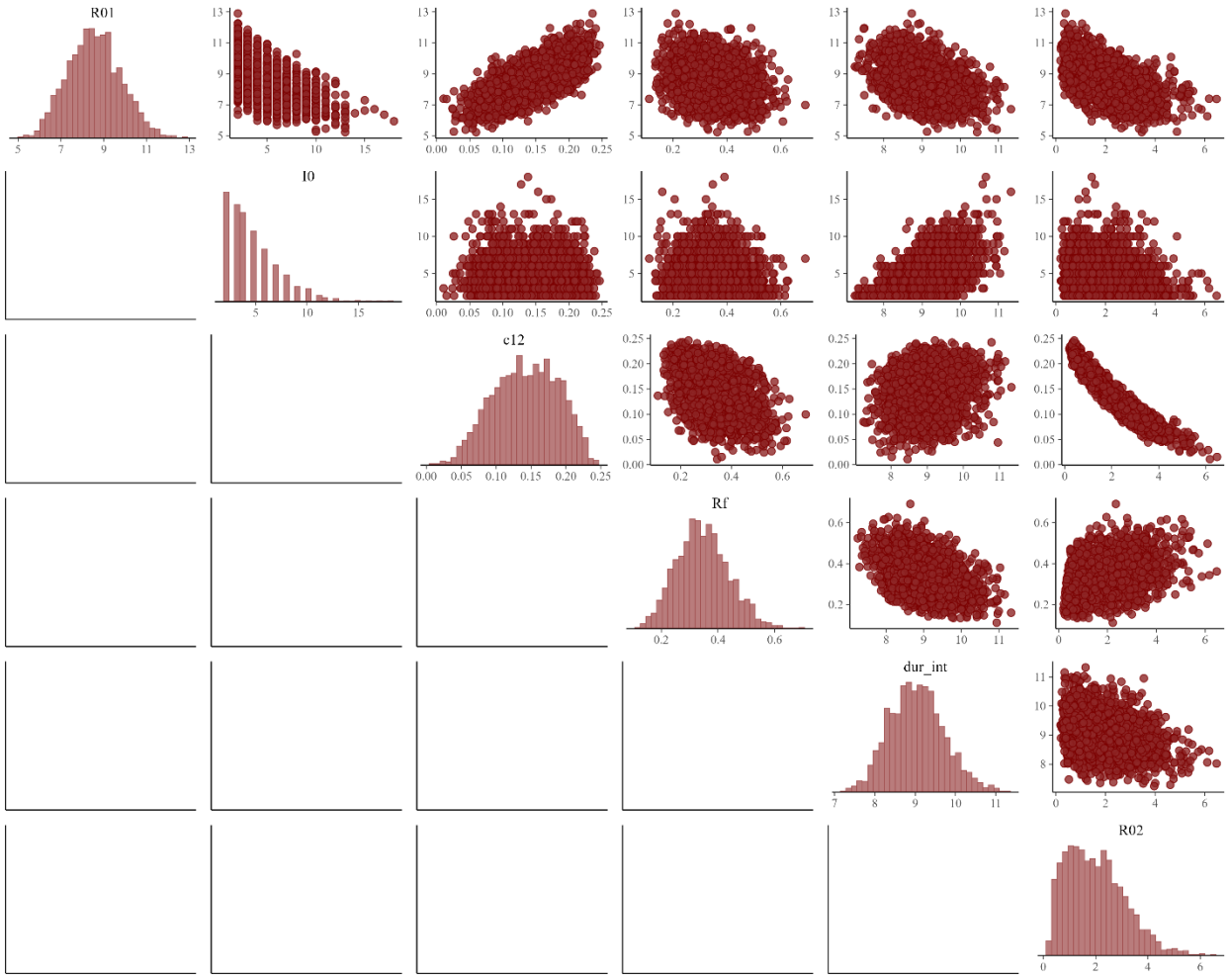
$$\sqrt{\sum_{j=1}^{n} \left(2 * y_{s,j}\right)^2},$$

where y_j represent incidence of confirmed case on day j for Shincheonji ($s = 1$) and non-Shincheonji ($s = 2$), respectively. $\epsilon_{2..20}^{s}$ values were determined by setting the minimum values, ϵ_{20}^{s}, as $0.06 \times \epsilon_1^s$ and dividing 20 equidistance pieces. Finally, final two value were manually adjusted to provide good fit between data and the model predictions through trial and error. Below are actual values used but were rounded for presentation.

$\epsilon_{1..G}^{1} = \{2140, 2035, 1928, 1823, 1717, 1611, 1505, 1399, 1293, 1187, 1081, 975, 869,$
$764, 658, 552, 446, 340, 290, 250\}$, $\epsilon_{1..G}^{2} =$
$\{1325, 1260, 1194, 1129, 1063, 998, 932, 867, 801, 736, 670, 604, 539, 473, 408, 342, 277, 211, 190, 180\}$.

Prior distributions for the parameters $\theta = \left(R_{0,1}, R_{0,2}I_0, c_{12}, d, R^{\text{final}}\right)$ were defined as uniform distribution as follows: $R_{0,1} \sim U(1,20), R_{0,2} \sim U(1,20), I_0 \sim U(1,20)$, $c_{12} \sim U(0.000001, 1)$, $d \sim U(1,30), R^{\text{final}} \sim U(1,20)$.

Figure S2. Posterior distribution of model parameters (n=2000). For each of 10 random seeds, 200 samples were generated.



Growth rate r and basic reproduction number R_0

For the differential equation-based SIR model, the initial (*i.e.*, the entire population is susceptible) epidemic growth rate r^* can be given as $\beta - \gamma$, (Ma, 2020) where β and γ represent transmission rate and recovery rate, respectively, as we defined in our model. Similarly, for a differential equation-based $SEIR$ model r^* is given as below (Ma, 2020; Ma et al., 2014):

$$r^* = \frac{1}{2}\left[-(\epsilon + \gamma) + \sqrt{(\epsilon - \gamma)^2 + 4\beta\epsilon}\right],$$

where ϵ represent the rate at which the exposed individuals become infectious (*i.e.*, $\frac{1}{\epsilon}$ = mean latent period) as we defined in the main text. The above equation gives β and therefore R_0 for given r^*, γ, ϵ. Assuming $r^* = r$, which is the growth rate we calculated in the main text, we can see to what value of R_0 the doubling times we calculated in the main text are translated and qualitatively see if R_0 estimates from the current study are reasonable.

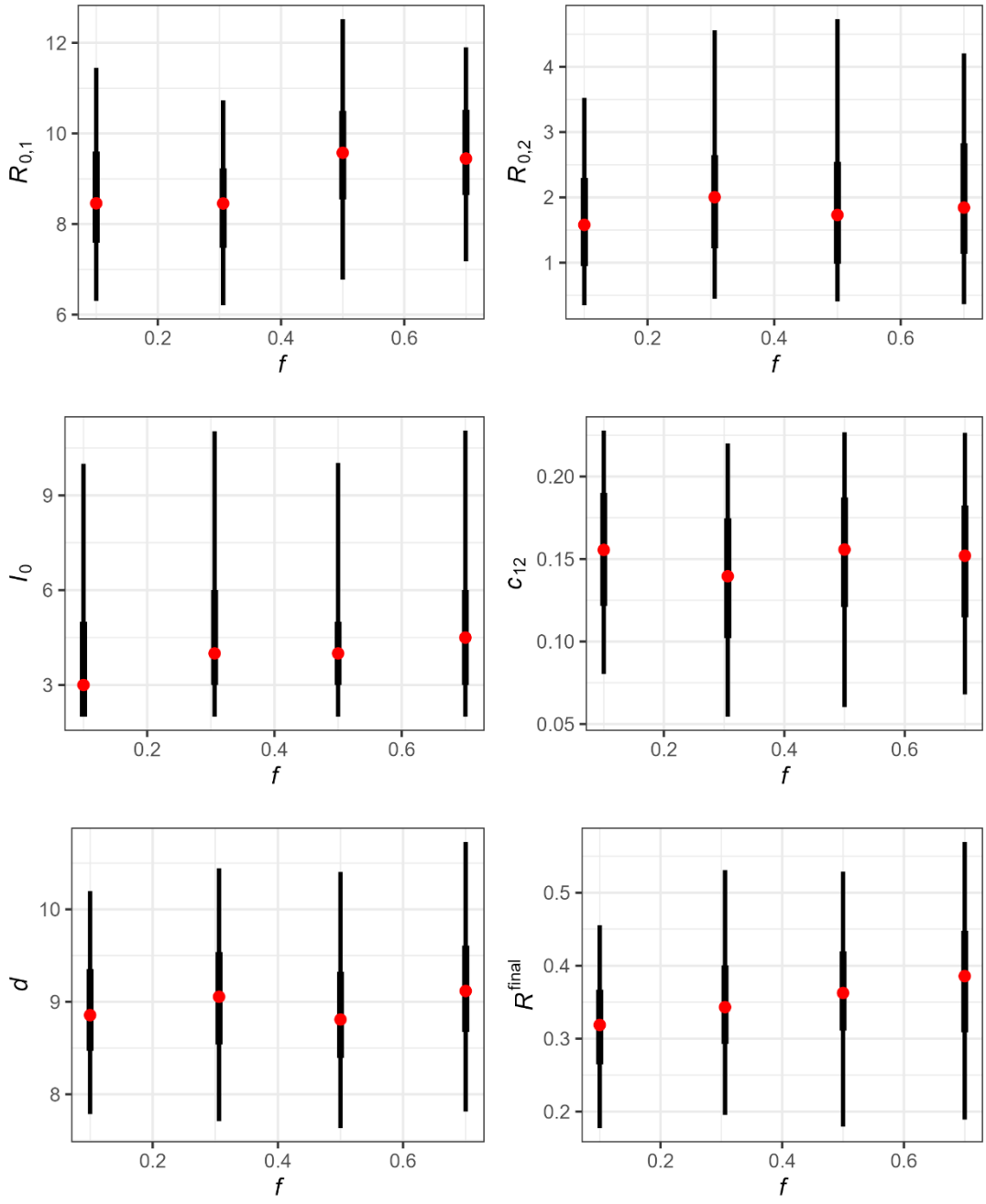
Table S2. Basic reproduction number, R_0, calculated by assuming the empirical daily or weekly growth rate r is the same as the r^* calculated for the SEIR model.

Date	*R_0 calculated from the daily doubling time		R_0 calculated from the weekly rolling doubling time	
	Shincheonji	Non-Shincheonji	Shincheonji	Non-Shincheonji
2020-02-18	-	-	-	-
2020-02-19	127.0	-	-	-
2020-02-20	29.5	51.6	-	-
2020-02-21	23.9	17.0	-	-
2020-02-22	17.0	4.1	-	-
2020-02-23	14.7	38.0	-	-
2020-02-24	9.6	2.1	-	-
2020-02-25	1.9	6.3	23.9	
2020-02-26	3.9	14.7	13.0	14.7
2020-02-27	5.9	14.7	9.6	11.6
2020-02-28	6.6	1.5	7.5	8.8
2020-02-29	6.6	17.0	6.3	10.5
2020-03-01	5.9	1.1	5.6	6.6
2020-03-02	2.7	5.1	4.5	7.5
2020-03-03	1.9	5.6	4.5	7.5

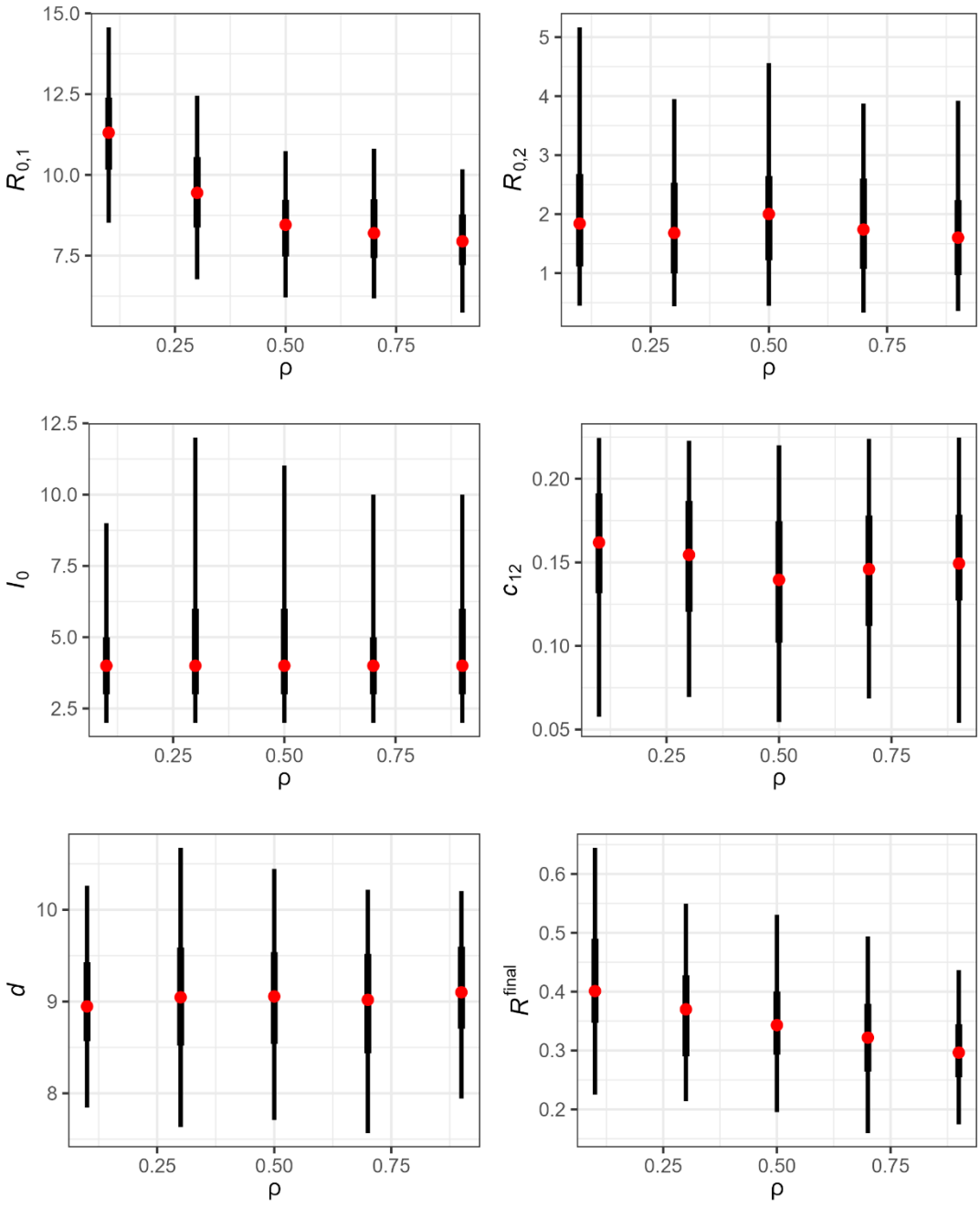
2020-03-04	2.1	2.5	4.2	5.6
2020-03-05	2.4	1.8	3.7	4.0

Figure S3. Sensitivity of our parameter estimates to simplifying assumption of three selected parameters

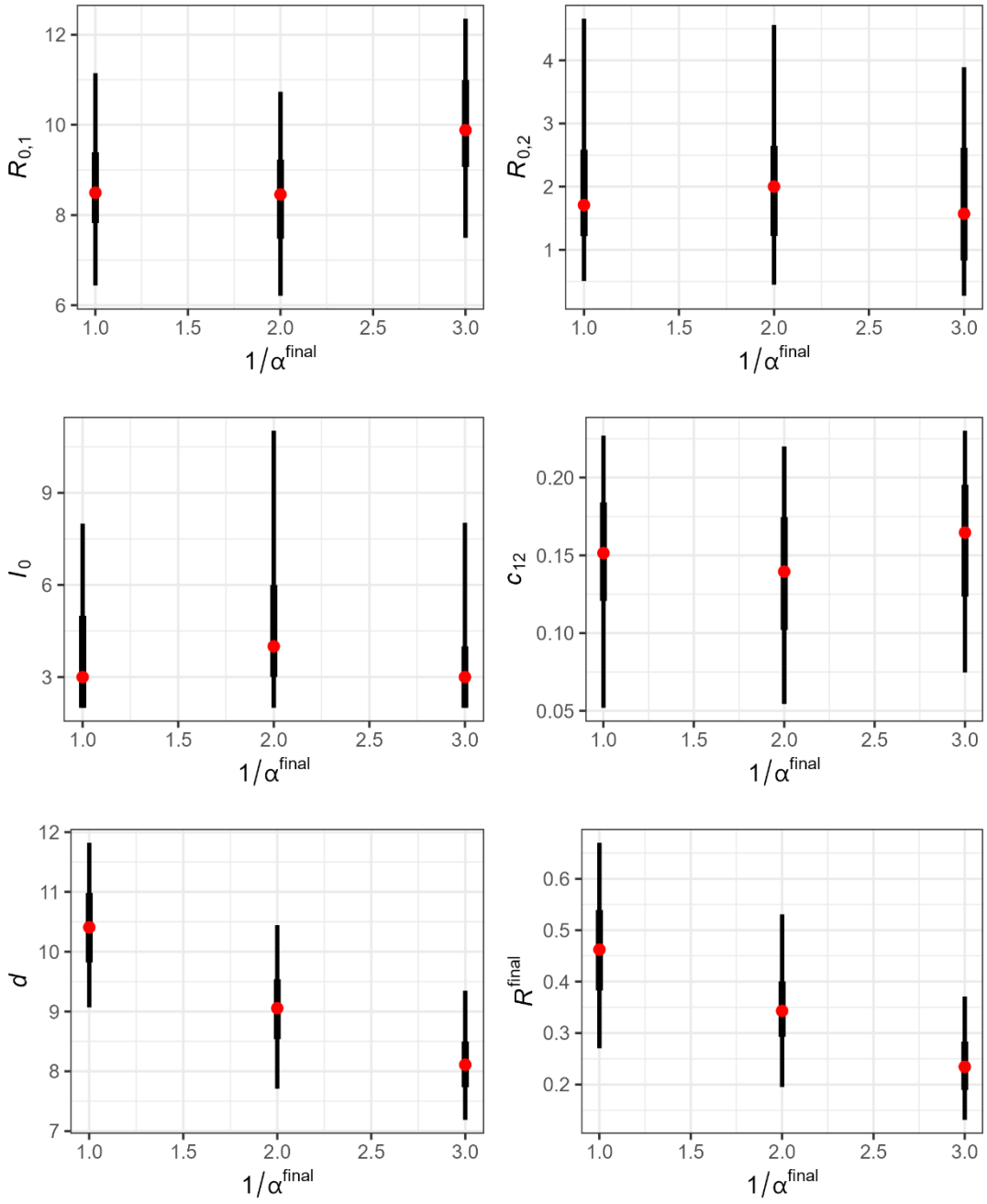
1) Fraction of asymptomatic infection f



2) Relative rate ρ of isolation of asymptomatic patients compared to the symptomatic patients before the intervention.



3) Time from symptom onset to isolation during the peak of the intervention, $1/\alpha^{\text{final}}$



References

Ma, J., 2020. Estimating epidemic exponential growth rate and basic reproduction number. Infect Dis Model 5, 129-141.

Ma, J., Dushoff, J., Bolker, B.M., Earn, D.J., 2014. Estimating initial epidemic growth rates. Bull Math Biol 76, 245-260.

Minter, A., Retkute, R., 2019. Approximate Bayesian Computation for infectious disease modelling. Epidemics 29, 100368.