OPEN

Estimating the Size of a COVID-19 Epidemic from Surveillance Systems

Mu Yue, Hannah E. Clapham, and Alex R. Cook

Abstract: Public health policy makers in countries with Coronavirus Disease 2019 (COVID-19) outbreaks face the decision of when to switch from measures that seek to contain and eliminate the outbreak to those designed to mitigate its effects. Estimates of epidemic size are complicated by surveillance systems that cannot capture all cases, and by the need for timely estimates as the epidemic is ongoing. This article provides a Bayesian methodology to estimate outbreak size from one or more surveillance systems such as virologic testing of pneumonia cases or samples from a network of general practitioners.

Keywords: Bayesian inference; COVID-19; Epidemic size; Surveillance

(*Epidemiology* 2020;31: 567–569)

As the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1] spreads around the world from its initial focus in Wuhan, China,[2] causing local Coronavirus Disease 2019 (COVID-19) epidemics, public health policy makers in countries or territories face the decision of when to switch from containment to mitigation measures.[3] This decision rests upon an accurate estimate of the size of the local outbreak. Where intensive contact tracing has been undertaken, such as in Singapore,[4] or mass testing, such as South Korea, there may be some degree of confidence that most cases have been identified and thus that the order of magnitude of the outbreak is known. Otherwise, however, policy makers may be reliant on passive surveillance streams to infer the size of the outbreak. Such inference may be challenged by incompleteness in coverage and the rapid growth of the outbreak, which coupled with the lag between onset of symptoms and being detected by the surveillance system, requires statistical inflation to correct the estimates.

This note outlines a simple Bayesian model designed to estimate the outbreak size during the exponential growth phase of the COVID-19 epidemic from one or two surveillance streams providing counts of cases meeting various criteria. We illustrate it through scenarios based on virologic surveillance from a network of influenza-like illness consultations in primary care and from pneumonia cases in hospitals, but the approach generalizes to other surveillance streams such as mortalities.

METHODS

We assume that we remain in the initial phase of the epidemic when both the total and the new number of cases (regardless of whether they are imported or autochthonous) grows exponentially,[5] prior to herd immunity taking hold. Let the number of new cases on day t be $n_t = n_0 \exp bt$. We assume that growth has a constant exponent, as it might if control has been implemented to a constant degree. Time 0 is arbitrary but may be set to the day the alarm was first raised in Wuhan, which coincidentally was 31 December 2019, allowing t to represent the day of the year in 2020. Also let the number of new cases detected by surveillance stream s on day t be the Poisson variable, x_t^s. We assume that a fraction of cases p_s enter the surveillance system at an average lag of L_s after onset, which we assume does not change over time. For instance, a fraction of cases may develop pneumonia or may present to a primary care clinic that is part of a virologic surveillance network. The likelihood function obtains from $X_t^s \sim Po(n_{t-L_s} p_s)$. Altogether the parameter space is $2 + 2|\{s\}|$ dimensional, at the early phase of the outbreak, the scarcity of local data may necessitate fixing some parameters using knowledge obtained from elsewhere (say on the proportion of cases developing pneumonia) or from the nature of the surveillance system (say on the proportion of primary care clinics in the network). The target of inference is n_t and the total cases to date, $N_t = \sum_{i \le t} n_i$. These can be estimated by (1) setting noninformative prior distributions for the parameters we have information to estimate, informative or Dirac delta priors for those we have not, (2) running a standard Metropolis-Hastings algorithm,[6] and (3)

Submitted March 6, 2020; accepted April 2, 2020.

From the Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore.

This research is supported by the Singapore Ministry of Health's National Medical Research Council under the Centre Grant Programme - Singapore Population Health Improvement Centre (NMRC/CG/C026/2017_NUHS).

The authors report no conflicts of interest.

Code may be downloaded from https://github.com/yuemu1989/COVID-19-Outbreak-Size.

Correspondence: Alex R. Cook, Saw Swee Hock School of Public Health, 12 Science Drive 2 #10-01, Singapore 118177, Singapore. E-mail: alex.richard.cook@gmail.com.

ISSN: 1044-3983/20/3104-0567

DOI: 10.1097/EDE.0000000000001202

<zjs;Brief Report> • <zjss;XXX> • <zdoi;10.1097/EDE.0000000000001202>

transforming the primary estimates to obtain posterior distribution for n_t and N_t. We have developed example R code[7] to implement this algorithm which may be downloaded from https://github.com/yuemu1989/COVID-19-Outbreak-Size. We now illustrate the approach through two examples. The pattern of cases together with the dynamic estimates of the size of the outbreak for illustrative examples 1 and 2 are presented in the Table.

Example 1: The First Case of Pneumonia

In city X, all cases of pneumonia are being tested for SARS-CoV-2 infection. We assume that $L = 7$ and $p = 19\%$.[8] In this scenario, after 30 days of negative tests, a positive case is identified on day 31. The estimated number of infections is $N_{31} = 24$ (95% CrI = 2–93). This estimate will necessarily evolve over the next few days as more cases come in, or not. Should there be no new cases by day 35, the estimate would reduce to $N_{35} = 18$ (95% credible interval [CrI] = 2–64). Should the first case be followed by one more pneumonia a day for 4 days, the estimate would change to $N_{35} = 108$ (95% CrI = 26–311).

Example 2: A Smattering of Pneumonias and Influenza-like Illnesses

In country Y, two passive surveillance systems are used to detect COVID-19: all pneumonias are tested, and a network of primary care doctors take nasopharyngeal swabs from patients with influenza-like illness (ILI), which are tested virologically for SARS-CoV-2 in addition to influenza. The network covers approximately 4% of ILIs presenting to primary care. We assume that 43% of SARS-CoV-2 cases develop ILI[9] and consult an average of 5.5 days after onset,[10] i.e., $p_1 = 1.72\%$, $L_1 = 5.5$, as well as $p_2 = 19\%$ and $L_2 = 7$ as before.

DISCUSSION AND CONCLUSION

The method we outline can readily be implemented on a daily basis as new reports come in. It will be affected by delays in reporting, which should be accommodated through the lag parameter(s) or by revising previous estimates as cases are reported. In the early period of the outbreak, it may be necessary to use estimates of parameters such as the growth rate b from China[5] or the second wave of countries to be affected. As the estimates are dependent on the prior distribution assumed for these parameters, sensitivity analyses may be conducted to assess how robust the estimates are to misspecification of input parameters.[11] As the local outbreak continues, there may be sufficient information to permit localized parameterization and to use model predictive checks to assess whether its assumptions, for instance of exponential growth, are valid.

TABLE. Estimated Outbreak Size to Date Based on Pneumonias Surveillance System and Pneumonia and ILI Surveillance Systems

Pneumonia Surveillance System														
	Pneumonia Cases on Day													
Scenario	1	…	30	31	32	33	34	35	36	37	38	39	*t*	Estimated Size to *t*
1	0	…	0	1	-	-	-	-	-	-	-	-	31	24 (2–93)
2	0	…	0	1	0	0	0	0	-	-	-	-	35	18 (2–64)
3	0	…	0	1	1	1	1	1	-	-	-	-	35	108 (26–311)
4	0	…	0	1	0	1	1	2	2	3	-	-	37	314 (92–892)
5[a]	1	-	-	-	-	-	-	-	-	-	-	-	1	18 (3–43)
6[a]	-	-	-	1	-	-	-	-	-	-	-	-	31	204 (25–560)
Pneumonia and ILI Surveillance Systems (Counts: Pneumonia/ILI)														
	Pneumonia and ILI Cases on Day													
Scenario	1	…	30	31	32	33	34	35	36	37	38	39	*t*	Estimated Size to *t*
1	0/0	…	0/0	1/0	-	-	-	-	-	-	-	-	31	21 (2–77)
2	0/0	…	0/0	1/0	0/0	-	-	-	-	-	-	-	32	20 (2–74)
3	0/0	…	0/0	1/0	0/0	0/0	-	-	-	-	-	-	33	19 (2–65)
4	0/0	…	0/0	1/0	0/0	0/0	1/0	-	-	-	-	-	34	33 (5–112)
5	0/0	…	0/0	1/0	0/0	0/0	1/0	1/0	-	-	-	-	35	57 (10–210)
6	0/0	…	0/0	1/0	0/0	0/0	1/0	1/0	0/0	-	-	-	36	44 (9–129)
7	0/0	…	0/0	1/0	0/0	0/0	1/0	1/0	0/0	2/0	-	-	37	88 (22–252)
8	0/0	…	0/0	1/0	0/0	0/0	1/0	1/0	0/0	2/0	1/0	-	38	98 (29–251)
9	0/0	…	0/0	1/0	0/0	0/0	1/0	1/0	0/0	2/0	1/0	2/1	39	180 (62–424)

Estimated sizes are posterior mean and 95% credible intervals. - indicates no data are available on that day.
[a]Uses fixed exponential growth rate $b = 0.11$ based on Reference 5.

As countries repatriate their citizens from areas of heightened transmission, the growth in allochthonous and autochthonous infections may diverge, and thus the exclusion of the former may be warranted when using this method.

REFERENCES

1. Gorbalenya AE, Baker SC, Baric RS, et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol*. 2020;5:536–544.
2. Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N Engl J Med*. 2020;382:1199–1207.
3. Heymann DL, Shindo N; WHO Scientific and Technical Advisory Group for Infectious Hazards. COVID-19: what is next for public health? *Lancet*. 2020;395:542–545.
4. Wong JE, Leo YS, Tan CC. COVID-19 in Singapore—current experience: critical global issues that require attention and action. *JAMA* 2020; 323:1243–1244.
5. 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. *Lancet*. 2020;395:689–697.
6. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. Boca Raton, FL: CRC Press; 2013.
7. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2014.
8. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) — China, 2020. *China CDC Wkly*. 2020;2:113–122.
9. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*. 2020. doi: 10.1056/NEJMoa2002032
10. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Published February 2020. Available at: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf. Accessed 6 March 2020.
11. Lee VJ, Chen MI, Yap J, et al. Comparability of different methods for estimating influenza infection rates over a single epidemic wave. *Am J Epidemiol*. 2011;174:468–478.