Evaluating uncertainties associated with early estimates of the basic reproduction number during the novel coronavirus (2019-nCoV) outbreak

Sang Woo Park[1] Jonathan Dushoff[2,3]

1 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, USA
2 Department of Biology, McMaster University, Hamilton, Ontario, Canada
3 Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada

Abstract

1 Introduction

Since the end of December 2019, a novel coronavirus (2019-nCoV) continues to spread in China and in other parts of the world. As of January 27th, 2020, the World Health Organization has confirmed 2798 cases, including 37 confirmed cases in 11 different countries, outside China (World Health Organization (WHO), 2020b). Although the virus is suspected to have originated from animal reservoirs, a recent case from Viet Nam demonstrated its ability to transmit between humans (World Health Organization (WHO), 2020a), posing a greater threat for a wider spread.

Many researchers have been rushing to publish their analysis of the outbreak (Imai et al., 2020; Riou and Althaus, 2020; Read et al., 2020; Zhao et al., 2020; Majumder and Mandl, 2020; Liu et al., 2020) and, in particular, their estimates of the basic reproductive number $\mathcal{R}_0$ (i.e., the average number of secondary cases generated by a primary case in a fully susceptible population). The basic reproductive number is of particular interest because it allows prediction about the final size of an epidemic. While their efforts are valuable, their analyses rely on several assumptions that could immediately affect their estimates of $\mathcal{R}_0$ and the associated uncertainties.

Here, we present a simple framework for evaluating uncertainties associated with parameter estimates across a wide range of models. Our results indicate that most published estimates of $\mathcal{R}_0$ are likely to be overly confident. We also lay down several principles that needs to be taken into consideration.

2 Methods and results

Early in an outbreak, $\mathcal{R}_0$ cannot be estimated directly. Instead, estimates of $\mathcal{R}_0$ are often inferred from the exponential growth rate r, which can be estimated reliably from incidence data. Given estimates of the exponential growth rate r and the distribution $g(\tau)$ of generation intervals (i.e., the time between when a person become infected and that person infects another person), the basic reproduction number can be estimated via the Euler-Lotka equation:

$$1/\mathcal{R}_0 = \int \exp(-r\tau)g(\tau)d\tau. \tag{1}$$

In other words, estimates of $\mathcal{R}_0$ will depend on the assumptions about the exponential growth rate r and the shape of generation-interval distribution $g(\tau)$.

We use the gamma approximation framework to first characterize the amount of uncertainty present in the underlying parameters and then to assess the degree to which these uncertainties affect the estimate of $\mathcal{R}_0$. Assuming that generation intervals follow a gamma distribution with the mean $\bar{G}$ and the squared coefficient of variation κ, we have

$$\mathcal{R}_0 = \left(1 + \kappa r \bar{G}\right)^{1/\kappa}. \tag{2}$$

This equation demonstrates that a generation-interval distribution that assumes a larger mean (higher $\bar{G}$) or is less variable (lower κ) will give a higher estimate of $\mathcal{R}_0$ for the same value of r.

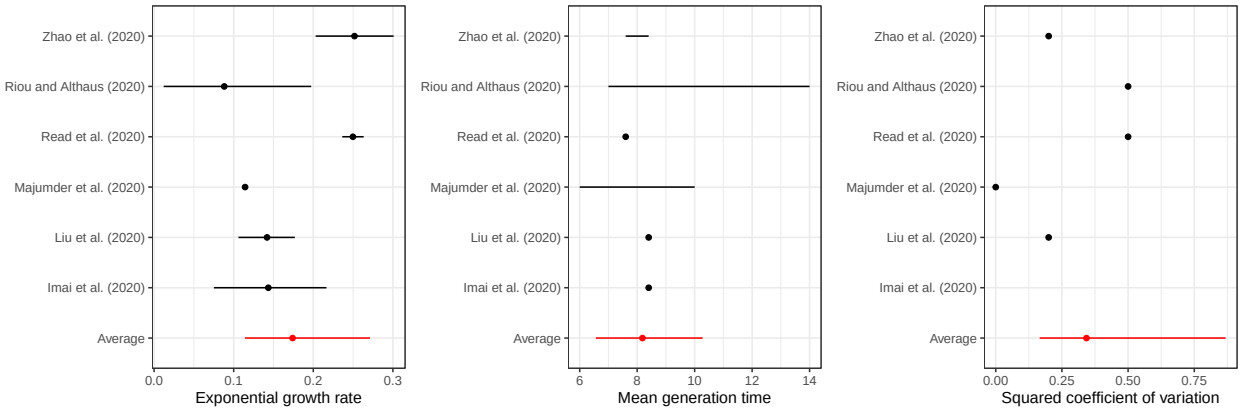

Figure 1: Early estimates of $\mathcal{R}_0$ and associated assumptions about r and $g(\tau)$.

First, we gather information on estimates of $\mathcal{R}_0$ and their assumptions about the underlying generation-interval distributions from preprint servers (e.g., bioRxiv, medRxiv, and SSRN). For each study, we model $\mathcal{R}_0$, $\bar{G}$, and κ with an appropriate probability distribution (either uniform or gamma) that roughly matches the estimated or assumed values and uncertainty ranges. For example, Imai et al. (2020) estimated $\mathcal{R}_0 = 2.6$ (95% CI: 1.5–3.5); we model this parameter as a gamma distribution with mean of 2.6 and a shape parameter of 28, which has a 95.2% probability of containing a value between 1.5 and 3.5. Then, we draw 100,000 random samples from these distributions and calculate the exponential growth rate r via the inverse of (2):

$$r = \frac{\mathcal{R}_0^{\kappa} - 1}{\kappa \bar{G}} \tag{3}$$

To define ranges of appropriate uncertainties for each parameter (r, $\bar{G}$, and κ), we use a Bayesian multilevel modeling approach by assuming that all *observed* (i.e., estimated or assumed) parameter values come from the same distribution:

$$\begin{aligned} r_i &\sim \text{Gamma}(\text{mean} = \mu_r, \text{shape} = \alpha_r), \\ \bar{G}_i &\sim \text{Gamma}(\text{mean} = \mu_G, \text{shape} = \alpha_G), \\ \kappa_i &\sim \text{Exponential}(\text{mean} = \mu_\kappa), \\ (\mu_r, \alpha_r, \mu_G, \alpha_G, \mu_\kappa) &\sim \text{Gamma}(\text{shape} = 0.1, \text{rate} = 0.1), \end{aligned} \tag{4}$$

where i represents the index across 6 studies that we consider. We use an exponential distribution to model κ_i to account for the fact that the IDEA model, which Majumder and Mandl (2020) uses, assumes that the generation-interval distribution is fixed ($\kappa = 0$). Hyperparameters μ_r, α_r, μ_G, α_G, and μ_κ are estimated with a Markov Chain Monte Carlo (MCMC) by drawing a random set for r_i, $\bar{G}_i$, and κ_i at each MCMC step from the 100,000 samples that we calculated previously. We run 4 parallel chains that consist of 100,000 burnin steps and 100,000 sampling steps. Posterior samples are thinned every 200 steps. Convergence is assessed by ensuring that the Gelman-Rubin statistic is below 1.01 for all parameters (Gelman et al., 1992).

References

Gelman, A., D. B. Rubin, et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science* 7(4), 457–472.

Imai, N., A. Cori, I. Dorigatti, M. Baguelin, C. A. Donelly, S. Riley, and N. M. Ferguson (2020). Report 3: Transmissibility of 2019-nCoV. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-2019-nCoV-transmissibility.pdf.

Liu, T., J. Hu, M. Kang, L. Lin, H. Zhong, J. Xiao, G. He, T. Song, Q. Huang, Z. Rong, A. Deng, W. Zeng, X. Tan, S. Zeng, Z. Zhu, J. Li, D. Wan, J. Lu, H. Deng, J. He, and W. Ma (2020). Transmission dynamics of 2019 novel coronavirus (2019-nCoV). https://www.biorxiv.org/content/10.1101/2020.01.25.919787v1.

Majumder, M. and K. D. Mandl (2020). Early transmissibility assessment of a novel coronavirus in Wuhan, China. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3524675.

Read, J. M., J. R. Bridgen, D. A. Cummings, A. Ho, and C. P. Jewell (2020). Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. https://www.medrxiv.org/content/10.1101/2020.01.23.20018549v1.

Riou, J. and C. L. Althaus (2020). Pattern of early human-to-human transmission of wuhan 2019-nCoV. https://www.biorxiv.org/content/10.1101/2020.01.23.917351v1.

World Health Organization (WHO) (2020a). Novel Coronavirus (2019-nCoV) Situation Report - 6. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200126-sitrep-6-2019–ncov.pdf?sfvrsn=beaeee0c_4.

World Health Organization (WHO) (2020b). Novel Coronavirus (2019-nCoV) Situation Report - 7. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200127-sitrep-7-2019–ncov.pdf?sfvrsn=98ef79f5_2.

Zhao, S., J. Ran, S. S. Musa, G. Yang, Y. Lou, D. Gao, L. Yang, and D. He (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. https://www.biorxiv.org/content/10.1101/2020.01.23.916395v1.