Monitoring changes in COVID-19 transmission over time

Appendix

Estimation of the time-varying reproduction number

If the incidence of infection at time t is $I(t)$, then the renewal equation is:

$$I(t) = R_t(t) \int_{-\infty}^{t} w(t-\tau) I(\tau)\ d\tau.$$

In practice, as in the examples here, time is discretized into daily units both for reporting of cases and for analysis, so that we obtain this estimate for the time-varying (also referred to as instantaneous in Cori et al. (1)) reproduction number:

$$R_t(t) = \frac{I(t)}{\sum_{\tau=-\infty}^{t} w(t-\tau) I(\tau)},$$

with τ being the time period over which R_t is estimated and $w(t-\tau)$, the discrete distribution of the serial interval.

The serial interval is the elapsed time between symptom onsets of a primary case and a secondary case in a chain of transmission. Its distribution for COVID-19 is calculated here as a gamma distribution with a mean value of 5.12 days and a standard deviation of 4.28 days (see below).

To capture variation in COVID-19 transmission at the national and state-levels, the median and interquartile ranges of daily sub-regional estimates (among states for U.S. national estimates, and among counties for state estimates) are weighted by $\sum_{\tau=-\infty}^{t} w(t-\tau) I(\tau)$.

Back-calculation of onset dates

The R_t estimation relies on the onset date but as this information is not available for all cases in the CDC COVID-19 line list, we use the deconvolution method to impute missing onset dates. We used the distribution between symptom onset and reporting for all laboratory-confirmed and probable cases where both dates are known and the backprojNP function from the R package **surveillance** (2). The backprojection method is implemented at the US level but can be applied at the state level to capture spatial heterogeneities in reporting delays. Caution should be applied when using these methods on locations with a high proportion of non laboratory-confirmed cases as the delay distribution of these probable cases may have longer right tails than laboratory-confirmed cases. The overall median reporting delay was 9 days (95% CI: 0.5 - 79) and the proportion of cases with both dates is 45% as of August 8, 2020. The method is an adaptation of the expectation-maximization-smoothing (EMS) algorithm initially developed for work on AIDS data by Becker et al. (3). This method could also be applied to estimate exposure dates for R_t estimation, using the incubation period distribution. Further work is under way to take temporal heterogeneity in onset to reporting delays into account. The median reporting delay is 9 days (95%CI: 0.5 - 79) for data with hospitalization status.

Additionally nowcasting methods can be applied to account for reporting delays for estimation of the time-dependent R_t.

Calculation of a serial interval meta-estimate

Search for serial intervals estimates

The serial interval is defined as the time between symptom onsets of a primary case and a secondary case in a chain of transmission. We searched for estimates of serial intervals on PubMed, medRxiv, arXiv and Google scholar using the keywords: "Serial intervals" AND "COVID-19" OR "SARS-CoV-2" OR "novel coronavirus" OR "2019-nCoV". Criteria for inclusion required provision of:

- The number of infector-infectee pairs used for the estimate
- The mean and standard deviations (SDs) for the estimate OR shape and scale parameters that allow calculation of SDs (see Appendix Table 1).

Estimates not fitted to a statistical distribution and those that rely on a subsample of the population (e.g., young adults) were excluded. Several limitations should be noted. First, serial intervals may be underestimated, because identification of earliest transmission events might be overrepresented, although some studies adjusted for right truncation of the data (4). Additionally, because serial interval estimates rely on the use of publicly available data, we used only one of two estimates (the one with highest number of infector-infectee pairs), in cases where both estimates were based on the same set of pairs. We indicated in Table 1 studies that used publicly available data for which unique identifiers could not be traced to verify whether infector-infectee pairs were used in other studies. Finally, in some instances serial intervals may have been adjusted with left censoring to exclude negative values that may be generated when a data set includes cases of pre-symptomatic transmission. We did not account for this possibility because information about such adjustments was not always available.

Pooling distributions

We assume that there are k serial interval distribution estimates to be pooled, each with known mean, m_i, standard deviation, s_i and study sample size, n_i, respectively ($i = 1 \ldots k$). Since each moment of the pooled distribution is a mean of the respective moment from each distribution weighted by its sample size, the mean of the pooled distribution, μ, is

$$\mu = \frac{\sum_i m_i n_i}{\sum_i n_i}$$

and (since the second moment of each distribution is $s_i^2 + m_i^2$) the standard deviation of the pooled distribution, σ, is given by

$$\sigma^2 + \mu^2 = \frac{\sum_i (s_i^2 + m_i^2) n_i}{\sum_i n_i}.$$

Using this method we calculate a serial interval meta-estimate with a mean of 5.12 (SD = 4.28).

If the pooled distribution is to be approximated to a gamma distribution, the shape parameter, k, and the scale parameter, θ, may be estimated using the method of moments so that:

$$k = \mu^2 / \sigma^2$$

and

$$\theta = \sigma^2 / \mu.$$

This serial interval pooled estimate relies on currently available data and can be revised as new information becomes available. A second method for calculating the serial interval pooled estimate that uses random resampling from each individual distribution, using the number of pairs in each study, and fits a gamma distribution to the pooled distribution yields the same mean, with a similar variance.

APPENDIX TABLE 1. List of studies included in the serial interval meta-estimate for COVID-19 used for the United States national and state R_t estimate, March-July 2020.

First Author	Last Date for case used	Location (sublocation)	Number of infector-infectee pairs	Mean	SD	Distribution
Nishiura et al. (4)	2020-02-12	China, Taiwan, Republic of Korea, Vietnam, Singapore, Germany	28	4.9	2.80	gamma
Du et al.* (5)	2020-02-08	China (93 cities)	468	4.0	4.75	normal
Bi et al. (6)	2020-02-12	China (Shenzhen)	48	6.3	4.20	gamma
Zhang et al.*† (7)	2020-02-17	China (outside Hubei)	34	5.1	2.68	gamma
Ping et al. (8)	2020-02-16	China (Guizhou)	57	6.4	4.15	weibull
Ganyani et al. (9)	2020-02-26	Singapore	54	5.2	4.32	gamma
Ganyani et al. (9)	2020-02-27	China (Tianjin)	114	4.0	4.24	gamma
Wang et al.* (10)	2020-02-12	China (Tianjin)	112	4.8	2.70	gamma
Cheng et al. (11)	2020-02-26	Mainland China	14	5.6	2.86	gamma
Cheng et al. (11)	2020-02-26	Hong Kong, Macau, Republic of Korea, Vietnam, Singapore, Germany	25	5.1	6.02	gamma
Cheng et al. (11)	2020-02-26	Taiwan	9	7.0	6.41	gamma
Wu et al. (12)	2020-02-10	China, Taiwan, Singapore, Malaysia, Vietnam, USA	43	7.0	4.50	gamma
Lavezzo et al. (13)	2020-03-08	Italy (Vo)	120	6.9	2.85	gamma
Zhao et al.† (14)	2020-02-15	Hong Kong	21	4.4	3.00	gamma
Chan et al. (15)	2020-04-02	Hong Kong	47	6.5	4.70	gamma
Kwok et al. (16)	2020-02-13	Hong Kong	26	4.8	3.32	gamma
Li et al.† (17)	2020-02-29	China (outside Hubei)	337	5.8	3.95	gamma

*SD generated from extracted shape and scale parameters.

† Studies that used publicly available data for which unique identifiers could not be traced to verify whether infector-infectee pairs were used in other studies.

Data reporting by onset date and hospitalization

Estimation of R_t relies on the availability of symptom onset dates. Similarly, ajusted R_t estimation uses data on hospitalization status. As completeness/availability of these data may vary by state and may affect the accuracy of R_t estimates, we caution users to pay attention to the availability of these data at the local level.

APPENDIX TABLE 2. Percentage of cases with reported onset date among laboratory-confirmed and probable COVID-19 cases in the CDC line list, by US state, March 1st to July 31st, 2020*,†.

State	Mar	Apr	May	Jun	Jul
AK	66.67% (3)	90.23% (174)	84.88% (258)	62.96% (378)	26.92% (2017)
AL	86.2% (906)	76.73% (6125)	60.8% (10965)	54.51% (20359)	20.15% (49226)
AR	98.39% (991)	69.81% (1633)	95.46% (3324)	95.81% (10084)	93.38% (11964)
AZ	70.94% (671)	69.87% (5509)	60.66% (10774)	40.98% (50461)	24.63% (96940)
CA	99.93% (14920)	99.91% (26507)	99.98% (37543)	99.98% (68271)	99.95% (65997)
CO	91.94% (2294)	76.82% (11402)	68.14% (10323)	54.88% (5858)	28.35% (13729)
CT	NA	NA	NA	23.42% (43868)	7.36% (3979)
DE	12.54% (295)	0.12% (4156)	NA	NA	NA
FL	85.19% (54)	82.88% (23594)	64.79% (19659)	64.61% (27395)	56.41% (56217)
GA	35.57% (3264)	24.69% (12950)	20.81% (11443)	24.54% (28637)	13.29% (72298)
HI	61.29% (31)	26.54% (407)	18.45% (103)	21.99% (191)	29.42% (571)
IA	91.3% (69)	87.49% (3901)	81.4% (9124)	69.91% (11557)	75.65% (10179)
ID	92.32% (573)	92.39% (1301)	79.84% (744)	71.86% (2999)	46.74% (13545)
IL	99.36% (157)	99.77% (2205)	99.69% (2260)	99.99% (138948)	99.78% (37389)
IN	88.64% (44)	100% (2)	100% (1)	NA	NA
KS	95.72% (421)	79.81% (3749)	57.54% (4917)	65.9% (1428)	57.74% (17933)
KY	91.12% (439)	85.85% (2863)	52.63% (2069)	55.04% (3481)	66.67% (21)
LA	60.78% (6366)	39.4% (9390)	21.71% (3787)	NA	NA
MA	83.37% (415)	49.7% (51347)	31.21% (47589)	32.12% (9450)	32.03% (9169)
MD	36.83% (687)	19.46% (9322)	9.46% (17863)	5.26% (14030)	6.79% (36284)
ME	88.58% (254)	86.91% (833)	72.51% (1004)	69.4% (840)	70.59% (578)
MI	95.35% (8987)	87.3% (14793)	80.1% (14128)	69.96% (12445)	78.65% (18014)
MN	100% (37)	98.23% (1188)	82.53% (21034)	85.91% (11595)	82.43% (20636)
MO	84% (50)	80% (30)	41.94% (11798)	25.05% (2858)	18.3% (3334)
MS	92.31% (130)	90.48% (6115)	77.76% (7604)	76.06% (7764)	70.27% (10312)
MT	80% (220)	86.62% (284)	69.05% (42)	79.07% (473)	76.93% (2544)
NC	84.37% (3263)	60.83% (9953)	60.73% (21763)	53.66% (44577)	36.98% (52690)
ND	99.56% (225)	98.69% (536)	100% (369)	100% (133)	NA
NE	53.85% (104)	51.23% (2805)	64.58% (3450)	66.7% (3180)	67.33% (3848)
NH	83.33% (246)	80.46% (1694)	70.76% (2353)	63.99% (1433)	64.94% (770)
NJ	46.43% (28)	37.65% (810)	52.73% (143181)	45.03% (27784)	56.91% (17674)
NM	55.29% (293)	69.06% (627)	47.37% (6669)	42.39% (4598)	41.32% (7563)
NV	85.26% (2191)	74.92% (3776)	76.56% (3762)	55.65% (13379)	27.45% (28819)
NY	46.78% (57194)	20.59% (97932)	5.09% (78769)	1.28% (34967)	2.3% (23526)
NYC	5.95% (47621)	2% (122260)	6.09% (34874)	14.65% (12140)	30.19% (9664)
OH	91.37% (1553)	60.49% (13684)	66.1% (14474)	62.13% (12001)	60.71% (39408)
OK	4.32% (324)	NA	NA	NA	NA
OR	100% (663)	99.94% (1811)	99.94% (1719)	99.98% (4405)	99.98% (9759)
PA	59.6% (4604)	46.23% (39566)	43.58% (25169)	46.71% (13996)	58.63% (22030)
RI	87.93% (381)	56.94% (706)	NA	NA	NA
SC	77.78% (9)	85.42% (5171)	70.98% (4548)	56.56% (20296)	36.73% (58285)
SD	81.4% (43)	100% (6)	NA	NA	NA
TN	93.24% (621)	70.36% (3634)	55.98% (18697)	59.15% (20005)	44.41% (62908)
TX	92.11% (266)	90.97% (1096)	85.08% (9509)	79.93% (6378)	52.74% (7831)

State	Mar	Apr	May	Jun	Jul
UT	100% (854)	99.56% (2937)	99.9% (6216)	99.93% (13143)	99.87% (19411)
VA	84.82% (1528)	72.03% (14552)	61.1% (26568)	53.73% (17431)	35.71% (28523)
VI	100% (2)	100% (1)	NA	50% (4)	54.55% (22)
VT	16.67% (30)	83.74% (855)	80.79% (151)	36.06% (208)	51.24% (242)
WA	94.74% (4146)	99.68% (7751)	99.7% (6405)	99.66% (10247)	99.58% (26186)
WI	87.09% (1526)	78.59% (5778)	63.91% (10194)	71.63% (5438)	73.94% (15922)
WV	71.43% (14)	32.14% (56)	71.83% (213)	77.19% (890)	88.53% (584)
WY	87.5% (24)	89.63% (376)	68.79% (141)	72.88% (756)	NA

*Total number of records by month and state in parenthesis, if a date and state of residence were reported.

† As of July 31, 2020, 96% of cases in the line list are laboratory-confirmed.

APPENDIX TABLE 3. Percentage of cases with reported hospitalization status (hospitalization + ICU) among laboratory-confirmed and probable COVID-19 cases in the CDC line list, by US state, March 1st to July 31st, 2020*,†.

State	Mar	Apr	May	Jun	Jul
AK	100% (3)	100% (174)	100% (258)	100% (378)	100% (2017)
AL	100% (906)	100% (6125)	100% (10965)	100% (20359)	100% (49226)
AR	99.8% (991)	100% (1633)	100% (3324)	100% (10084)	100% (11964)
AZ	100% (671)	100% (5509)	100% (10774)	100% (50461)	100% (96940)
CA	100% (14920)	100% (26507)	100% (37543)	100% (68271)	100% (65997)
CO	99.48% (2294)	99.88% (11402)	100% (10323)	100% (5858)	100% (13729)
CT	NA	100% (1)	NA	100% (43868)	100% (3979)
DE	100% (295)	100% (4156)	100% (3672)	100% (1875)	100% (3150)
FL	100% (54)	100% (23594)	100% (19659)	100% (27395)	100% (56217)
GA	100% (3264)	100% (12950)	100% (11443)	100% (28637)	100% (72298)
HI	100% (31)	100% (407)	100% (103)	100% (191)	100% (571)
IA	100% (69)	97.85% (3901)	94.27% (9124)	95.99% (11557)	99.54% (10179)
ID	100% (573)	100% (1301)	100% (744)	100% (2999)	100% (13545)
IL	99.36% (157)	96.15% (2205)	94.65% (2260)	61.4% (138948)	46.01% (37389)
IN	100% (44)	100% (2)	100% (1)	NA	NA
KS	99.52% (421)	89.92% (3749)	80.94% (4917)	85.22% (1428)	75.33% (17933)
KY	98.86% (439)	100% (2863)	100% (2069)	100% (3481)	100% (21)
LA	63.05% (6366)	40.92% (9390)	25.03% (3787)	NA	NA
MA	90.36% (415)	59.1% (51347)	41.14% (47589)	44.81% (9450)	43.32% (9169)
MD	100% (687)	100% (9322)	100% (17863)	100% (14030)	100% (36284)
ME	100% (254)	100% (833)	100% (1004)	100% (840)	100% (578)
MI	100% (8987)	100% (14793)	100% (14128)	100% (12445)	100% (18014)
MN	100% (37)	100% (1188)	100% (21034)	100% (11595)	100% (20636)
MO	94% (50)	96.67% (30)	100% (11798)	100% (2858)	100% (3334)
MS	99.23% (130)	100% (6115)	100% (7604)	100% (7764)	100% (10312)
MT	100% (220)	100% (284)	100% (42)	100% (473)	100% (2544)
NC	91.3% (3263)	82.28% (9953)	81.47% (21763)	70.19% (44577)	48.79% (52690)
ND	100% (225)	99.81% (536)	100% (369)	100% (133)	NA
NE	100% (104)	100% (2805)	100% (3450)	100% (3180)	100% (3848)
NH	100% (246)	99.06% (1694)	92.61% (2353)	94.21% (1433)	93.77% (770)
NJ	50% (28)	52.1% (810)	53.98% (143181)	46.53% (27784)	68.94% (17674)
NM	100% (293)	100% (627)	100% (6669)	100% (4598)	100% (7563)
NV	99.82% (2191)	99.87% (3776)	99.81% (3762)	99.19% (13379)	99.66% (28819)
NY	99.93% (57194)	99.88% (97932)	99.96% (78769)	99.99% (34967)	100% (23526)

State	Mar	Apr	May	Jun	Jul
NYC	37.65% (47621)	27.37% (122260)	22.32% (34874)	13.02% (12140)	21.99% (9664)
OH	99.94% (1553)	99.86% (13684)	99.81% (14474)	99.77% (12001)	99.79% (39408)
OK	99.69% (324)	100% (3673)	100% (2857)	100% (9247)	100% (24755)
OR	100% (663)	100% (1811)	100% (1719)	100% (4405)	100% (9759)
PA	66.12% (4604)	52.09% (39566)	52.14% (25169)	54.97% (13996)	68.54% (22030)
RI	83.99% (381)	65.58% (706)	NA	NA	NA
SC	100% (9)	100% (5171)	100% (4548)	100% (20296)	100% (58285)
SD	100% (43)	100% (6)	NA	NA	NA
TN	99.84% (621)	97.52% (3634)	95.48% (18697)	98.15% (20005)	98.29% (62908)
TX	98.87% (266)	100% (1096)	100% (9509)	100% (6378)	100% (7831)
UT	100% (854)	100% (2937)	100% (6216)	99.89% (13143)	98.93% (19411)
VA	100% (1528)	100% (14552)	100% (26568)	100% (17431)	100% (28523)
VI	100% (2)	100% (1)	100% (1)	100% (4)	90.91% (22)
VT	100% (30)	100% (855)	100% (151)	100% (208)	100% (242)
WA	100% (4146)	100% (7751)	100% (6405)	100% (10247)	100% (26186)
WI	100% (1526)	100% (5778)	100% (10194)	100% (5438)	100% (15922)
WV	100% (14)	100% (56)	100% (213)	100% (890)	99.83% (584)
WY	100% (24)	91.49% (376)	77.3% (141)	64.29% (756)	NA

*Total number of records by month and state in parenthesis, if a date and state of residence were reported.

† As of July 31, 2020, 96% of cases in the line list are laboratory-confirmed.

References

1. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. American Journal of Epidemiology. 2013;178(9):1505–12.

2. Meyer S, Held L, Höhle M. Spatio-temporal analysis of epidemic phenomena using the R package surveillance. Journal of Statistical Software. 2017;77(11):1–55.

3. Becker NG, Watson LF, Carlin JB. A method of non-parametric back-projection and its application to aids data. Statistics in Medicine [Internet]. 1991;10(10):1527–42. Available from: https://doi.org/10.1002/sim.4780101005

4. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. International Journal of Infectious Diseases. 2020;93:284–6.

5. Du Z, Xu X, Wu Y, Wang L, Cowling B, Meyers LA. Serial interval of covid-19 among publicly reported confirmed cases. Emerging infectious diseases. 2020;26(6):1341.

6. 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. The Lancet Infectious Diseases.

7. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: A descriptive and modelling study. The Lancet Infectious Diseases.

8. Ping K, Mingyu L, Yun G, Ying Tao, Yan H. Epidemiologic characteristics of COVID-19 in Guizhou, China. medRxiv [Internet]. 2020; Available from: https://www.medrxiv.org/content/early/2020/03/06/2020.03.01.20028944

9. Ganyani T, Kremer C, Chen D, Torneri A, Faes C, Wallinga J, et al. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. Eurosurveillance. 2020;25(17).

10. Wang Y, Teunis P. Strongly heterogeneous transmission of COVID-19 in mainland China: Local and regional variation. Frontiers in Medicine. 2020;7:329–9.

11. Cheng H-Y, Jian S-W, Liu D-P, Ng T-C, Huang W-T, et al. High transmissibility of COVID-19 near symptom onset. medRxiv [Internet]. 2020; Available from: https://www.medrxiv.org/content/early/2020/03/19/2020.03.18.20034561

12. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, Salazar PM de, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine [Internet]. 2020;26(4):506–10. Available from: https://doi.org/10.1038/s41591-020-0822-7

13. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of COVID-19 outbreak in the municipality of Vo, Italy. medRxiv [Internet]. 2020; Available from: https://www.medrxiv.org/content/early/2020/04/18/2020.04.17.20053157

14. Zhao S, Gao D, Zhuang Z, Chong M, Cai Y, Ran J, 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. medRxiv [Internet]. 2020; Available from: https://www.medrxiv.org/content/early/2020/02/25/2020.02.21.20026559

15. Chan Y, Flasche S, Lam T, Leung M, Wong M, Lam H, et al. Transmission dynamics, serial interval and epidemiology of COVID-19 diseases in Hong Kong under different control measures [version 1; peer review: Awaiting peer review]. Wellcome Open Research. 2020;5(91).

16. Kwok KO, Wong VWY, Wei WI, Wong SYS, Tang JW-T. Epidemiological characteristics of the first 53 laboratory-confirmed cases of COVID-19 epidemic in Hong Kong, 13 February 2020. Eurosurveillance [Internet]. 2020;25(16). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.16.2000155

17. Li M, Liu K, Song Y, Wang M, Wu J. Serial interval and generation interval for respectively the imported and local infectors estimated using reported contact-tracing data of COVID-19 in China. medRxiv [Internet]. 2020; Available from: https://www.medrxiv.org/content/early/2020/04/19/2020.04.15.20065946