



35 **Author Contributions**

36 EK and GL designed the study. GL, AB, MP, YY, AG, EH and EK developed the model, ran the  
37 computations, and analyzed the data. KT, OG, and ES conducted the literature search and data  
38 collection. GL, KT, and OG created all the figures. GL, AS, DM, SL, and EK interpreted the  
39 results. GL and EK drafted the Article. All authors edited and revised the Article.

## 40 **Abstract**

41 Understanding the transmission dynamics of the severe acute respiratory syndrome coronavirus 2  
42 (SARS-COV-2) is critical to inform sound policy decisions. We demonstrate how the  
43 transmission of undetected cases with pre-symptomatic, asymptomatic and mild symptoms,  
44 which are typically underreported due to lower testing capacity, explained the “bomb-like”  
45 behavior of exponential growth in the coronavirus disease 2019 (COVID-19) cases during early  
46 stages before the effects of large-scale, non-pharmaceutical interventions such as social  
47 distancing, school closures, or lockdowns. Using a Bayesian approach to epidemiological  
48 compartmental modeling, we captured the initial stages of the pandemic resulting in the  
49 explosion of cases and compared the parameter estimation with empirically measured values  
50 from the current knowledgebase. Parameter estimation was conducted using Markov chain  
51 Monte Carlo (MCMC) sampling methods with a Bayesian inference framework to estimate the  
52 proportion of undetected cases. Using data from the exponential phase of the pandemic prior to  
53 the implementation of interventions we estimated the basic reproductive number ( $R_0$ ) and  
54 symptomatic rates in Italy, Spain, South Korea, New York City, and Chicago. From this  
55 modeling study,  $R_0$  was estimated to be 3.25 (95% CrI, 1.09-29.77), 3.62 (95% CrI, 1.13-34.89),  
56 2.75 (95% CrI, 1.04-22.44), 3.31 (95% CrI, 1.69-20.55), and 3.46 (95% CrI, 1.01-34.41),  
57 respectively. For all locations, 3-25% of infected patients were identified with moderate and  
58 severe symptoms in the early stage of the COVID-19 pandemic. Our modeling results support  
59 the mounting evidence that potentially large fractions of the infected population were undetected  
60 with asymptomatic and mild symptoms. Furthermore, a significant number of models of  
61 transmission that do not account for these asymptomatic cases may lead to an underestimation of  
62  $R_0$  and, subsequently, policies that do not sufficiently reduce transmission to contain the spread

63 of the virus. Detecting asymptomatic transmission can help slow down the spread of SARS-  
64 CoV-2.

65

## 66 **Author Summary**

67 The spread of SARS-CoV-2 has led to a global pandemic that is still spreading across countries.

68 We fitted a mathematical model to reported infections in Spain, Italy, South Korea, New York,

69 and Chicago before any large-scale interventions, such as lockdowns and school closures, and

70 found that undetected infected individuals drove the accelerated pace of transmissions. Due to

71 the limited capacity in testing in many of the five locations, undetected cases were most likely

72 asymptomatic and mild to moderate symptomatic infections. Given the explosive nature in the

73 number of cases during the early phase of the pandemic and the latest serological surveys, our

74 study suggested that most active cases were undetected. Other cohort studies have shown that a

75 significant proportion of cases reported little or no symptoms. We also showed that early

76 detection of asymptomatic and mild symptomatic cases can lead to a slower spread of SARS-

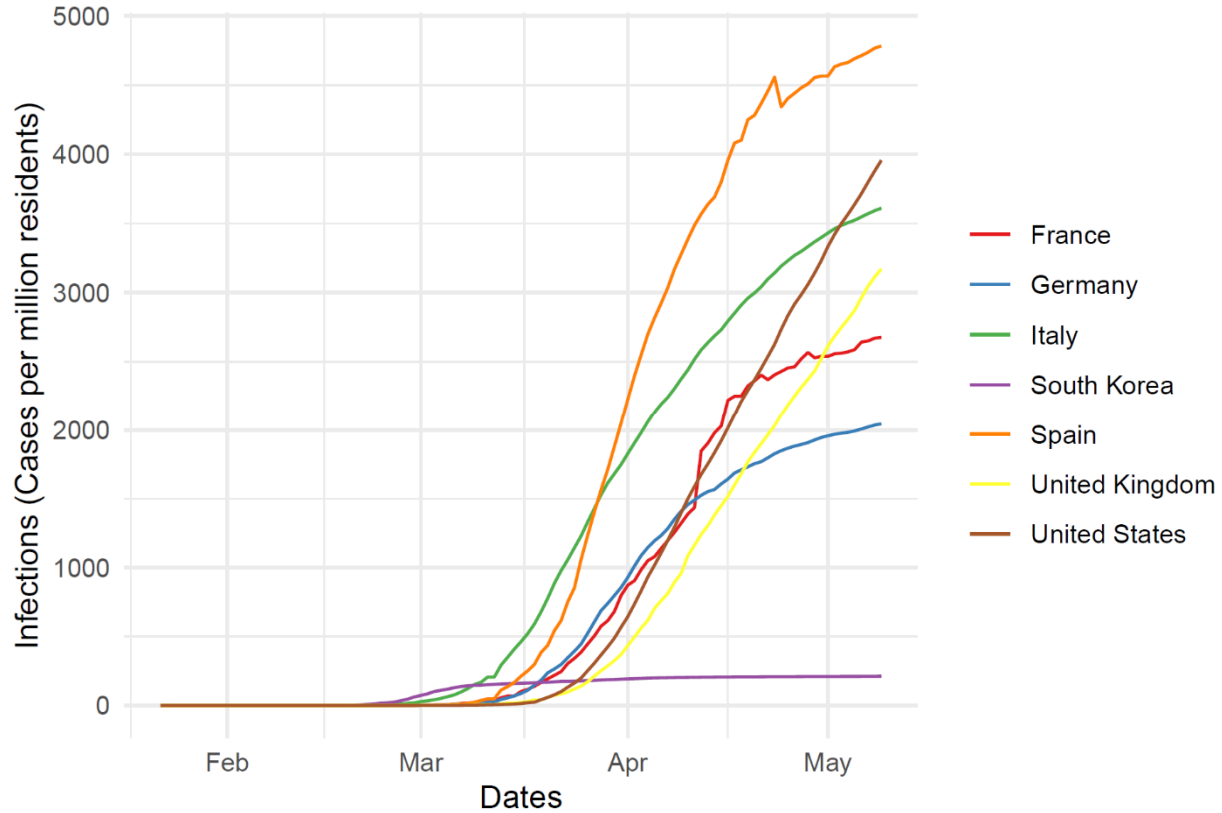
77 CoV-2 as evident in South Korea. Policies targeting symptomatic individuals, such as travel

78 restrictions on affected areas or quarantines of sick individuals, are not as effective because they

79 neglect asymptomatic transmission events.

## 80 **Introduction**

81 The spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was declared  
82 a global pandemic by the World Health Organization on March 11, and confirmed cases of  
83 coronavirus disease 2019 (COVID-19) have since grown exponentially on every inhabited  
84 continent [1]. Despite the rising numbers of cases and deaths, vital questions about the dynamics  
85 of the spread of COVID-19, particularly the “bomb-like” dynamics of the disease, remain. In  
86 most countries, states, and cities, the number of confirmed cases remained low for weeks and  
87 then suddenly exploded, with the number of confirmed cases exponentially increasing in a matter  
88 of days (Figure 1). There are two possible explanations for this dynamic: (i) the growth rate is  
89 intrinsic to the disease or (ii) the rapid growth is due to undercounting of asymptomatic or mild  
90 cases of the disease. The  $R_0$  measure is important for capturing the magnitude and speed that the  
91 population will be infected with SARS-CoV-2. Initial estimates of  $R_0$  ranged widely from 1.4 to  
92 7.1 [2–6]. The variability in reported  $R_0$  between locations could be explained by different  
93 detection rates of asymptomatic and mild symptomatic cases.



94

95 **Figure 1. Confirmed cases of COVID-19 per million residents for France, Germany,**  
96 **Italy, South Korea, Spain, United Kingdom, and the United States.** The rate of  
97 exponential increase are observed for all countries except South Korea where early  
98 surveillance was employed. Source: Center for Systems Science and Engineering [1] at  
99 Johns Hopkins Universty.

100

101 Understanding the percentage of the population infected with SARS-CoV-2 is critical for  
102 developing policies for containment, particularly as the threat of a second wave of disease  
103 appears inevitable. Most countries have instituted extended shutdowns to limit social contacts  
104 and opportunities for disease spread; the lifting of these restrictions in the continued presence of  
105 cases may expose these countries to explosive increases in infections. However, questions  
106 remain regarding the most effective strategies to contain the pandemic in the short- and long-

107 term. Countries that have contained or slowed the spread of disease include China, Singapore,  
108 Hong Kong, Taiwan, and South Korea [7–11]. Through comprehensive strategies including early  
109 detection, extensive testing, shelter-in-place orders, and isolation of infected individuals, these  
110 countries have demonstrated that a combination of control measures can effectively slow  
111 transmission from affected individuals whether they are severely sick or asymptotically  
112 infected. In contrast, isolated measures like broad travel restrictions or enforced quarantine have  
113 proven less effective, as the ability to achieve “zero risk” through these measures is virtually  
114 unattainable in most contexts [12,13]. Moreover, the effectiveness of these strategies, which  
115 primarily target individuals with noticeable symptoms of COVID-19, is questionable given  
116 evidence of asymptomatic and mildly symptomatic carriers as important vectors for community  
117 transmission of SARS-CoV-2 [14–17].

118 Emerging evidence from China [17], Germany [18], Taiwan [19], Iceland [20], and other  
119 places [21] suggests a larger fraction of the population may be asymptotically or mildly  
120 infected than previously thought. The first report of genomic differences in the virus from  
121 Washington State in the United States suggested widespread community transmission as early as  
122 late January [22], which presumably generated asymptomatic or mild cases that were not  
123 identified. Given that the spread of the disease coincided with peak influenza season, it is  
124 reasonable to assume that mild COVID-19 infections could have been misdiagnosed and  
125 undetected. Furthermore, data on SARS-CoV-2’s effect on children have been almost entirely  
126 lacking in the numbers of cases and hospitalizations reported [23]. Some evidence suggests  
127 children are being infected at the same rate as adults, albeit with lesser severity [24,25], implying  
128 that at least 15% of the population in many countries may be asymptomatic when infected, with  
129 potential to be much higher. Studies have also emerged showing similar viral loads between

130 asymptomatic and symptomatic cases, which suggests that there is potential for similar  
131 transmissibility between the two cohorts [18,26].

132 Here we assess the value of  $R_0$  using a Bayesian framework to estimate the role of  
133 asymptomatic cases in driving disease dynamics. Our goal was to explain the data associated  
134 with emerging outbreaks in countries where initial effects of distancing and other measures to  
135 control the disease were largely absent. During this early, “bomb-like” phase, infection count  
136 data are assumed to be largely representative of the transmission dynamics, with some proportion  
137 of the infected population being masked due to asymptomatic and mild symptomatic cases and  
138 shortages in testing. We examined data from Italy and Spain, both of which had large outbreaks  
139 that were not well contained and did not institute early lockdowns. These countries are compared  
140 to South Korea, which largely contained its outbreak through extensive testing and no hard  
141 lockdown. In addition, we contrasted these country-wide outbreaks with two metropolitan areas  
142 in the United States, New York City and Chicago, as dynamics at the city level may differ from  
143 those at the country level (see S1 Appendix for metropolitan area classification). Understanding  
144 the transmission patterns of SARS-CoV-2 can help policymakers predict critical moments in its  
145 progression, such as peak infection, the point when herd immunity has been reached, and local  
146 peaks during periods of relaxed or tightened restrictions on activity.

147



## 148 **Methods**

149 We adapted the Kermack and McKendrick compartmental model to the reported disease  
150 dynamics of COVID-19 [27]. In our model, we assume there is (i) an incubation period for  
151 susceptible individuals that become infected; and (ii) a fraction of individuals who are  
152 asymptomatic or mildly symptomatic and neither tested nor counted as confirmed cases. For a  
153 given population  $N$ , the model is described by the following set of ordinary differential  
154 equations,

155

$$\begin{aligned} \dot{S} &= -\alpha\beta\frac{SC}{N} - \beta\frac{SI}{N} \\ \dot{E} &= \alpha\beta\frac{SC}{N} + \beta\frac{SI}{N} - \mu E \\ (1) \quad \dot{C} &= (1 - \theta)\mu E - \gamma_1 C \\ \dot{I} &= \theta\mu E - \gamma_2 I \\ \dot{R} &= \gamma_1 C + \gamma_2 I \end{aligned}$$

156

157 where  $S$  is the population of susceptible individuals, and  $E$  are exposed individuals incubating  
158 the disease that eventually become infected. After an incubation period,  $\mu^{-1}$ , we assume a  
159 proportion of individuals, defined by  $\theta$ , will transition to a state  $I$  with moderate to severe  
160 symptoms which would result in detection, while the other proportion  $C$  will remain undetected  
161 because they have mild symptoms or are asymptomatic. For undetected individuals, we assumed  
162 there is a reduction of  $\alpha$  which modifies the symptomatic transmission rate  $\beta$ . Finally,  $R$   
163 represents individuals that either recover, die, or remove themselves from transmission through

164 self-quarantine until they are no longer transmissible at rates  $\gamma_1$  for asymptomatic/mild cases and  
165  $\gamma_2$  for symptomatic cases.

### 166 **Bayesian estimation of model parameters**

167 While traditional unbiased curve-fitting methods yield a set of parameter estimations that capture  
168 observed data, they do not account for known prior belief on parameter ranges shown in Table 1.

169 A Bayesian approach to parameter estimation allows us to quantify the credibility of one set of  
170 model parameters. This approach is useful for this context as it provides a range of credible  
171 parameters that describes the observed data and allows us to infer underlying causal pathways  
172 and quantify uncertainty.

**Table 1.** Parameters estimation from other literature where the means and corresponding credible intervals are shown in the parenthesis

Parameter	Definition	Ages 0-64	Ages 65-79	Ages $\geq 80$	Calculation	Reference
$R_0$	<b>Basic Reproduction Number</b>	3.1 (1.9 – 6.5)	3.1 (1.9 – 6.5)	3.1 (1.9 – 6.5)		[6]
$\theta$	Detection (symptomatic) rate	0.038 (0.011-0.121)	0.038 (0.011-0.121)	0.038 (0.011-0.121)	$\frac{\text{Number of symptomatic cases}}{\text{Number of positive cases}}$	[31,32]
$\alpha$	Reduction of infection rate for asymptomatic/mild transmission <sup>†</sup>	0.053 *(0.102)	0.053 *(0.102)	0.053 *(0.102)	$\frac{\text{Asymptomatic transmission coefficient, } \beta_A}{\text{Symptomatic transmission coefficient, } \beta_S}$	[38]
$\beta$	Infection rate for symptomatic transmissions <sup>†</sup>	0.579 (0.443-0.917)	0.672 (0.514-1.065)	0.660 (0.506-1.046)	$\frac{R_0}{\text{Infectious Period}_{Severe}}$	[33,34,39]
$\gamma_1$	Clearance rate for asymptomatic/mild cases <sup>†</sup>	0.204 (0.169-0.303)	0.237 (0.197-0.352)	0.233 (0.193-0.346)	$\frac{1}{\text{Infectious Period}_{Mild}}$	[33,34]
$\gamma_2$	Clearance rate for symptomatic non-hospitalized cases <sup>†</sup>	0.202 (0.168-0.300)	0.235 (0.195-0.348)	0.231 (0.192-0.342)	$\gamma_1 \times \frac{\text{Recovery Time}_{Mild}}{\text{Recovery Time}_{Mod}}$	[33,34]
$\gamma_3$	Clearance rate for symptomatic hospitalized cases <sup>†</sup>	0.145 (0.120-0.215)	0.168 (0.140-0.250)	0.165 (0.137-0.246)	$\gamma_1 \times \frac{\text{Recovery Time}_{Mild}}{\text{Recovery Time}_{Severe}}$	[33,34]
$\varphi$	Hospitalization Rate <sup>‡</sup>	0.202 (0.167-0.237)	0.361 (0.286-0.435)	0.471 (0.308-0.634)	$\frac{\text{Number of patients hospitalized}}{\text{Number of symptomatic cases}}$	[40,41]
$\mu$	Incubation rate	0.202 (0.101-0.364)	0.202 (0.101-0.364)	0.202 (0.101-0.364)	$\frac{1}{\text{Incubation Time}}$	[29,30]

<sup>†</sup> Age groups are: 0-59, 60-69, 70+; <sup>‡</sup> Age groups are: 0-59, 60-69, 70+; 0-64, 65-74, 65+

\* represents standard deviation of a normal distribution

173 To estimate the posterior distribution of the parameters, we obtained data on outbreaks in  
174 Italy, Spain, South Korea, and Chicago from the Center for Systems Science and Engineering at  
175 Johns Hopkins University [1], and for New York City we fitted the model to reported  
176 hospitalizations of COVID-19 patients obtained from the New York Department of Health and  
177 Mental Health [28]. We trained the model to fit the observed data assuming that the first  
178 observed case was in the more severe, symptomatic group. For each initial symptomatic case, we  
179 seeded an additional twenty people in the exposed compartment. We determined these  
180 timeframes in 2020 to be January 31 to March 21 for Italy, February 1 to March 25 for Spain,  
181 January 22 to March 3 for South Korea, January 22 to April 22 for Chicago, and February 17 to  
182 April 10 for New York City, which coincide with the first observed case.

183 We used Markov Chain Monte Carlo (MCMC) methods assuming a uniform prior  
184 density, specifically the Metropolis-Hasting Algorithm, to approximate a posterior distribution of  
185 the parameter set. To explore all possibilities and demonstrate robustness in the estimation  
186 results, we ran three cases for each locale with different Bayesian priors for the symptomatic rate  
187 ( $\theta$ ). Reported parameter estimates are the medians of the posterior distributions and the 95%  
188 credible intervals (CrI) from quantiles of the posterior distribution. We provide a full description  
189 of the mathematical structure of our models and estimation procedures in the S1 Appendix. The  
190 model was fit to the number of cumulative infections and the likelihood was estimated based on  
191 the sum of squared residuals (SSR). For New York City, we included the number of cumulative  
192 hospitalizations into the SSR error metric in order to estimate the hospitalization rate. The error  
193 between the results of the model yielding the expected number of observed cases and the  
194 observed number of cases is considered to be distributed with  $\mathcal{N}(0, \sigma^2)$ .

195

## 196 **Biological Parameters and Prior Beliefs**

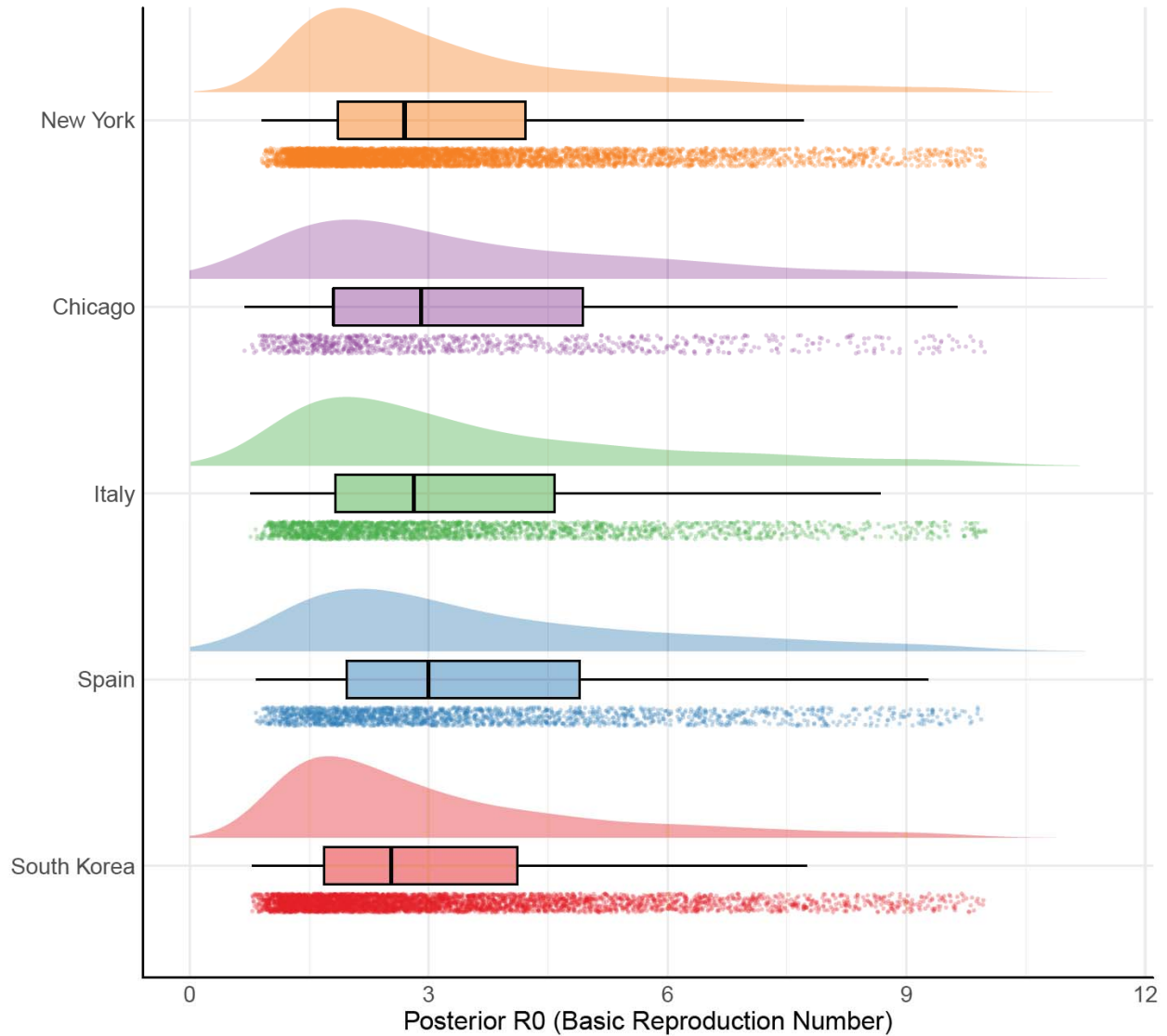
197 In order to assess fit, determine model parameter priors, and provide context for the analysis, we  
198 conducted a literature search of the biology and transmission of SARS-CoV-2 using peer-  
199 reviewed literature and non-peer-reviewed literature available on pre-print servers medRxiv,  
200 bioRxiv and SSRN's First Look (Table 1). From the literature, we found a mean incubation  
201 period, defined as the time from exposure to onset of illness, of 4.95 (CI, 2.75-9.90) [29,30]. The  
202 proportion of the population with no/mild/moderate symptoms was estimated to be as high as  
203 96.2% of the population [31,32]. Based on the dynamics of earlier coronaviruses [33], the  
204 recovery periods for symptomatic and asymptomatic cases were estimated to be ~4.9 days while  
205 hospitalized cases were around 6.9 days [33,34]. In our model,  $\gamma_1$ ,  $\gamma_2$ , and  $\gamma_3$  represented the  
206 infectiousness period, which is not necessarily equivalent to the recovery rate since  
207 severe/moderate symptomatic and hospitalized cases are effectively removed due to self-  
208 quarantine. Data on confirmed cases reported in the US estimate the hospitalization rate for all  
209 ages to range from 20.7%-31.4%, with lower rates observed among younger populations and  
210 higher rates observed among older age groups. Finally, we estimated the transmission rates by  
211 asymptomatic,  $\alpha\beta$ , and symptomatic,  $\beta$ , persons based on calculations of the basic reproduction  
212 number,  $R_0$ , of COVID-19, clearance estimates, and contact rates. Values for  $R_0$  in the literature  
213 range from 1.4 to 7.1 [2-6]; given these values we assumed the transmission rate for  $\beta$  was likely  
214 between 0.01 and 2.

215 We conducted MCMC sampling for three cases using different bounds on weakly-  
216 informed priors for the symptomatic rate in order to test for robustness. The main results are  
217 presented as the fitted parameter values of the posterior distribution with unconstrained  
218 symptomatic rate ( $\theta$ ) priors and their relative relationship to the known biology. In addition to

219 the unconstrained prior of  $\theta$ , we refitted the data twice using two different prior assumptions of  
220 the symptomatic fraction:  $\theta$  is less than 50%, and  $\theta$  is less than 10% with the results located in S1  
221 Appendix. In our parameter estimation, we assumed that transmission from our unobserved  
222 group was lower by a factor of  $\alpha$  than transmission from the symptomatic observed group,  $\beta$ , as  
223 mild/asymptomatic cases are likely to shed fewer viruses and thus be less transmissible. Thus,  
224 we constrained  $\alpha$  between 0.01-0.99 and  $\beta$  between 0.01-2.00. As the process is stochastic, we  
225 evaluated the median and credible intervals of the posterior distribution obtained from this  
226 procedure to guide our understanding of the parameters and to assess the ability of the model to  
227 fit actual data with biologically plausible values.

## 228 **Results**

229 Assuming a uniform prior with no constraint on the fraction of the population that was  
230 symptomatic,  $\theta$ , the parameter estimation from our MCMC sampling resulted in a posterior  
231 median for the symptomatic rate of 0.04 (95% CrI, 0.01-0.41), 0.03 (95% CrI, 0.01-0.32), 0.18  
232 (95% CrI, 0.02-0.85) for Italy, Spain, and South Korea, respectively, while the symptomatic rates  
233 of metropolitan areas of New York City and Chicago were 0.25 (95% CrI, 0.01-0.89) and 0.03  
234 (95% CrI, 0.01-0.36), respectively. The median posterior values of these parameter ranges  
235 resulted in a calculated  $R_0$  of 3.25 (95% CrI, 1.09-29.77), 3.62 (95% CrI, 1.13-34.89), 2.75 (95%  
236 CrI, 1.04-22.44), 3.31 (95% CrI, 1.69-20.55), and 3.46 (95% CrI, 1.01-34.41) for Italy, Spain,  
237 South Korea, New York City, and Chicago, respectively (Table 2 and Figure 2).



238

239 **Figure 2. Estimated Basic Reproduction Number  $R_0$  for New York, Chicago, Italy,**

240 **Spain, and South Korea from Bayesian distribution. Each curve is the posterior**

241 **distribution generated from the MCMC sampling for the  $R_0$  of each location are shown. For**

242 **each location, the raw data, box and whisker plot (median, interquartile ranges, 95%**

243 **credible intervals), and probability density are displayed top-to-bottom.**

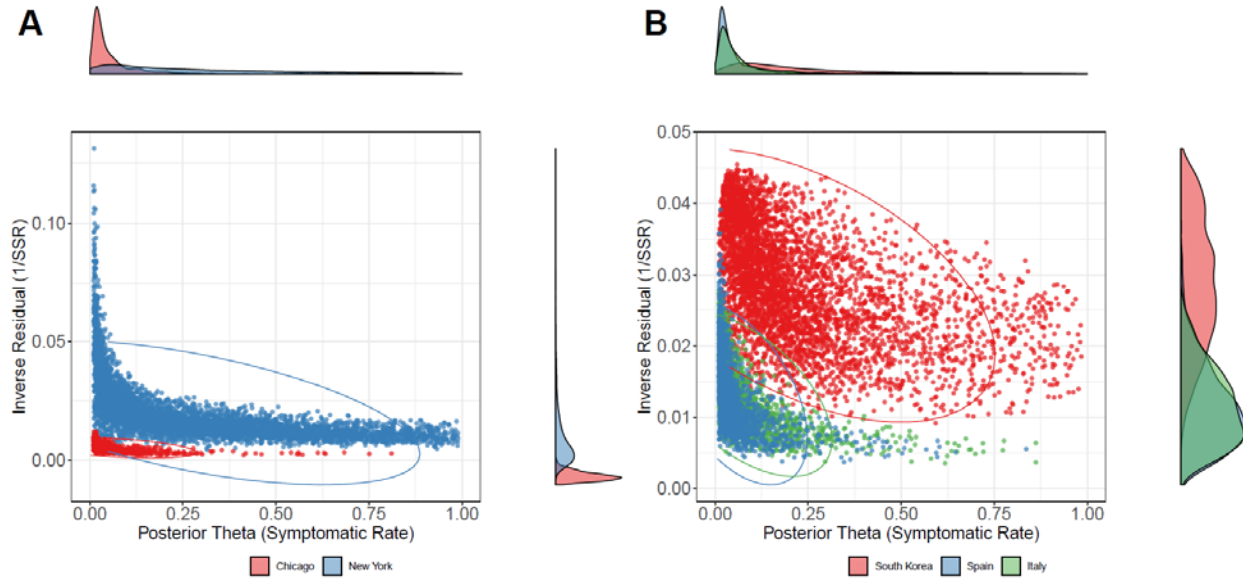
**Table 2.** Posterior parameters estimation for US cities from the Bayesian analysis using the epidemiological model where the medians and corresponding 95% credible intervals are shown in the parenthesis.

Variable	Prior Distribution	Countries			Cities	
		Italy	Spain	South Korea	New York	Chicago
$R_0$	--	<b>3.25</b> (1.09-29.77)	<b>3.62</b> (1.13-34.89)	<b>2.75</b> (1.04-22.44)	<b>3.31</b> (1.69-20.55)	<b>3.46</b> (1.01-34.41)
$\theta$	U(0.01,0.99) <sup>†</sup>	<b>0.04</b> (0.01-0.41)	<b>0.03</b> (0.01-0.32)	<b>0.18</b> (0.02-0.85)	<b>0.25</b> (0.01-0.89)	<b>0.03</b> (0.01-0.36)
$\alpha$	U(0.01,0.99)	0.67 (0.24-0.97)	0.68 (0.22-0.97)	0.58 (0.09-0.96)	0.61 (0.09-0.97)	0.61 (0.16-0.97)
$\beta$	U(0.01,2)	1.38 (0.53-1.96)	1.36 (0.51-1.96)	1.29 (0.43-1.95)	1.35 (0.58-1.96)	1.27 (0.4-1.95)
$\gamma_1$	U(0.01,0.99)	0.22 (0.02-0.84)	0.18 (0.02-0.78)	0.33 (0.03-0.94)	0.36 (0.03-0.93)	0.17 (0.02-0.85)
$\gamma_2$	U(0.01,0.30)	0.49 (0.03-0.96)	0.5 (0.03-0.96)	0.37 (0.03-0.95)	0.38 (0.03-0.95)	0.5 (0.03-0.97)
$\gamma_3$	U(0.01,0.30)	--	--	--	0.46 (0.03-0.95)	--
$\varphi$	U(0.01,0.50)	--	--	--	0.34 (0.13-0.49)	--
$\mu$	U(0.01,0.99)	0.31 (0.07-0.92)	0.26 (0.05-0.89)	0.25 (0.05-0.92)	0.42 (0.1-0.94)	0.12 (0.02-0.83)
Inverse Sum of Squared Residuals (SSR <sup>-1</sup> )	--	0.013 (0.006-0.024)	0.013 (0.006-0.028)	0.029 (0.014-0.042)	0.012 (0.007-0.026)	0.005 (0.003-0.01)



245           Variation between the locales that had higher symptomatic infections was primarily  
246 governed by differences in the length of the infectiousness of the asymptomatic group. In the  
247 high symptomatic scenarios of South Korea and New York City, the fit of the parameter  
248 produced median infectious periods of 3 days in contrast to lower symptomatic locales where the  
249 infectious period was estimated to be about 5-6 days. The symptomatic group at all locales had  
250 median infectious periods of 2-3 days.

251           Assuming stronger priors with constrained uniform priors of the symptomatic rate  $\theta$  to be  
252 under 50% or 10%, the inverse sum of squared residuals ( $SSR^{-1}$ ) was relatively lower, resulting  
253 in a slightly better fit for all locations (Tables S1 and S2 in S1 Appendix). However, the  
254 parameter estimation of constrained priors still resulted in similar values as to their  
255 unconstrained counterparts. The posterior density of  $\theta$  was truncated for South Korea and New  
256 York City if we constrained  $\theta$ 's prior to be less than 10 percent (Tables S31 and S32 in S1  
257 Appendix). Figure 3 illustrates a negative relationship between symptomatic rate  $\theta$  and  $SSR^{-1}$ ,  
258 indicating that lower estimates of symptomatic rates resulted in higher posterior likelihoods. In  
259 all projections using the accepted posterior parameter samples, the simulations overestimated  
260 actual case numbers following the initial period of infection, as non-pharmaceutical interventions  
261 such as lockdowns artificially reduced disease spread (Figure 4).



262

263

264

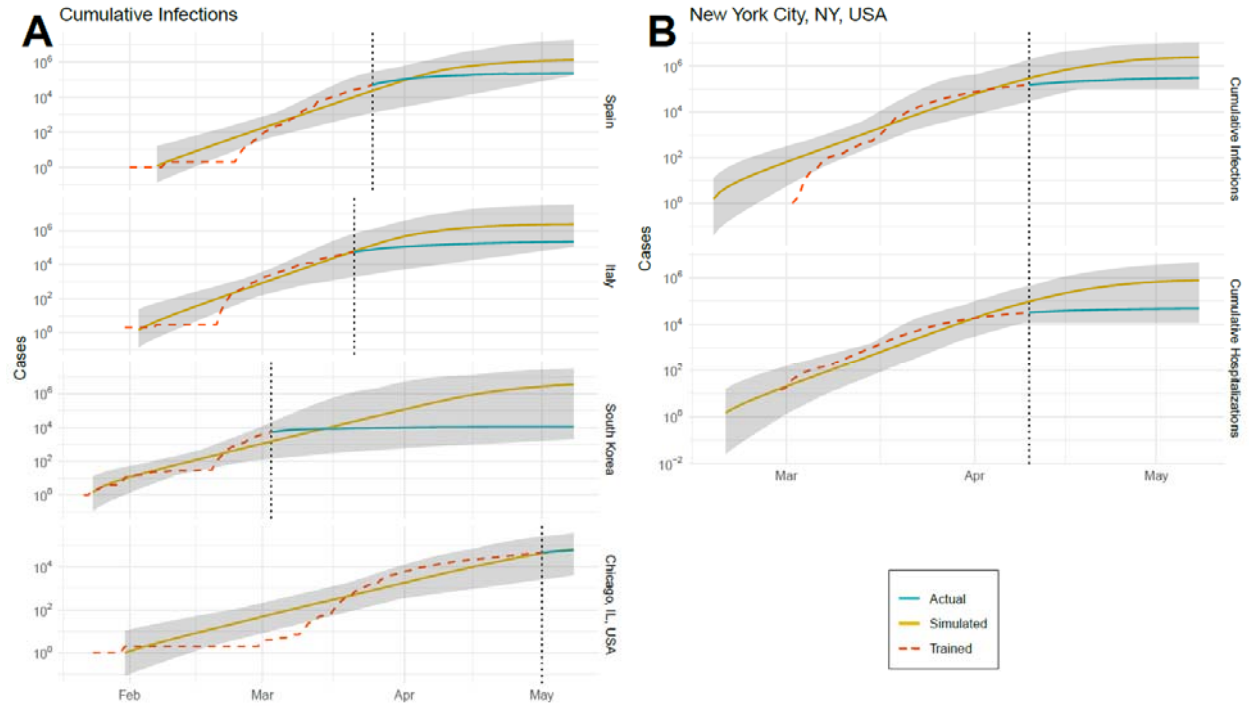
265

266

267

268

**Figure 3. Scatterplot of inverse residual (goodness of fit) versus posterior sample of symptomatic rate.** The posterior distribution of symptomatic rate  $\theta$  and the inverse of sum of squared residuals ( $SSR^{-1}$ ) for Chicago and New York City (A) and the same set of priors for Italy, Spain, and South Korea (B). The marginal posterior density of  $\theta$  and  $SSR^{-1}$  are displayed on the right and top of the scatter plot, respectively. The ellipses encircle 95% of simulation runs for each location.



269

270

271

272

273

274

275

276

277

278

279

280

281 **Discussion**

**Figure 4. Simulated versus actual cumulative infections and hospitalization.** (A) The simulated and actual cumulative infections for each location (Spain, Italy, South Korea, Chicago, and New York City are shown) for three different prior distributions of the symptomatic rate ( $\theta$ ). The turquoise line (simulated) corresponds with the simulated outputs for cumulative confirmed cases, which were fitted to the yellow line (trained) representing the actual data. The red dashed line represents the actual data after the explosion date, which is demarcated by the vertical black dotted line. The grey ribbon represents the upper and lower bounds of the MCMC sampling for all cases for each day. (B) Simulated and actual cumulative hospitalizations and cumulative infections for New York City.

282 With no end in sight to the ongoing COVID-19 pandemic, understanding the rate of disease  
283 spread and the extent of people infected is vital to designing effective containment policies. This  
284 analysis contributes to available disease models, which play an integral role in defining the  
285 policy choices of government officials, by providing accurate evidence that conforms to  
286 observed data. Given the growth rate in infection cases in the exponential phase and assuming  
287 the force of infection for undetected cases is lower, our results for all scenarios and countries  
288 suggest that the undetected population is significantly larger than the symptomatic population,  
289 the latter of which accounts for only 3-25% of the total infected population. The value of  $R_0$   
290 varies from 2.8-3.6, which corresponds with a growing consensus that  $R_0$  is approximately 3.1  
291 (Table 1). The variation in symptomatic cases is almost certainly a function of testing, which has  
292 been quite variable among cities and countries, yet even in South Korea, which had extremely  
293 high rates of testing during the beginning of the pandemic, our models still suggest large  
294 fractions of cases were likely missed due to asymptomatic/mild infections. Nevertheless, higher  
295 testing rates in South Korea also led to a slower transmission of SARS-CoV-2. This is also true  
296 when you compare testing in New York City versus Chicago, where testing per capita during the  
297 exponential phase was higher in New York City.

298 If the proportion of individuals that are asymptotically infected is higher than initially  
299 assumed, policies targeting symptomatic individuals, such as travel restrictions on affected areas  
300 or quarantines of sick individuals, are not productive. By the time these policies can be  
301 implemented, a large proportion of the population may already be infected but not yet infectious  
302 due to the long incubation period [35]. Evidence from China suggests that the observed infection  
303 data lagged reported infections by about two weeks -- despite a dramatic drop in transmission  
304 between January 15 and January 25, the rate of newly confirmed cases did not begin to level off

305 country-wide and within each local city until two weeks later. As of May 18, 2020, the uptick of  
306 COVID-19 cases has forced northeast China to reinitiate lockdown measures [36], which  
307 indicates the sensitivities and caution around case reemergence. Accurate assessments of the risk  
308 of community spread are needed to inform these decisions about when and where to re-impose  
309 restrictions to contain the spread of SARS-CoV-2. Our study found that asymptomatic cases may  
310 be easily missed without extensive testing, leading to the continued, undetected spread of disease  
311 despite possibly lower transmission rates among asymptomatic individuals compared to more  
312 severe cases. Thus, identifying asymptomatic cases is imperative for reducing widespread  
313 transmission and explosive growth of the disease, especially if a large majority of the population  
314 remains susceptible.

315 A potentially large asymptotically infected population has important implications for  
316 policy regarding herd immunity. In recent reports, the vast majority of available evidence  
317 suggests that infected individuals develop some level of immunity to circulating strains of  
318 SARS-CoV-2, but there are potential differences in immunity between symptomatic and  
319 asymptomatic cases [37]. As more of the population gains immunity, the susceptible population  
320 will decrease and disease spread will slow. When transmissibility between contacts falls because  
321 of widespread immunity, the effective reproductive number is reduced, the disease spreads more  
322 slowly, and the threat posed by widespread numbers of infected individuals fades away. Such  
323 widespread immunity would allow restrictions to be lifted sooner, as most individuals would not  
324 pose a transmission risk to the general public and would abrogate the possibility of a second peak  
325 in the future. Our analysis underscores the likelihood of a large, hitherto undetected population  
326 of immune individuals in areas hit hard by the pandemic. Widespread, representative serological  
327 surveys are therefore essential to understanding the extent of disease spread and the necessity of

328 mandatory social distancing policies to contain the spread of the virus. Understanding the  
329 potential risks requires serological surveys to be conducted as soon as possible across  
330 representative populations of disease burden to improve our understanding of disease  
331 transmission and to inform better policies regarding quarantines.

332         As with most COVID-19 modeling studies, the limitations of epidemiological models  
333 have been constrained by data and testing quality regarding the true prevalence of SARS-CoV-2  
334 infections. We assumed for the beginning stages of the pandemic, the confirmed cases only  
335 reflected cases that were mostly moderate and severely symptomatic, while asymptomatic and  
336 mild cases were mostly overlooked due to the limited supply of testing kits. The uncertainty of  
337 these testing differences was implicitly captured in the Bayesian framework. The population size  
338 for each country are simplified and assumed to be static. Nevertheless, the parsimonious model  
339 provides a conservative estimate of undetected cases since traveling individuals make up a  
340 negligible proportion population, and if these travelers are infected and undetected in early stages  
341 of the pandemic, it will further support our claim. Furthermore, the estimation of transmission is  
342 assumed to be homogenous population mixing, while in reality contact networks are  
343 heterogeneous with varying contact patterns.

## 344 **References**

- 345 1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real  
346 time. *Lancet Infect Dis.* 2020;0. doi:10.1016/S1473-3099(20)30120-1
- 347 2. Majumder MS, Mandl KD. Early in the epidemic: impact of preprints on global discourse  
348 about COVID-19 transmissibility. *Lancet Glob Health.* 2020;0.  
349 doi:[https://doi.org/10.1016/S2214-109X\(20\)30113-3](https://doi.org/10.1016/S2214-109X(20)30113-3)
- 350 3. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-Based Analysis, Modelling and  
351 Forecasting of the COVID-19 outbreak. *medRxiv.* 2020; 2020.02.11.20022186.  
352 doi:<https://doi.org/10.1101/2020.02.11.20022186>
- 353 4. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is  
354 higher compared to SARS coronavirus. *J Travel Med.* 2020;27.  
355 doi:<https://doi.org/10.1093/jtm/taaa021>
- 356 5. Distanto C, Piscitelli P, Miani A. Covid-19 Outbreak Progression in Italian Regions:  
357 Approaching the Peak by March 29th. *medRxiv.* 2020; 2020.03.30.20043612.  
358 doi:<https://doi.org/10.1101/2020.03.30.20043612>
- 359 6. Park M, Cook AR, Lim JT, Sun Y, Dickens BL. A Systematic Review of COVID-19  
360 Epidemiology Based on Current Evidence. *J Clin Med.* 2020;9: 967.  
361 doi:10.3390/jcm9040967
- 362 7. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New  
363 Technology, and Proactive Testing. *JAMA.* 2020 [cited 5 Apr 2020].  
364 doi:<https://doi.org/10.1001/jama.2020.3151>
- 365 8. Tang B, Xia F, Bragazzi NL, Wang X, He S, Sun X, et al. Lessons drawn from China and  
366 South Korea for managing COVID-19 epidemic: insights from a comparative modeling  
367 study. *medRxiv.* 2020; 2020.03.09.20033464.  
368 doi:<https://doi.org/10.1101/2020.03.09.20033464>
- 369 9. Balilla J. Assessment of COVID-19 Mass Testing: The Case of South Korea. *SSRN.* 2020  
370 [cited 5 Apr 2020]. doi:<http://dx.doi.org/10.2139/ssrn.3556346>
- 371 10. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of  
372 non-pharmaceutical interventions against COVID-19 and influenza in Hong Kong: an  
373 observational study. *medRxiv.* 2020; 2020.03.12.20034660.  
374 doi:<https://doi.org/10.1101/2020.03.12.20034660>
- 375 11. Tariq A, Lee Y, Roosa K, Blumberg S, Yan P, Ma S, et al. Real-time monitoring the  
376 transmission potential of COVID-19 in Singapore, February 2020. *medRxiv.* 2020;  
377 2020.02.21.20026435. doi:<https://doi.org/10.1101/2020.02.21.20026435>

- 378 12. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of  
379 travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak.  
380 Science. 2020;368: 395–400. doi:10.1126/science.aba9757
- 381 13. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al.  
382 Quarantine alone or in combination with other public health measures to control  
383 COVID-19: a rapid review. Cochrane Database Syst Rev. 2020 [cited 18 Jun 2020].  
384 doi:10.1002/14651858.CD013574
- 385 14. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the Asymptomatic Proportion  
386 of 2019 Novel Coronavirus onboard the Princess Cruises Ship, 2020. medRxiv. 2020;  
387 2020.02.20.20025866. doi:<https://doi.org/10.1101/2020.02.20.20025866>
- 388 15. Cheng H-Y, Jian S-W, Liu D-P, Ng T-C, Huang W-T, Team TC-19 outbreak investigation,  
389 et al. High transmissibility of COVID-19 near symptom onset. medRxiv. 2020;  
390 2020.03.18.20034561. doi:10.1101/2020.03.18.20034561
- 391 16. Tao Y, Cheng P, Chen W, Wan P, Chen Y, Yuan G, et al. High incidence of asymptomatic  
392 SARS-CoV-2 infection, Chongqing, China. medRxiv. 2020; 2020.03.16.20037259.  
393 doi:<https://doi.org/10.1101/2020.03.16.20037259>
- 394 17. Bai Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, et al. Presumed Asymptomatic Carrier  
395 Transmission of COVID-19. JAMA. 2020 [cited 3 Apr 2020].  
396 doi:<https://doi.org/10.1001/jama.2020.2565>
- 397 18. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission  
398 of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med.  
399 2020;382: 970–971. doi:<https://doi.org/10.1056/NEJMc2001468>
- 400 19. Liu Y-C, Liao C-H, Chang C-F, Chou C-C, Lin Y-R. A Locally Transmitted Case of  
401 SARS-CoV-2 Infection in Taiwan. N Engl J Med. 2020;382: 1070–1072.  
402 doi:<https://doi.org/10.1056/NEJMc2001573>
- 403 20. CNN TJ. Iceland lab’s testing suggests 50% of coronavirus cases have no symptoms. In:  
404 CNN [Internet]. [cited 3 Apr 2020]. Available:  
405 <https://www.cnn.com/2020/04/01/europe/iceland-testing-coronavirus-intl/index.html>
- 406 21. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection. Ann Intern Med.  
407 2020 [cited 23 Jun 2020]. doi:10.7326/M20-3012
- 408 22. Bedford T. Cryptic transmission of novel coronavirus revealed by genomic epidemiology.  
409 In: Cryptic transmission of novel coronavirus revealed by genomic epidemiology [Internet].  
410 2 Mar 2020 [cited 3 Apr 2020]. Available: [https://bedford.io/blog/ncov-cryptic-](https://bedford.io/blog/ncov-cryptic-transmission/)  
411 [transmission/](https://bedford.io/blog/ncov-cryptic-transmission/)
- 412 23. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus  
413 Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From



- 414 the Chinese Center for Disease Control and Prevention. *JAMA*. 2020 [cited 3 Apr 2020].  
415 doi:<https://doi.org/10.1001/jama.2020.2648>
- 416 24. Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiological Characteristics of  
417 2143 Pediatric Patients With 2019 Coronavirus Disease in China. *Pediatrics*. 2020 [cited 5  
418 Apr 2020]. doi:<https://doi.org/10.1542/peds.2020-0702>
- 419 25. Kelvin AA, Halperin S. COVID-19 in children: the link in the transmission chain. *Lancet*  
420 *Infect Dis*. 2020;0. doi:[https://doi.org/10.1016/S1473-3099\(20\)30236-X](https://doi.org/10.1016/S1473-3099(20)30236-X)
- 421 26. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in  
422 Upper Respiratory Specimens of Infected Patients. *N Engl J Med*. 2020;382: 1177–1179.  
423 doi:10.1056/NEJMc2001737
- 424 27. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics—I.  
425 *Bull Math Biol*. 1991;53: 33–55. doi:10.1007/BF02464423
- 426 28. COVID-19: Data - NYC Health. [cited 2 May 2020]. Available:  
427 <https://www1.nyc.gov/site/doh/covid/covid-19-data.page>
- 428 29. Liu T, Hu J, Xiao J, He G, Kang M, Rong Z, et al. Time-varying transmission dynamics of  
429 Novel Coronavirus Pneumonia in China. *Systems Biology*; 2020 Jan.  
430 doi:10.1101/2020.01.25.919787
- 431 30. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation  
432 Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed  
433 Cases: Estimation and Application. *Ann Intern Med*. 2020;172: 577–582.  
434 doi:10.7326/M20-0504
- 435 31. Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, et al. COVID-19  
436 Antibody Seroprevalence in Santa Clara County, California. *medRxiv*. 2020;  
437 2020.04.14.20062463. doi:10.1101/2020.04.14.20062463
- 438 32. Sutton D, Fuchs K, D’Alton M, Goffman D. Universal Screening for SARS-CoV-2 in  
439 Women Admitted for Delivery. *New England Journal of Medicine*. Massachusetts Medical  
440 Society; 2020. doi:10.1056/NEJMc2009316
- 441 33. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission  
442 dynamics of SARS-CoV-2 through the postpandemic period. *Science*. 2020;368: 860–868.  
443 doi:10.1126/science.abb5793
- 444 34. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of  
445 COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a  
446 retrospective cohort study. *Lancet Infect Dis*. 2020;0. doi:10.1016/S1473-3099(20)30287-5
- 447 35. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral  
448 shedding and transmissibility of COVID-19. *Nat Med*. 2020;26: 672–675.  
449 doi:10.1038/s41591-020-0869-5

- 450 36. Over 100 Million in China's Northeast Face Renewed Lockdown. Bloomberg.com. 18 May  
451 2020. Available: [https://www.bloomberg.com/news/articles/2020-05-18/over-100-million-](https://www.bloomberg.com/news/articles/2020-05-18/over-100-million-in-china-s-northeast-thrown-back-under-lockdown)  
452 [in-china-s-northeast-thrown-back-under-lockdown](https://www.bloomberg.com/news/articles/2020-05-18/over-100-million-in-china-s-northeast-thrown-back-under-lockdown). Accessed 18 May 2020.
- 453 37. Long Q-X, Tang X-J, Shi Q-L, Li Q, Deng H-J, Yuan J, et al. Clinical and immunological  
454 assessment of asymptomatic SARS-CoV-2 infections. *Nat Med*. 2020; 1–5.  
455 doi:10.1038/s41591-020-0965-6
- 456 38. Ogbunugafor CB, Miller-Dickson MD, Meszaros VA, Gomez LM, Murillo AL, Scarpino  
457 SV. The intensity of COVID-19 outbreaks is modulated by SARS-CoV-2 free-living  
458 survival and environmental transmission. *medRxiv*. 2020; 2020.05.04.20090092.  
459 doi:10.1101/2020.05.04.20090092
- 460 39. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and  
461 international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling  
462 study. *The Lancet*. 2020;395: 689–697. doi:10.1016/S0140-6736(20)30260-9
- 463 40. CDC COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus  
464 Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. *Morb Mortal*  
465 *Wkly Rep*. 2020;69: 343–346. doi:10.15585/mmwr.mm6912e2
- 466 41. Garg S. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-  
467 Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020.  
468 *MMWR Morb Mortal Wkly Rep*. 2020;69: 458–464. doi:10.15585/mmwr.mm6915e30
- 469  
470  
  
471

472 **Supporting information**

473 **S1 Appendix. Model Specifications, Deriving the Next Generation Matrix and Calculating**

474 **R0, Bayesian Inference of Model Parameters, Population Inclusion for US Cities, and**

475 **Supplement Plots**