Modeling the "Bomb-Like" Dynamics of COVID-19 with Undetected Transmissions and the Implications for Policy

Gary Lin, PhD[1,2,*], Anindya Bhaduri, PhD[3], Alexandra T. Strauss, MD, MIE[2,4], Maxwell Pinz, MS[3], Diego A. Martinez, PhD[1,2], Katie K. Tseng, MPH[5], Oliver Gatalo, MPH[6], Andrew T. Gaynor, PhD[6], Efrain Hernandez-Rivera, PhD[6], Emily Schueller, MA[5], Yupeng Yang, MS[5], Simon A. Levin, PhD[7,8], Eili Y. Klein, PhD[1,2,5,9], For the CDC MInD-Healthcare Program

1. Department of Emergency Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA
2. Center for Data Science in Emergency Medicine, Johns Hopkins University, Baltimore, MD, USA
3. Department of Civil and Systems Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, USA
4. Division of Gastroenterology and Hepatology, Department of Medicine, Johns Hopkins University, Baltimore, MD, USA
5. Center for Disease Dynamics, Economics, and Policy, Silver Spring, MD, USA
6. U.S. Army Research Laboratory, Aberdeen, MD, USA
7. Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA
8. Princeton Environmental Institute, Princeton University, Princeton, NJ, USA
9. Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

*** Correspond Author:**

Gary Lin

Email: glin15@jhu.edu

Author Contributions

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

Abstract

Understanding the transmission dynamics of the severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) is critical to inform sound policy decisions. We demonstrate how the transmission of undetected cases with pre-symptomatic, asymptomatic and mild symptoms, which are typically underreported due to lower testing capacity, explained the "bomb-like" behavior of exponential growth in the coronavirus disease 2019 (COVID-19) cases during early stages before the effects of large-scale, non-pharmaceutical interventions such as social distancing, school closures, or lockdowns. Using a Bayesian approach to epidemiological compartmental modeling, we captured the initial stages of the pandemic resulting in the explosion of cases and compared the parameter estimation with empirically measured values from the current knowledgebase. Parameter estimation was conducted using Markov chain Monte Carlo (MCMC) sampling methods with a Bayesian inference framework to estimate the proportion of undetected cases. Using data from the exponential phase of the pandemic prior to the implementation of interventions we estimated the basic reproductive number (R_0) and symptomatic rates in Italy, Spain, South Korea, New York City, and Chicago. From this modeling study, R_0 was estimated to be 3·25 (95% CrI, 1·09-29·77), 3·62 (95% CrI, 1·13-34·89), 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), respectively. For all locations, 3-25% of infected patients were identified with moderate and severe symptoms in the early stage of the COVID-19 pandemic. Our modeling results support the mounting evidence that potentially large fractions of the infected population were undetected with asymptomatic and mild symptoms. Furthermore, a significant number of models of transmission that do not account for these asymptomatic cases may lead to an underestimation of R_0 and, subsequently, policies that do not sufficiently reduce transmission to contain the spread

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

Author Summary

The spread of SARS-CoV-2 has led to a global pandemic that is still spreading across countries. We fitted a mathematical model to reported infections in Spain, Italy, South Korea, New York, and Chicago before any large-scale interventions, such as lockdowns and school closures, and found that undetected infected individuals drove the accelerated pace of transmissions. Due to the limited capacity in testing in many of the five locations, undetected cases were most likely asymptomatic and mild to moderate symptomatic infections. Given the explosive nature in the number of cases during the early phase of the pandemic and the latest serological surveys, our study suggested that most active cases were undetected. Other cohort studies have shown that a significant proportion of cases reported little or no symptoms. We also showed that early detection of asymptomatic and mild symptomatic cases can lead to a slower spread of SARS-CoV-2 as evident in South Korea. Policies targeting symptomatic individuals, such as travel restrictions on affected areas or quarantines of sick individuals, are not as effective because they neglect asymptomatic transmission events.

Introduction

The spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was declared a global pandemic by the World Health Organization on March 11, and confirmed cases of coronavirus disease 2019 (COVID-19) have since grown exponentially on every inhabited continent [1]. Despite the rising numbers of cases and deaths, vital questions about the dynamics of the spread of COVID-19, particularly the "bomb-like" dynamics of the disease, remain. In most countries, states, and cities, the number of confirmed cases remained low for weeks and then suddenly exploded, with the number of confirmed cases exponentially increasing in a matter of days (Figure 1). There are two possible explanations for this dynamic: (i) the growth rate is intrinsic to the disease or (ii) the rapid growth is due to undercounting of asymptomatic or mild cases of the disease. The R_0 measure is important for capturing the magnitude and speed that the population will be infected with SARS-CoV-2. Initial estimates of R_0 ranged widely from 1·4 to 7·1 [2–6]. The variability in reported R_0 between locations could be explained by different detection rates of asymptomatic and mild symptomatic cases.

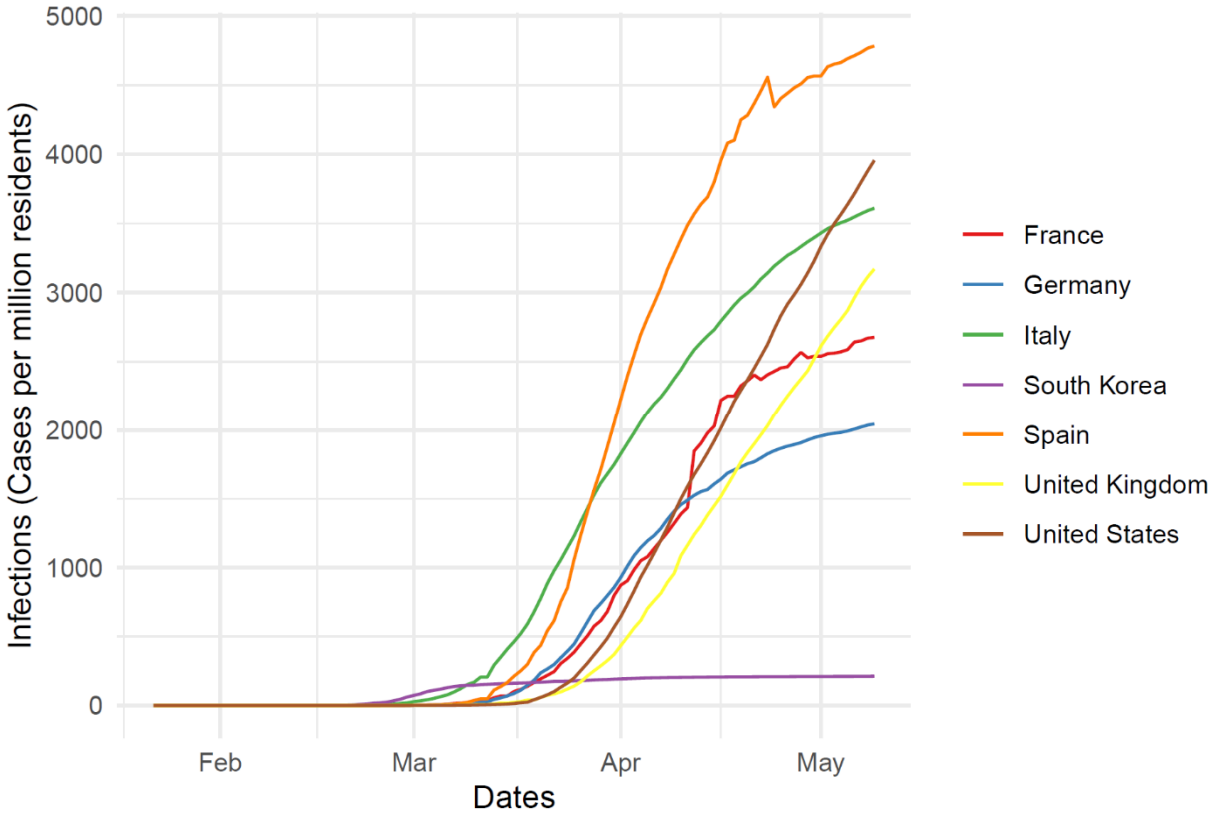

Figure 1. Confirmed cases of COVID-19 per million residents for France, Germany, Italy, South Korea, Spain, United Kingdon, and the United States. The rate of exponential increase are observed for all countries except South Korea where early surveillence was employed. Source: Center for Systems Science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

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

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

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

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

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

Methods

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

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

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

self-quarantine until they are no longer transmissible at rates γ_1 for asymptomatic/mild cases and γ_2 for symptomatic cases.

Bayesian estimation of model parameters

While traditional unbiased curve-fitting methods yield a set of parameter estimations that capture observed data, they do not account for known prior belief on parameter ranges shown in Table 1. A Bayesian approach to parameter estimation allows us to quantify the credibility of one set of model parameters. This approach is useful for this context as it provides a range of credible parameters that describes the observed data and allows us to infer underlying causal pathways 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 ≥ 80	Calculation	Reference
R_0	**Basic Reproduction Number**	3.1 (1.9 – 6.5)	3.1 (1.9 – 6.5)	3.1 (1.9 – 6.5)		[6]
θ	Detection (symptomatic) rate	0.038 (0.011-0.121)	0.038 (0.011-0.121)	0.038 (0.011-0.121)	$\frac{Number\ of\ symptomatic\ cases}{Number\ of\ positive\ cases}$	[31,32]
α	Reduction of infection rate for asymptomatic/mild transmission†	0.053 *(0.102)	0.053 *(0.102)	0.053 *(0.102)	$\frac{Asymptomatic\ transmission\ coefficient,\ \beta_A}{Symptomatic\ transmission\ coefficient,\ \beta_S}$	[38]
β	Infection rate for symptomatic transmissions†	0.579 (0.443-0.917)	0.672 (0.514-1.065)	0.660 (0.506-1.046)	$\frac{R_0}{Infectious\ Period_{Severe}}$	[33,34,39]
γ_1	Clearance rate for asymptomatic/mild cases†	0.204 (0.169-0.303)	0.237 (0.197-0.352)	0.233 (0.193-0.346)	$\frac{1}{Infectious\ Period_{Mild}}$	[33,34]
γ_2	Clearance rate for symptomatic non-hospitalized cases†	0.202 (0.168-0.300)	0.235 (0.195-0.348)	0.231 (0.192-0.342)	$\gamma_1 \times \frac{Recovery\ Time_{Mild}}{Recovery\ Time_{Mod}}$	[33,34]
γ_3	Clearance rate for symptomatic hospitalized cases†	0.145 (0.120-0.215)	0.168 (0.140-0.250)	0.165 (0.137-0.246)	$\gamma_1 \times \frac{Recovery\ Time_{Mild}}{Recovery\ Time_{Severe}}$	[33,34]
φ	Hospitalization Rate‡	0.202 (0.167-0.237)	0.361 (0.286-0.435)	0.471 (0.308-0.634)	$\frac{Number\ of\ patients\ hospitalized}{Number\ of\ symptomatic\ cases}$	[40,41]
μ	Incubation rate	0.202 (0.101-0.364)	0.202 (0.101-0.364)	0.202 (0.101-0.364)	$\frac{1}{Incubation\ Time}$	[29,30]

† Age groups are: 0-59, 60-69, 70+; ‡ Age groups are: 0-59, 60-69, 70+; 0-64, 65-74, 65+
* represents standard deviation of a normal distribution

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

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

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, γ_1, γ_2, and γ_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, β, 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 β 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 (θ) priors and their relative relationship to the known biology. In addition to

the unconstrained prior of θ, we refitted the data twice using two different prior assumptions of the symptomatic fraction: θ is less than 50%, and θ is less than 10% with the results located in S1 Appendix. In our parameter estimation, we assumed that transmission from our unobserved group was lower by a factor of α than transmission from the symptomatic observed group, β, as mild/asymptomatic cases are likely to shed fewer viruses and thus be less transmissible. Thus, we constrained α between 0·01-0·99 and β between 0·01-2·00. As the process is stochastic, we evaluated the median and credible intervals of the posterior distribution obtained from this procedure to guide our understanding of the parameters and to assess the ability of the model to fit actual data with biologically plausible values.

Results

Assuming a uniform prior with no constraint on the fraction of the population that was symptomatic, θ, the parameter estimation from our MCMC sampling resulted in a posterior median for the symptomatic rate of 0·04 (95% CrI, 0·01-0·41), 0·03 (95% CrI, 0·01-0·32), 0·18 (95% CrI, 0·02-0·85) for Italy, Spain, and South Korea, respectively, while the symptomatic rates of metropolitan areas of New York City and Chicago were 0·25 (95% CrI, 0·01-0·89) and 0·03 (95% CrI, 0·01-0·36), respectively. The median posterior values of these parameter ranges 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% 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, South Korea, New York City, and Chicago, respectively (Table 2 and Figure 2).

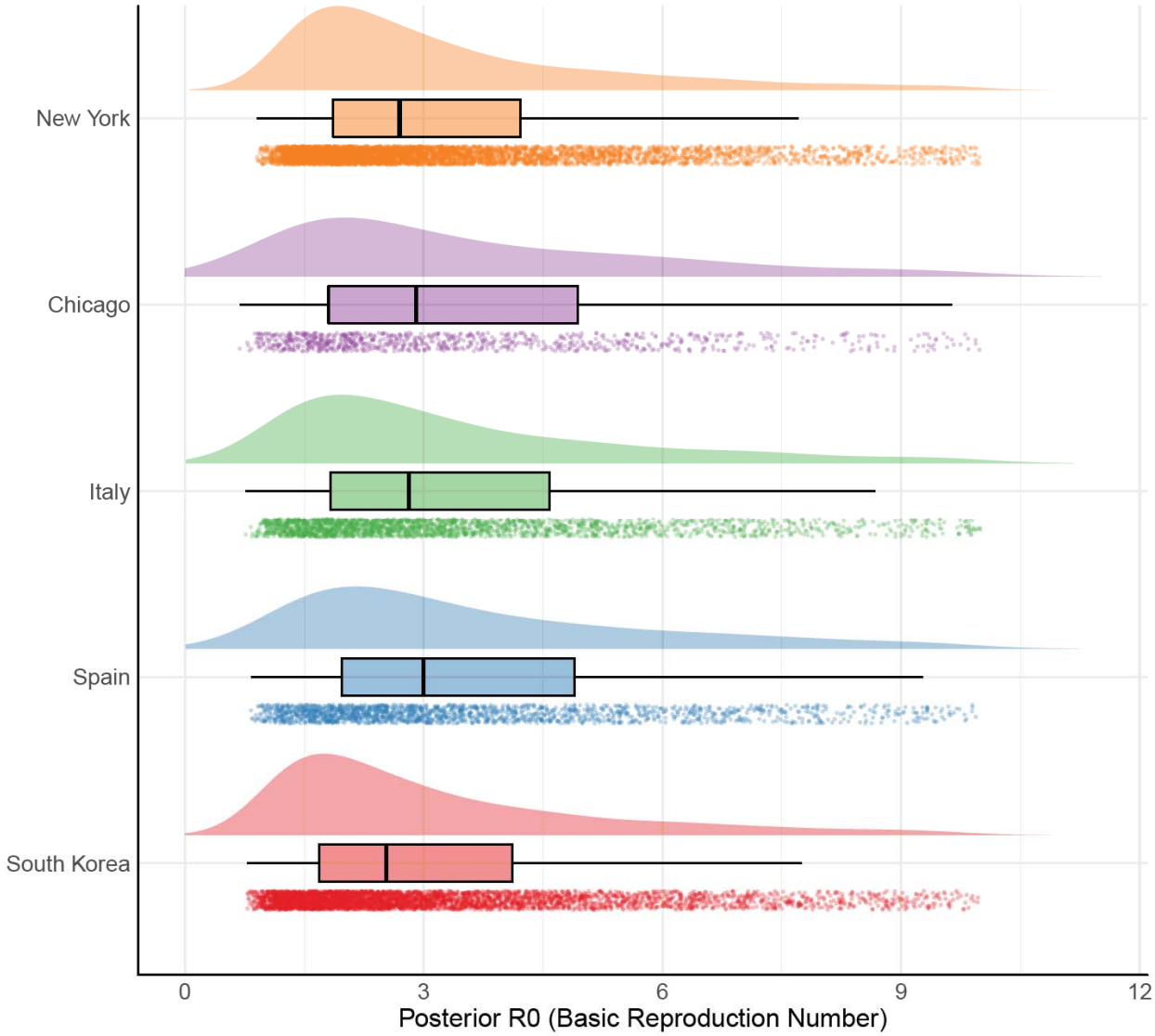

Figure 2. Estimated Basic Reproduction Number R_0 for New York, Chicago, Italy, Spain, and South Korea from Bayesian distribution. Each curve is the posterior distribution generated from the MCMC sampling for the R_0 of each location are shown. For each location, the raw data, box and whisker plot (median, interquartile ranges, 95% 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	--	**3.25 (1.09-29.77)**	**3.62 (1.13-34.89)**	**2.75 (1.04-22.44)**	**3.31 (1.69-20.55)**	**3.46 (1.01-34.41)**
θ	**U(0.01,0.99)**[†]	**0.04 (0.01-0.41)**	**0.03 (0.01-0.32)**	**0.18 (0.02-0.85)**	**0.25 (0.01-0.89)**	**0.03 (0.01-0.36)**
α	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)
β	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)
γ_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)
γ_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)
γ_3	U(0.01,0.30)	--	--	--	0.46 (0.03-0.95)	--
φ	U(0.01,0.50)	--	--	--	0.34 (0.13-0.49)	--
μ	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^{-1})	--	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)

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

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

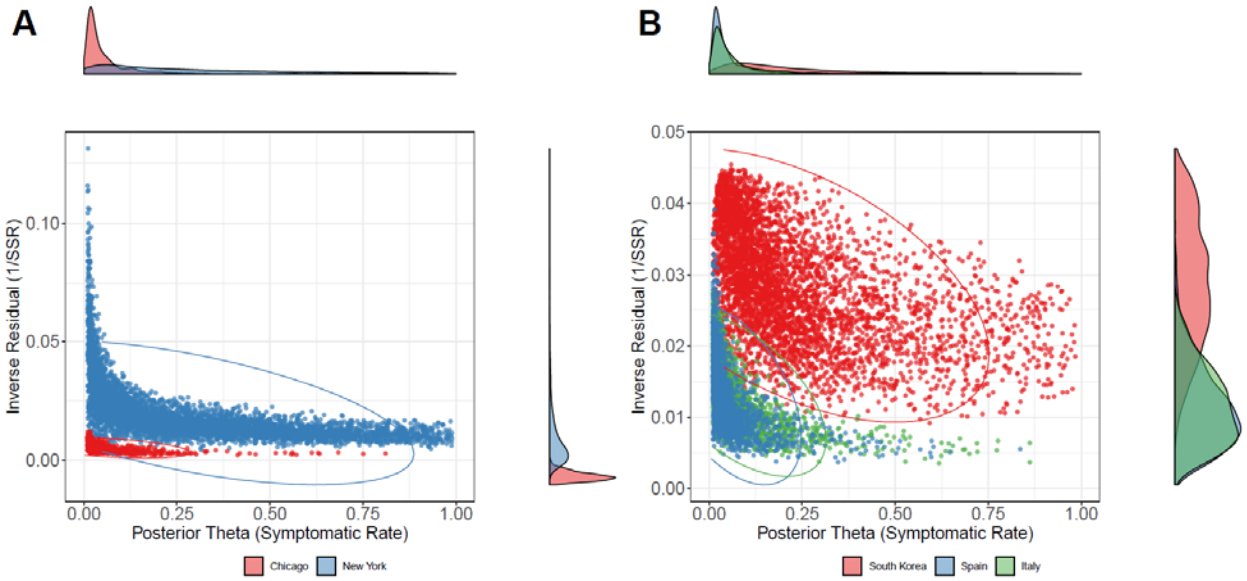

Figure 3. Scatterplot of inverse residual (goodness of fit) versus posterior sample of symptomatic rate. The posterior distribution of symptomatic rate θ and the inverse of sum of squared residuals (SSR^{-1}) for Chicago and New York City (A) and the same set of priors for Italy, Spain, and South Korea (B). The marginal posterior density of θ 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.

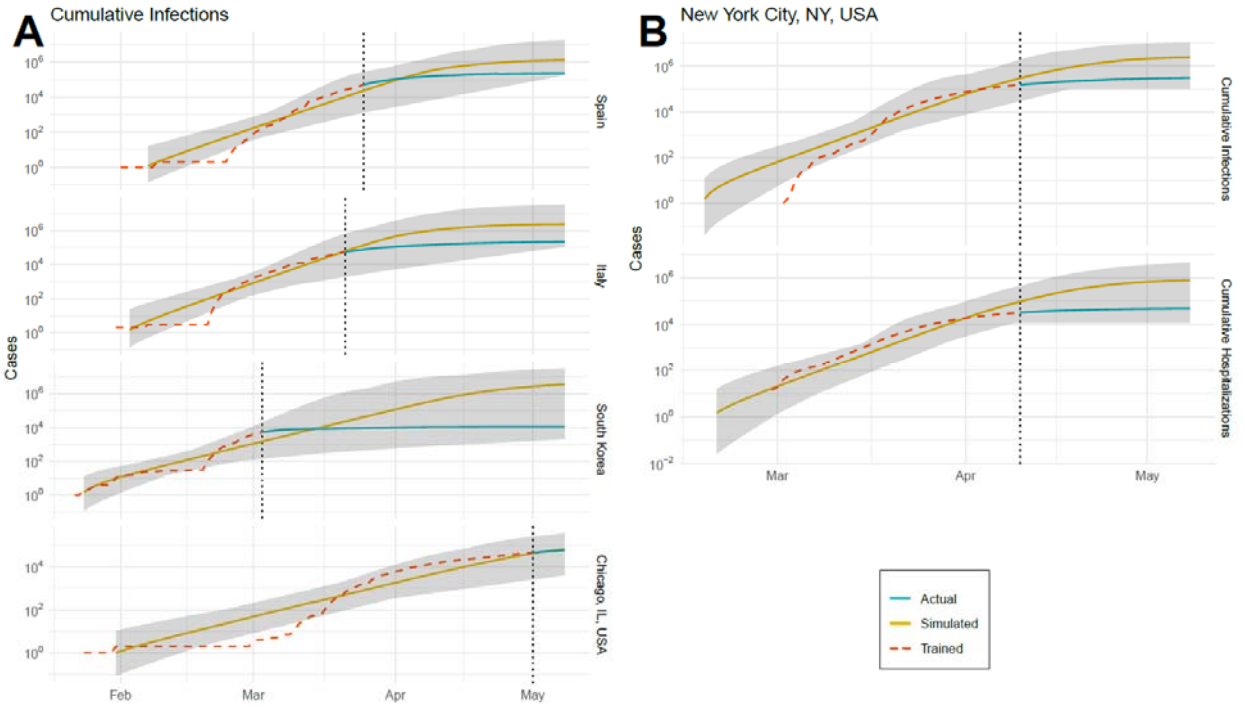

Figure 4. Simulated versus actual cumulative infections and hospitalization. (A) The simulated and actual cumulative infections for each location (Spain, Italy, South Korea, Chicago, and New York City are shown) for three different prior distributions of the symptomatic rate (θ). The turquoise line (simulated) 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.

Discussion

With no end in sight to the ongoing COVID-19 pandemic, understanding the rate of disease spread and the extent of people infected is vital to designing effective containment policies. This analysis contributes to available disease models, which play an integral role in defining the policy choices of government officials, by providing accurate evidence that conforms to observed data. Given the growth rate in infection cases in the exponential phase and assuming the force of infection for undetected cases is lower, our results for all scenarios and countries suggest that the undetected population is significantly larger than the symptomatic population, the latter of which accounts for only 3-25% of the total infected population. The value of R_0 varies from 2·8-3·6, which corresponds with a growing consensus that R_0 is approximately 3·1 (Table 1). The variation in symptomatic cases is almost certainly a function of testing, which has been quite variable among cities and countries, yet even in South Korea, which had extremely high rates of testing during the beginning of the pandemic, our models still suggest large fractions of cases were likely missed due to asymptomatic/mild infections. Nevertheless, higher testing rates in South Korea also led to a slower transmission of SARS-CoV-2. This is also true when you compare testing in New York City versus Chicago, where testing per capita during the exponential phase was higher in New York City.

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

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

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

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

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

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;0. doi:10.1016/S1473-3099(20)30120-1

2. Majumder MS, Mandl KD. Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility. Lancet Glob Health. 2020;0. doi:https://doi.org/10.1016/S2214-109X(20)30113-3

3. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-Based Analysis, Modelling and Forecasting of the COVID-19 outbreak. medRxiv. 2020; 2020.02.11.20022186. doi:https://doi.org/10.1101/2020.02.11.20022186

4. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med. 2020;27. doi:https://doi.org/10.1093/jtm/taaa021

5. Distante C, Piscitelli P, Miani A. Covid-19 Outbreak Progression in Italian Regions: Approaching the Peak by March 29th. medRxiv. 2020; 2020.03.30.20043612. doi:https://doi.org/10.1101/2020.03.30.20043612

6. Park M, Cook AR, Lim JT, Sun Y, Dickens BL. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. J Clin Med. 2020;9: 967. doi:10.3390/jcm9040967

7. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA. 2020 [cited 5 Apr 2020]. doi:https://doi.org/10.1001/jama.2020.3151

8. Tang B, Xia F, Bragazzi NL, Wang X, He S, Sun X, et al. Lessons drawn from China and South Korea for managing COVID-19 epidemic: insights from a comparative modeling study. medRxiv. 2020; 2020.03.09.20033464. doi:https://doi.org/10.1101/2020.03.09.20033464

9. Balilla J. Assessment of COVID-19 Mass Testing: The Case of South Korea. SSRN. 2020 [cited 5 Apr 2020]. doi:http://dx.doi.org/10.2139/ssrn.3556346

10. Cowling BJ, Ali ST, Ng TWY, Tsang TK, Li JCM, Fong MW, et al. Impact assessment of non-pharmaceutical interventions against COVID-19 and influenza in Hong Kong: an observational study. medRxiv. 2020; 2020.03.12.20034660. doi:https://doi.org/10.1101/2020.03.12.20034660

11. Tariq A, Lee Y, Roosa K, Blumberg S, Yan P, Ma S, et al. Real-time monitoring the transmission potential of COVID-19 in Singapore, February 2020. medRxiv. 2020; 2020.02.21.20026435. doi:https://doi.org/10.1101/2020.02.21.20026435

12. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020;368: 395–400. doi:10.1126/science.aba9757

13. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID‐19: a rapid review. Cochrane Database Syst Rev. 2020 [cited 18 Jun 2020]. doi:10.1002/14651858.CD013574

14. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the Asymptomatic Proportion of 2019 Novel Coronavirus onboard the Princess Cruises Ship, 2020. medRxiv. 2020; 2020.02.20.20025866. doi:https://doi.org/10.1101/2020.02.20.20025866

15. Cheng H-Y, Jian S-W, Liu D-P, Ng T-C, Huang W-T, Team TC-19 outbreak investigation, et al. High transmissibility of COVID-19 near symptom onset. medRxiv. 2020; 2020.03.18.20034561. doi:10.1101/2020.03.18.20034561

16. Tao Y, Cheng P, Chen W, Wan P, Chen Y, Yuan G, et al. High incidence of asymptomatic SARS-CoV-2 infection, Chongqing, China. medRxiv. 2020; 2020.03.16.20037259. doi:https://doi.org/10.1101/2020.03.16.20037259

17. Bai Y, Yao L, Wei T, Tian F, Jin D-Y, Chen L, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA. 2020 [cited 3 Apr 2020]. doi:https://doi.org/10.1001/jama.2020.2565

18. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med. 2020;382: 970–971. doi:https://doi.org/10.1056/NEJMc2001468

19. Liu Y-C, Liao C-H, Chang C-F, Chou C-C, Lin Y-R. A Locally Transmitted Case of SARS-CoV-2 Infection in Taiwan. N Engl J Med. 2020;382: 1070–1072. doi:https://doi.org/10.1056/NEJMc2001573

20. CNN TJ. Iceland lab's testing suggests 50% of coronavirus cases have no symptoms. In: CNN [Internet]. [cited 3 Apr 2020]. Available: https://www.cnn.com/2020/04/01/europe/iceland-testing-coronavirus-intl/index.html

21. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection. Ann Intern Med. 2020 [cited 23 Jun 2020]. doi:10.7326/M20-3012

22. Bedford T. Cryptic transmission of novel coronavirus revealed by genomic epidemiology. In: Cryptic transmission of novel coronavirus revealed by genomic epidemiology [Internet]. 2 Mar 2020 [cited 3 Apr 2020]. Available: https://bedford.io/blog/ncov-cryptic-transmission/

23. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From

the Chinese Center for Disease Control and Prevention. JAMA. 2020 [cited 3 Apr 2020]. doi:https://doi.org/10.1001/jama.2020.2648

24. Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiological Characteristics of 2143 Pediatric Patients With 2019 Coronavirus Disease in China. Pediatrics. 2020 [cited 5 Apr 2020]. doi:https://doi.org/10.1542/peds.2020-0702

25. Kelvin AA, Halperin S. COVID-19 in children: the link in the transmission chain. Lancet Infect Dis. 2020;0. doi:https://doi.org/10.1016/S1473-3099(20)30236-X

26. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N Engl J Med. 2020;382: 1177–1179. doi:10.1056/NEJMc2001737

27. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics—I. Bull Math Biol. 1991;53: 33–55. doi:10.1007/BF02464423

28. COVID-19: Data - NYC Health. [cited 2 May 2020]. Available: https://www1.nyc.gov/site/doh/covid/covid-19-data.page

29. Liu T, Hu J, Xiao J, He G, Kang M, Rong Z, et al. Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China. Systems Biology; 2020 Jan. doi:10.1101/2020.01.25.919787

30. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020;172: 577–582. doi:10.7326/M20-0504

31. Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, et al. COVID-19 Antibody Seroprevalence in Santa Clara County, California. medRxiv. 2020; 2020.04.14.20062463. doi:10.1101/2020.04.14.20062463

32. Sutton D, Fuchs K, D'Alton M, Goffman D. Universal Screening for SARS-CoV-2 in Women Admitted for Delivery. New England Journal of Medicine. Massachusetts Medical Society; 2020. doi:10.1056/NEJMc2009316

33. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science. 2020;368: 860–868. doi:10.1126/science.abb5793

34. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis. 2020;0. doi:10.1016/S1473-3099(20)30287-5

35. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020;26: 672–675. doi:10.1038/s41591-020-0869-5

36. Over 100 Million in China's Northeast Face Renewed Lockdown. Bloomberg.com. 18 May 2020. Available: https://www.bloomberg.com/news/articles/2020-05-18/over-100-million-in-china-s-northeast-thrown-back-under-lockdown. Accessed 18 May 2020.

37. Long Q-X, Tang X-J, Shi Q-L, Li Q, Deng H-J, Yuan J, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat Med. 2020; 1–5. doi:10.1038/s41591-020-0965-6

38. Ogbunugafor CB, Miller-Dickson MD, Meszaros VA, Gomez LM, Murillo AL, Scarpino SV. The intensity of COVID-19 outbreaks is modulated by SARS-CoV-2 free-living survival and environmental transmission. medRxiv. 2020; 2020.05.04.20090092. doi:10.1101/2020.05.04.20090092

39. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. The Lancet. 2020;395: 689–697. doi:10.1016/S0140-6736(20)30260-9

40. CDC COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. Morb Mortal Wkly Rep. 2020;69: 343–346. doi:10.15585/mmwr.mm6912e2

41. Garg S. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020. MMWR Morb Mortal Wkly Rep. 2020;69: 458–464. doi:10.15585/mmwr.mm6915e30

Supporting information

S1 Appendix. Model Specifications, Deriving the Next Generation Matrix and Calculating R0, Bayesian Inference of Model Parameters, Population Inclusion for US Cities, and Supplement Plots