Estimating effective reproduction number using generation time versus serial interval, with application to COVID-19 in GTA Canada

Jesse Knight[1] and Sharmistha Mishra[1,2,3,4]

[1]Institute of Medical Science, University of Toronto
[2]MAP Centre for Urban Health Solutions, Unity Health Toronto
[3]Department of Medicine, Division of Infectious Disease, University of Toronto
[4]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto

May 8, 2020

Abstract

TODO

1 Introduction

The effective reproduction number $R_e(t)$ provides an instantaneous measure of potential for epidemic growth, after considering herd immunity and interventions. Cori et al. [1] provide a method to estimate $R_e(t)$ in quasi-real time based on only two inputs: an incidence time series,[1] and the generation time distribution. The generation time is defined as the time between infection events in an infector-infectee pair.

When estimating $R_e(t)$ in previous epidemics [1–3], and in COVID-19 [4–7], the generation time has been approximated by the serial interval, defined as the time between symptom onset in an infector-infectee pair, since only the latter is directly observable. This approximation is reasonable for infectious diseases where onset of infectiousness and symptoms is simultaneous [1]. However, potential pre-symptomatic transmission of COVID-19 [8, 9] renders approximation of generation time by serial interval problematic. Namely, while generation time is strictly positive, the serial interval can be negative due to variability in the incubation period, such as in 59 of 468 (12.6%) reported cases in [9]. As a result, there have been differences in whether or not non-negative distributions are fit to COVID-19 serial interval data [9–12].

We aimed to show that it might be unnecessary to fit non-negative distributions to serial interval data for the purpose of calculating $R_e(t)$, since a non-negative generation time distribution can be inferred based on the incubation period and serial interval distributions. We also explored the implications for the estimated $R_e(t)$ of using non-negative serial interval distributions versus an inferred generation time distribution.

2 Methods

First, we designed a simple method to recover the generation time distribution from parametric definitions of the incubation period and the serial interval distributions. Then, we applied this method to COVID-19 using distributions reported in the literature. Finally, we compared $R_e(t)$ estimated for the Greater Toronto Area (GTA) region of Canada between March 08 and April 15, 2020 using the estimated generation time distribution versus serial interval distributions reported in the literature.

[1] Incidence may be further stratified by imported versus locally generated cases to quantify local transmission dynamics.

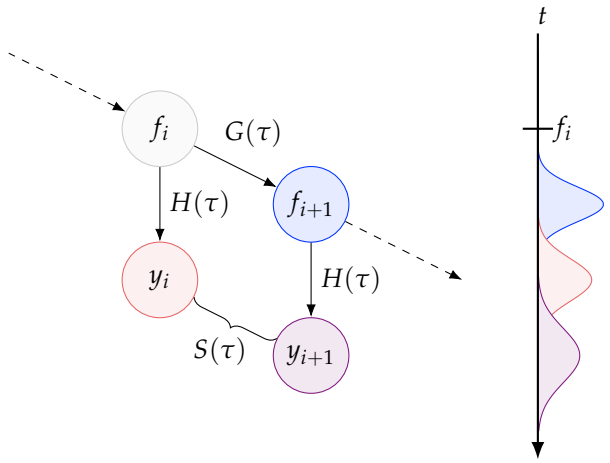

Figure 1: Random variables involved in the serial interval

Notation — i: infector index; $i+1$: infectee index; f_i: time of infection; y_i: time of symptom onset; $G(\tau)$: generation time distribution; $H(\tau)$: incubation time distribution; $S(\tau)$: serial interval distribution.

2.1 Estimating the Generation Time Distribution

Let i and $i+1$ be the indices of an infector-infectee pair. Let f_i and f_{i+1} be the respective times of infection, such that $g_i = [f_{i+1} - f_i] \sim G(\tau)$ is the generation time. Let y_i be the time of symptom onset in case i, such that $h_i = [y_i - f_i] \sim H(\tau)$ is the incubation period. Finally, let $s_i = [y_{i+1} - y_i] \sim S(\tau)$ be the serial interval. Figure 1 illustrates the variables and distributions graphically. We assume all distributions are independent, although previous work has shown that the generation time and incubation period may be correlated, for example in the context of measles [13].

Rearranging, we have: $s_i = g_i + h_{i+1} - h_i$. The probability distribution of the sum of independent random variables is the convolution of their respective distributions [14], where convolution $*$ is defined as:

$$[H * F](\tau) = \int_{-\infty}^{+\infty} H(z)F(\tau - z)\, dz \tag{1}$$

Thus $S(\tau)$ can be defined as:

$$S(\tau) = G(\tau) * H(\tau) * H(-\tau) \tag{2}$$

and $G(\tau)$ can be recovered using deconvolution $*^{-1}$ as:

$$G(\tau) = \left[S(\tau) *^{-1} H(\tau)\right] *^{-1} H(-\tau) \tag{3}$$

Some definitions of $S(\tau)$ and $H(\tau)$ may yield forms of $G(\tau)$ via deconvolution in (3) which are implausible or intractable. So, we defined a parametric form $\hat{G}(\tau \mid \theta)$, and found parameters θ^* that minimized the Kullback-Leibler divergence between the observed $S(\tau)$ and $\hat{S}(\tau \mid \theta)$ obtained via (2) using $\hat{G}(\tau \mid \theta^*)$. It can be shown that such parameters θ^* provide the maximum likelihood estimate (MLE) of $S(\tau)$ under $\hat{G}(\tau \mid \theta)$.

2.2 Application

2.2.1 Generation Time

We identified several parameterizations of the COVID-19 incubation period and serial interval following the review by Park et al. [15] (Table A.1).[2] For our analysis, we used the negative-permitting serial interval from [9] ($N = 468$):

$$S(\tau) = \text{Norm}\,(\mu = 3.96,\ \sigma = 4.75) \tag{4}$$

and the incubation period from [16] ($N = 181$):

$$H(\tau) = \text{Gam}\,(\alpha = 5.807,\ \beta = 0.948) \tag{5}$$

We assumed a Gamma parametric form for the generation time distribution $\hat{G}(\tau \mid \theta)$, with $\theta = [\alpha\,(\text{shape}), \beta\,(\text{scale})]$, for consistency with downstream assumptions used in calculating $R_e(t)$. We then minimized the Kullback-Leibler divergence between $S(\tau)$ and $\hat{S}(\tau \mid \theta)$ using the Nelder-Mead optimization method in the `optimization` R package [3] to obtain the MLE generation time distribution parameters θ^*.

2.2.2 Effective Reproduction Number

In the model described by Cori et al. [1], the incidence I at time t is given by the integral over all previous infections, multiplied by their respective infectivity ω at time τ since infection, collectively multiplied by the effective reproductive number R_e at time t. This model yields the following definition of $R_e(t)$:

$$R_e(t) = I(t)\left[\int_0^t I(t-\tau)\omega(\tau)d\tau\right]^{-1} \tag{6}$$

The infectivity profile $\omega(\tau)$ is equivalent to the generation time distribution $G(\tau)$. Given $I(t)$ and $G(\tau)$, probabilistic estimates of $R_e(t)$ can then be resolved in a Bayesian framework, as implemented in the `EpiEstim` R package.[4]

[2] Some studies did not provide enough information to define a parametric form (e.g. only reported the mean).

[3] https://cran.r-project.org/web/packages/optimization

[4] https://cran.r-project.org/web/packages/EpiEstim

In order to quantify $R_e(t)$ of COVID-19 in GTA, Canada, we used reported cases in the region between March 08 and April 15 2020 as the incidence time series $I(t)$.[5] We smoothed $I(t)$ using a Gaussian kernel with $\sigma = 1$ day to reflect uncertainty in reporting delay. We then compared estimates of $R_e(t)$ using the MLE generation time distribution versus serial interval distributions reported in the literature, including negative-permitting (Normal [9]), and non-negative (Gamma [10], Log-Normal [11]) distributions.

3 Results

Figure 2 shows the serial interval and incubation period distributions from [9] and [16] respectively, and the generation time distribution estimated via the proposed method. The MLE parameters of $\hat{G}(\tau \mid \theta)$ were: shape $\alpha = 1.813$ and scale $\beta = 2.199$, yielding $\hat{S}(\tau \mid \theta^*)$ that closely approximated the target $S(\tau)$.

The estimated mean generation time of 3.99 was close to the mean serial interval of 3.96 based on the negative-permitting distribution [9], but shorter than mean serial interval based on non-negative distributions, such as 5.12 in [10] and 4.7 in [11]. The SD of the generation time distribution was smaller at 2.96 than the SD of the serial interval at 4.75. Overall, estimated mean and SD generation time were similar to those reported by Ganyani et al. [12] (Table A.1).

Figure 3 shows $R_e(t)$ for COVID-19 in GTA, Canada based on reported cases and estimated using the generation time distribution versus selected serial interval distributions reported in the literature.[6] The $R_e(t)$ using the generation time distribution estimated here was lower compared to using non-negative serial interval distributions. Conversely, $R_e(t)$ estimated using a negative-permitting serial interval distribution was the smallest.

4 Discussion

We have demonstrated how to estimate a non-negative generation time distribution based on negative serial interval and non-negative incubation period distributions. Using reported parametric COVID-19 distributions, we estimated a Gamma-distributed generation time with a mean of 4 days and 95% of transmissions occurring before 10 days. When approximating the generation time distribution by non-negative serial interval distribution, 95% of transmissions are still expected within 10 days [10, 11], but the proportion of pre-

[5] We did not stratify incidence by "imported" vs "local" transmission events, in order to quantify overall epidemic growth.

[6] Generation time and serial interval distributions are also illustrated in Figure A.1.

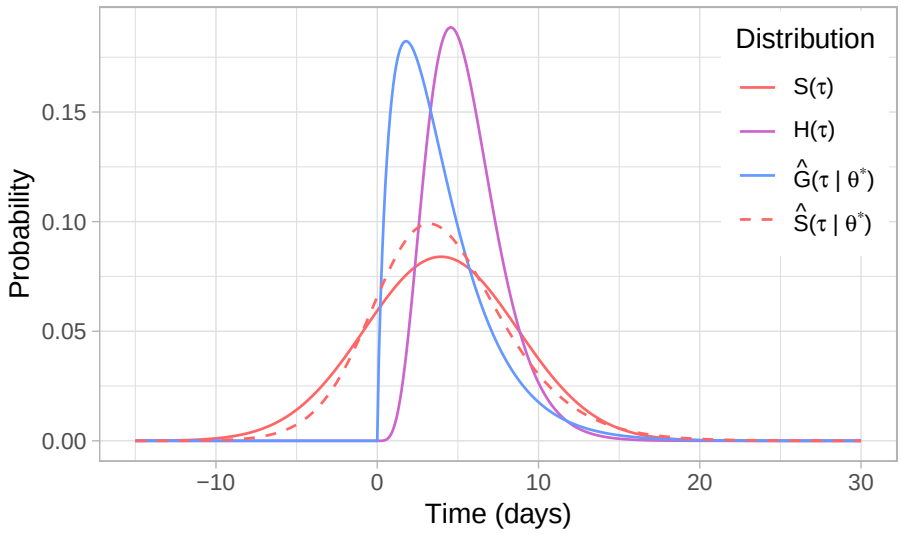

Figure 2: Recovery of the generation time distribution $G(\tau)$ from the serial interval $S(\tau)$ and incubation period $H(\tau)$ distributions

symptomatic infections may be underestimated due to higher mean serial interval versus generation time (Figure A.1) [12, 19].

Estimated effective reproduction number $R_e(t)$ based on the generation time distribution was lower than $R_e(t)$ based on non-negative serial interval distributions reported in the literature, suggesting that the latter may overestimate the infection transmission potential. By contrast, estimation of $R_e(t)$ using negative permitting serial intervals, such as that reported by Du et al. [9], may result in underestimated infection transmission potential.

Our analysis has three major limitations. First, we assumed that generation time and incubation period were independent, although Klinkenberg et al. [13] showed that this is false at least in measles. Second, we did not leverage line-listed (subject-level) data to estimate the generation time distribution, such as in [12] and [13]. Rather, we focused on parametric distributions, which could, for example, be obtained by meta-analysis. Finally, we did not propagate uncertainty in the reported serial interval and incubation period distributions through our results to obtain confidence intervals for the generation time parameterization. To model distribution uncertainty, future work could examine joint estimation of the generation time, serial interval, and $R_e(t)$ within the same Bayesian framework as described by [1].

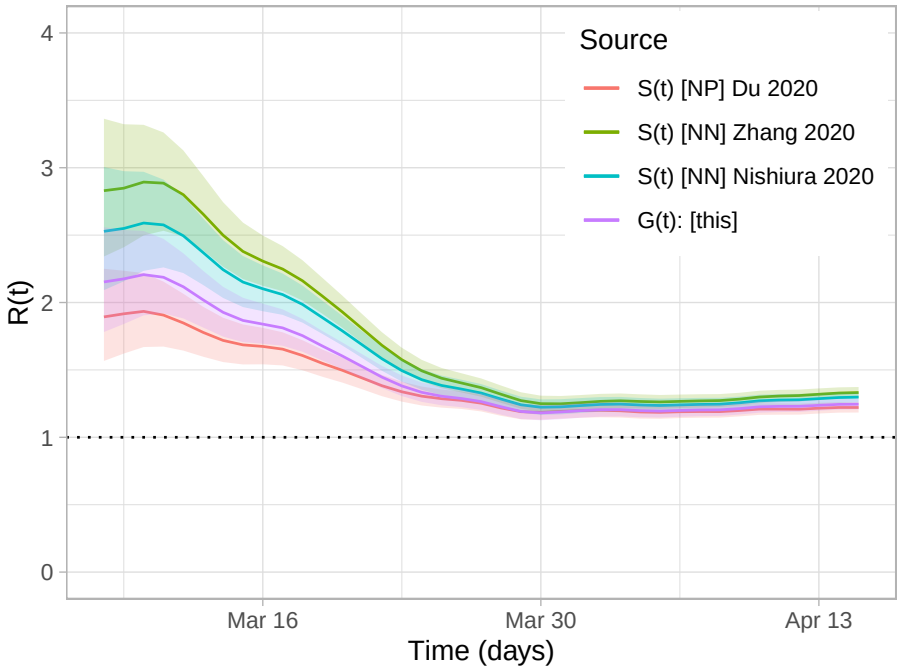

Figure 3: $R_e(t)$ of COVID-19 in GTA using serial interval versus generation time

Notation — $S(\tau)$: serial interval; $G(\tau)$: generation time; [NP]: negative-permitting; [NN]: non-negative.

References

[1] Anne Cori et al. "A new framework and software to estimate time-varying reproduction numbers during epidemics". In: *American Journal of Epidemiology* 178.9 (2013), pp. 1505–1512. DOI: 10.1093/aje/kwt133.

[2] Sheikh Taslim Ali, A. S. Kadi, and Neil M. Ferguson. "Transmission dynamics of the 2009 influenza A (H1N1) pandemic in India: The impact of holiday-related school closure". In: *Epidemics* 5.4 (2013), pp. 157–163. DOI: 10.1016/j.epidem.2013.08.001.

[3] Bruce Aylward et al. "Ebola virus disease in West Africa - The first 9 months of the epidemic and forward projections". In: *New England Journal of Medicine* 371.16 (2014), pp. 1481–1495. DOI: 10.1056/NEJMoa1411100.

[4] An Pan et al. "Association of Public Health Interventions with the Epidemiology of the COVID-19 Outbreak in Wuhan, China". In: *JAMA - Journal of the American Medical Association* (2020). DOI: 10.1001/jama.2020.6130.

[5] Benjamin J. Cowling et al. "Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study". In: *The Lancet Public Health* 5.5 (2020), e279–e288. DOI: 10.1016/S2468-2667(20)30090-6.

[6] Kathy Leung et al. "First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment". In: *The Lancet* 395.10233 (2020), pp. 1382–1393. DOI: 10.1016/S0140-6736(20)30746-7.

[7] Yang Liu, Sebastian Funk, and Stefan Flasche. "The contribution of pre-symptomatic infection to the transmission dynamics of COVID-2019". In: *Wellcome Open Research* 5 (2020), p. 58. DOI: 10.12688/wellcomeopenres.15788.1.

[8] Anne Kimball et al. *Asymptomatic and presymptomatic SARS-COV-2 infections in residents of a long-term care skilled nursing facility - King County, Washington, March 2020*. 2020. DOI: 10.15585/MMWR.MM6913E1.

[9] Zhanwei Du et al. "Serial Interval of COVID-19 among Publicly Reported Confirmed Cases". In: *Emerging infectious diseases* 26.6 (2020). DOI: 10.3201/eid2606.200357.

[10] Sheng Zhang et al. "Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis". In: *International Journal of Infectious Diseases* 93 (2020), pp. 201–204. DOI: 10.1016/j.ijid.2020.02.033.

[11] Hiroshi Nishiura, Natalie M. Linton, and Andrei R. Akhmetzhanov. "Serial interval of novel coronavirus (COVID-19) infections". In: *International Journal of Infectious Diseases* 93 (2020), pp. 284–286. DOI: 10.1016/j.ijid.2020.02.060.

[12] Tapiwa Ganyani et al. "Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020". In: *Eurosurveillance* 25.17 (2020), p. 2000257. DOI: 10.2807/1560-7917.ES.2020.25.17.2000257.

[13] Don Klinkenberg and Hiroshi Nishiura. "The correlation between infectivity and incubation period of measles, estimated from households with two cases". In: *Journal of Theoretical Biology* 284.1 (2011), pp. 52–60. DOI: 10.1016/j.jtbi.2011.06.015.

[14] Robert V Hogg, Joseph McKean, and Allen T Craig. *Introduction to mathematical statistics*. Pearson Education, 2005.

[15] Minah Park et al. "A Systematic Review of COVID-19 Epidemiology Based on Current Evidence". In: *Journal of Clinical Medicine* 9.4 (2020), p. 967. DOI: 10.3390/JCM9040967.

[16] Stephen A. Lauer et al. "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application". In: *Annals of Internal Medicine* (2020). DOI: 10.7326/m20-0504.

[17] Shi Zhao et al. "Estimating the serial interval of the novel coronavirus disease (COVID-19): A statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020". In: *medRxiv* (2020), p. 2020.02.21.20026559. DOI: 10.1101/2020.02.21.20026559.

[18] Qun Li et al. "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia". In: *The New England journal of medicine* 382.13 (2020), pp. 1199–1207. DOI: 10.1056/NEJMoa2001316.

[19] Lauren Tindale et al. "Transmission interval estimates suggest pre-symptomatic spread of COVID-19". In: *medRxiv* February (2020), p. 2020.03.03.20029983. DOI: 10.1101/2020.03.03.20029983.

[20] Natalie M. Linton et al. "Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data". In: *Journal of Clinical Medicine* 9.2 (2020), p. 538. DOI: 10.3390/jcm9020538.

[21] Jantien A Backer, Don Klinkenberg, and Jacco Wallinga. *Incubation period of 2019 novel coronavirus (2019- nCoV) infections among travellers from Wuhan, China, 20 28 January 2020*. 2020. DOI: 10.2807/1560-7917.ES.2020.25.5.2000062.

A Data & Code

All code for this project is available online at:
https://github.com/mishra-lab/covid19-Re

Table A.1 summarizes the reported parametric COVID-19 distributions, and Figure A.1 illustrates the serial interval and generation time distributions explored for estimating $R_e(t)$.

Table A.1: Summary of reported parametric COVID-19 distributions

	Ref	Author	N	Distribution	Mean	SD
$S(\tau)$	[9]	Du et al.	468	Norm $(\mu = 3.96,\ \sigma = 4.75)$	3.96	4.75
	[10]	Zhang et al.	35	Gam $(\alpha = 3.619,\ \beta = 1.416)$	5.12	2.69
	[11]	Nishiura et al.	28	LogN $(\mu = 4.7,\ \sigma = 2.9)$	4.7	2.9
	[17]	Zhao et al.	21	Gam $(\alpha = 2.151,\ \beta = 2.045)$	4.4	3.0
	[18]	Li et al.	6	Gam $(\alpha = 4.866,\ \beta = 1.541)$	7.5	3.4
	[19]	Tindale et al.	4	Norm $(\mu = 4.22,\ \sigma = 0.4)$	4.22	0.4
	[19]	Tindale et al.	4	Norm $(\mu = 4.56,\ \sigma = 0.95)$	4.56	0.95
$H(\tau)$	[16]	Lauer et al.	181	Gam $(\alpha = 5.807,\ \sigma = 0.948)$	5.51	2.28
	[19]	Tindale et al.	135	Weib $(\alpha = 2.25,\ \beta = 10.15)$	8.99	4.23
	[19]	Tindale et al.	93	Weib $(\alpha = 1.88,\ \beta = 7.97)$	7.07	3.91
	[20]	Linton et al.	52	LogN $(\mu = 5.6,\ \sigma = 2.8)$	5.6	2.8
	[21]	Backer et al.	88	Weib $(\alpha = 3.038,\ \beta = 7.163)$	6.4	2.3
	[18]	Li et al.	10	LogN $(\mu = 5.2,\ \sigma = 3.91)$	5.2	3.91
$G(\tau)$	this	—	—	Gam $(\alpha = 1.813,\ \beta = 2.199)$	3.99	2.96
	[12]	Ganyani et al.	*	Gam $(\alpha = 9.140,\ \beta = 0.569)$	5.20	1.72
	[12]	Ganyani et al.	*	Gam $(\alpha = 6.843,\ \beta = 0.577)$	3.95	1.51

Notation — N: sample size; *: indeterminate; $S(\tau)$: serial interval; $H(\tau)$: incubation time; $G(\tau)$: generation time; μ: mean; σ: std dev (SD); α: shape; β: scale.

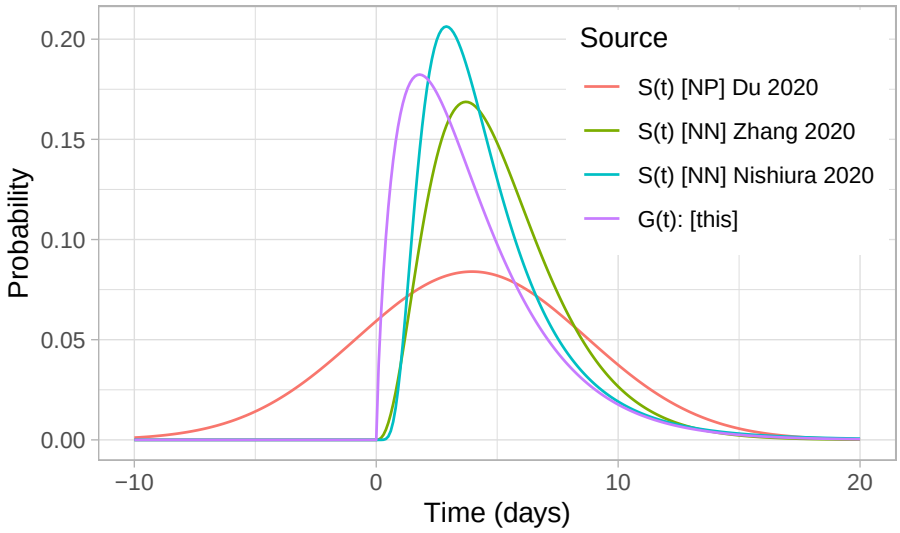

Figure A.1: Illustration of reported serial interval / generation time distributions used for calculating $R_e(t)$ in COVID-19