

**Supplementary web appendix:**

Supplement to: **Serial intervals and case isolation delays for COVID-19: a systematic review and meta-analysis**

**Authors:**

Sheikh Taslim Ali<sup>1,2</sup> PhD, Amy Yeung<sup>1</sup> BBiomedSc, Songwei Shan<sup>1,2</sup> MSc, Lin Wang<sup>3</sup> PhD, Huizhi Gao<sup>1</sup> MPH, Zhanwei Du<sup>1,2</sup> PhD, Xiao-Ke Xu<sup>4</sup> PhD, Peng Wu<sup>1,2</sup> PhD, Eric H. Y. Lau<sup>1,2</sup> PhD, Prof Benjamin J. Cowling<sup>1,2,†</sup> PhD.

**Affiliation:**

<sup>1</sup> WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong Special Administrative Region, China

<sup>2</sup> Laboratory of Data Discovery for Health, Hong Kong Science and Technology Park, Hong Kong Special Administrative Region, China

<sup>3</sup> Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

<sup>4</sup> College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China.

**† Corresponding author:**

Prof. Benjamin J Cowling, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, 7 Sassoon Road, Pokfulam, Hong Kong.

Tel: +852 3917 6711; Fax: +852 3520 1945; email: bcowling@hku.hk

## Contents:

	<i>Page No.</i>
<b>1. Search Strategies</b>	<i>3</i>
<b>2. Study inclusion and data extraction</b>	<i>3-5</i>
<b>3. Defining the variables for data extraction</b>	<i>6-7</i>
<b>4. Predefined rules of selecting estimates for temporal analysis on Mainland China data</b>	<i>7-8</i>
<b>5. List of the Figures/ Tables</b>	<i>9-13</i>
<b>6. References</b>	<i>13-14</i>

## 1. Search Strategies

Two independent and experienced review authors (AY and SS) have performed the articles search and data extraction from PubMed (including the preprints in *medRxiv*). The details of the article search results are presented in the following table S1. We focused for interval related parameters including serial intervals and isolation delay related intervals for COVID-19.

Table S1: The details of article search strategies and search results for epidemiological parameters (serial interval and isolation delay related intervals (onset-to-isolation and onset-to-hospitalization) for COVID-19) with search terms used to find the studies in PubMed.

Estimations	Timing	Search terms details	Number of articles in PubMed Search
Serial interval for COVID-19	<i>Search on:</i> 23-10-2020	#1: "serial interval" OR "generation interval" OR "generation time" OR "serial distribution"	242
	<i>Search duration:</i> 01-01-2020 to 22-10-2020	#2: "COVID-19" OR "coronavirus" OR "2019 nCoV" OR "SARS CoV 2" OR "SARS-CoV-2" OR "SARS-CoV" OR "SARS CoV" OR "2019 CoV" OR "Pneumonia"	73539
		#3: #1 AND #2. (Total number of studies at first search hit)	<b>91</b>
Isolation delay related interval for COVID-19	<i>Search on:</i> 23-10-2020	#1: "interval" OR "delay" OR "latency"	58552
		#2: "Isolation" OR "hospital admission" OR "containment" OR "quarantine"	26528
	<i>Search duration:</i> 01-01-2020 to 22-10-2020	#3: "COVID-19" OR "coronavirus" OR "2019 nCoV" OR "SARS CoV 2" OR "SARS-CoV-2" OR "SARS-CoV" OR "SARS CoV" OR "2019 CoV" OR "Pneumonia"	73774
		#4: #1 AND #2 AND #3 (Total number of studies at first search hit)	<b>441</b>

## 2. Study inclusion and data extraction

The details of the studies excluded and included under the above search are provided in the following table S2. Both the review authors (AY and SS) then performed the data extraction from these searched articles following predetermined data extraction form in MS Excel. The data extraction form includes the estimates of outcome variables, possible factors variables as well as

the information on the studies (as defined and mentioned in the below section 3 in supplementary web-appendix with the details).

Table S2: Summary of the exclusion strategies with the detailed criteria of excluding the studies from this review and the number of studies included in the review of epidemiological parameters (serial interval for and isolation delay related intervals (onset-to-isolation and onset-to-hospitalization) for COVID-19).

Estimations	Title/ Abstract/ Text Search	Exclusion details	Number of articles excluded in PubMed Search	Remarks
<b>Serial interval for COVID-19</b>	<b>Title &amp; Abstract</b>	Criteria-1: Duplicated paper/Author response/Pre-print of published paper/Review	17	Duplicated: Redacted versions of published paper
		Criteria-2: Modeling/Simulation/Data not used or not original	15	Analysis is not based on original data, either by modeling/predictions, or by deriving serial interval from other studies
		Criteria-3: Non-COVID	1	Related to other diseases (Pneumonia)
		<b>Number of studies excluded by title &amp; abstract</b>	33	
	<b>Full text</b>	Criteria-4: Modeling/Simulation/Data not used or not original	10	
		Criteria-5: Data was incomplete (i.e. missing estimates) No serial interval reported	4	Missing Estimates/time frame/uncertainty. Full text did not show any serial interval estimate
		<b>Number of studies excluded by full text</b>	14	
		<b>Number of studies excluded in total</b>	47	<b>44 studies included</b> (out of 91 studies screened) from initial search, <b>12 studies included</b> additionally (out of 27 studies) from existing reviews.  <b>In total of 56 unique studies included</b> in this review.

<b>Isolation delay related intervals for COVID-19</b>	<b>Title &amp; Abstract</b>	Criteria-1: Clinical symptoms/Case reports	54	Unique case reports of individuals or symptoms/markers that can be measured by clinicians
		Criteria-2: Diagnostics/Screening/Therapeutics	67	
		Criteria-3: Not COVID-focused	78	Focused on comorbidity management of COVID or not related to COVID at all
		Criteria-4: Behavioural response and Healthcare systems	65	Related to people's behaviour changes caused by COVID, e.g., lifestyle, or the burden of COVID on healthcare systems, e.g. hospital capacity
		Criteria-5: Duplicated paper/Author response/Pre-print of published paper/Review	20	Unique case reports of individuals or symptoms/markers that can be measured by clinicians
		Criteria-6: Modeling/Simulation/Data not used or not original	56	
		Criteria-7: Virology	13	Focused on virus properties alone
		Criteria-8: Risk factor analysis	27	Related to risk factors or other characteristics of patients
		<b>Number of studies excluded by title</b>	380	
		<b>Full text</b>	Criteria-9: No isolation estimate	51
	Criteria-10: Incomplete data		2	
	<b>Number of studies excluded by full text</b>		53	
	<b>Number of studies excluded in total</b>		433	<b>8 studies included finally</b> (out of 441 studies screened) from our initial search, <b>10 studies included</b> additionally (out of 18 studies) from existing reviews.  In total of <b>18 unique studies included</b> in this review

### 3. Defining the variables for data extraction

Based on our hypothesis and objectives, the information was retrieved from these included studies and broadly classified into following variables.

The *outcome-variables* are the estimates of serial interval and isolation delay related intervals with respective *uncertainty types* (measures), which often reported as 95% confidence intervals (CI), 95% credible intervals (CrI), standard deviation (sd), inter-quartile range (IQR) and range. For comparison, we defined the *uncertainty lengths* (the length between upper and lower uncertainty bounds), calculated from these uncertainty estimates, and further defined *adjusted uncertainty lengths* to make it as much as comparable. If mean and standard deviation are provided, the *adjusted uncertainty lengths* are calculated either as the  $2 \times 1.96 \times \frac{sd}{\sqrt{n}}$ , ( $n$  is sample size) when approximated by a normal distribution or by generating distribution using re-estimated shape and scale parameters for the distributions, which can't be a normal approximation. If it's an IQR, first we approximated mean as median and standard deviation as  $IQR/1.35$  and then calculated the *adjusted uncertainty lengths*. The *adjusted uncertainty lengths* simply calculated when only range provided as  $0.95 \times \text{range}$ .

The diverse estimates of these interval measures reported can be attributed by several factors, including methodological factors, temporal factors and spatial factors, defined as following factor variables.

#### **Variables for methodological factors:**

*Estimation types*: the variable to identify the nature of the reported estimates (mean or median).

*Distribution types*: the variable to indicate whether the estimates are derived empirically or using probabilistic distributions fitting (e.g. normal, Gumbel, Weibull, Gamma, lognormal, etc.).

*Truncation*: the variable specifies whether the data is truncated to address the possible biases for the estimating the outcome variables.

*Settings*: this is a variable to indicate whether the estimate are evaluated based on the transmission pairs for household or community settings.

*Sample sizes*: the variable to account the data volume, the number of transmission pairs (for serial interval) and number of cases (isolation delay related intervals) used to estimate the *outcome-variables*.

*Data types*: the variable to indicate the data used based on the time of either onset illness or case reports or confirmations or hospitalizations.

#### **Variables for temporal factors:**

*Start date*: the earliest date of the data window used to estimate the *outcome variables* in the studies.

*End date*: the latest date of the data window used to estimate the *outcome variables* in the studies.

*Mid date*: the mid date of the data window used to estimate the *outcome variables* in the studies.

*Duration*: the variable indicates the data length (days) for analysis.

For analysing the temporal factors, we defined the respective dummy variables considering the days since the first onset case.

#### **Variables for spatial factors:**

*Location*: the variable indicates the *country* and *provinces* (specific regions) for which the *outcome variables* were estimated.

Along with the information on above mentioned variables (defined or derived), we retrieved the data for the details on the studies including *Author* (as first author), *Journal* (publishing journal), *Publication date*. The conflicts over inclusion of the studies and estimates on these variables were resolved by reassessing the particular articles again by both the reviewing authors together. The details of the data on these variables are summarised in table S3 and S4 and available online at <https://osf.io/c37zh/>.

#### **4. Predefined rules of selecting estimates for temporal analysis on Mainland China data**

For further analysis on the estimates from Mainland China data, we considered following set of rules on selection of the estimates as there were reported several different estimates on outcomes and factors variables derived from same data set. These predefined rules were set to make these

estimates more comparable and hence interpretable, specifically for the temporal analysis of the outcome variables.

- i. If both mean and median reported, we preferred mean over median, as most of the reports are mean instead.
- ii. For uncertainty measure with mean estimates, we preferred CI/ CrI first, if not reported the adjusted uncertainty measures (as explained in above section 3) were opted.
- iii. For uncertainty measure with median estimates, we preferred adjusted uncertainty measures derived by IQR first over by the range.
- iv. We opted the estimates based on the theoretical (probabilistic) distribution models fitting over the empirical estimates. If several estimates on different model fittings were reported, we opted the estimates with best fit (e.g., lowest AIC) as mentioned in the text. If it was not clearly mentioned in the text, we preferred the estimates derived by fitting Normal distribution (allowing both positive and negative values) for serial intervals and Gamma distribution for isolation delay related intervals.
- v. If study reported both isolation and hospitalization, we opted isolation.
- vi. We avoid the estimate on sub-groups unless it is based on temporal scale.



## 5. List of the Figures/ Tables

**Figure S1:** Boxplots showing serial interval estimates of COVID-19 for the studies in all locations (a-e) and for mainland China (f-j) against various factors: sample size (categorised relative to median), duration (categorised relative to median), distribution types, settings (household or non-household) and truncation.

**Figure S2:** Boxplots showing uncertainty of the serial interval estimates of COVID-19 for the studies in all locations (a-b) and for mainland China (c-d) against the factors: sample size (categorised relative to median), duration (categorised relative to median).

**Figure S3:** Boxplots showing temporal pattern of the serial interval estimates of COVID-19 for the studies in mainland China estimated (based on mid dates of data window) during pre-peak period (Before 20 January 2020), peak period (20-31 January, 2020) and post-peak period (after 31 January, 2020).

**Figure S4:** Boxplots showing isolation delay related interval estimates of COVID-19 for the studies in all locations (a-d) and for mainland China (e-h) against various factors: sample size (categorised relative to median), duration (categorised relative to median), distribution types and truncation.

**Figure S5:** Boxplots showing uncertainty of the isolation delay related interval estimates of COVID-19 for the studies in all locations (a-b) and for mainland China (c-d) against the factors: sample size (categorised relative to median), duration (categorised relative to median).

**Figure S6:** Boxplots showing temporal pattern of the isolation delay related interval estimates of COVID-19 for the studies in mainland China estimated (based on mid dates of data window) during pre-peak period (Before 20 January 2020), peak period (20-31 January, 2020) and post-peak period (after 31 January, 2020).

**Figure S7:** The temporal variation in reported estimates on serial intervals for COVID-19 in mainland China. The plot showing the reported serial interval estimates (in red circles) over time by start dates of the data windows used for estimation the serial intervals. Where the horizontal bars indicate the data window (indicating start dates and end dates) of the individual experiments, with the colour gradient representing the sample sizes (transmission pairs), constructed for each data window (with shades, in light blue: log-value of smaller pair size, dark blue: log-value of larger pair size, grey: pair size was not available). The epidemic curve with the onsets of confirmed cases (in gray line) and epidemic curve with the onset of infectors and infectee in the transmission pairs (in teal columns as available from 7 January 2020 to 28 February 2020) for mainland China alone, shown for reference of the epidemic timing <sup>2,3</sup>.

**Figure S8:** The temporal variation in reported estimates on isolation delay related intervals for COVID-19 in mainland China. The plot showing the reported isolation delay related interval estimates (in red circles) over time by start dates of the data windows used for estimation the isolation delay related intervals. Where the horizontal bars indicate the data window (indicating start dates and end dates) of the individual experiments, with the colour gradient representing the sample sizes (number of cases), constructed for each data window (with shades, in light blue: log-value of smaller sample size, dark blue: log-value of larger sample size, grey: sample size was not available). The epidemic curve with the onsets of confirmed cases (in gray line) and epidemic curve with the onset of infectors and infectee of the transmission pairs (in teal columns as available from 7 January 2020 to 28 February 2020) for mainland China alone, shown for reference of the epidemic timing <sup>2,3</sup>.

**Figure S9:** The association between serial interval and case isolation (a sensitivity analysis accounting the weights due to varied number of studies reported during each weeks). The regression model prediction of estimated serial intervals (weekly pooled estimates by taking average) by the estimates of isolation delay related intervals (weekly pooled estimates by taking average) accounting for the weights by the number of in the mainland China. The black dots are scattered plot of weekly pooled serial interval and isolation delay related estimates. Blue line is the fitted serial intervals predicted by case isolation delay related intervals with 95% CI (in dashed red lines). Gray shaded region indicates the standard error for the liner prediction. The sensitivity analysis provides similar conclusion as for every one day reduction in the estimated isolation delay related intervals, the estimated serial intervals reduced by 0.40 (95% CI: 0.27, 0.53, p-value = 0.00057) days.

**Table S5:** Table showing the number of studies and estimates of outcome variables (serial interval and isolation delay related intervals) and their estimation types reported for COVID-19 on the data with the pooled their estimates for different countries.

Locations (Data used for estimations)	Estimates of serial intervals					Estimates of isolation delay related intervals				
	Number of studies	Number of estimates	Mean	Median	Pooled mean (n)	Number of studies	Number of estimates	Mean	Median	Pooled mean (n)
Mainland China	28	58	40	18	5.15 (39)	11	53	35	18	4.89 (20)
China+	5	14	8	6	6.44 (8)	-				
Hong Kong	2	13	12	1	5.15 (2)	4	16	15	1	5.12 (5)
South Korea	6	12	8	4	4.63 (6)	1	2	1	1	4.3 (1)
India	1	6	1	5	5.2 (1)					
Singapore	3	4	4		4.6 (3)	1	2	1	1	5.6 (1)
Taiwan	2	4	2	2	4.6 (2)					
Italy	2	4	4		6.9 (4)					
Argentina	1	4	2	2	5.5 (1)					
Brunei	2	2	2		4.85 (2)					
Iran	2	2	2		5.15 (2)					
Brazil	1	2	1	1	3 (1)					
Philippines	1	1	1		6.9 (1)					
Germany	1	1		1	4 (1)					
Vietnam	1	1	1		3.2 (1)					
UK						1	1		1	6.5 (1)
DPCS	1	1		1	2 (1)					4.99 (28)
<b>Total</b>	<b>56</b>	<b>129</b>	<b>88</b>	<b>41</b>	<b>5.20 (73)</b>	<b>18</b>	<b>74</b>	<b>52</b>	<b>22</b>	

Notes: UK: The United Kingdom, China+: Other countries along with mainland China, DPCS: Diamond Princess Cruise Ship, n: is the number of estimates (excluding multiple estimates from same data) used to calculate the pooled mean for each country.

**Table S6:** Table showing the pooled mean (range) for outcome variable by estimation types.

Type of estimates	Serial interval estimates		Isolation related interval estimates	
	Pooled mean (Range)	Pooled mean Uncertainty length (Range)	Pooled mean (Range)	Pooled mean Uncertainty length (Range)
<b>Mean</b>	5.40 (2.6, 9.9)	5.21 (0.7, 33.4)	5.34 (1.5, 12.5)	8.29 (0.3, 47.2)
<b>Median</b>	4.54 (1.9, 6.0)	4.26 (1.0, 27.7)	3.87 (1.0, 8.0)	4.17 (0.2, 40.9)
<b>P-values</b>	<0.001	0.44	0.02	0.11

Note: p-values are for t-test (Welch Two Sample t-test) to compare the difference between the estimations by types (mean and median)

**Table S7:** Table showing the correlation tests (Pearson's, Spearman rank and Kendall rank correlation tests) results (including the respective correlation coefficients and p-values) for the estimates of outcome variables, serial interval and corresponding uncertainty for sample size, duration, start date and mid date for the studies for all countries and mainland China. We first considered the earliest start/ mid date as the origin and calculated the interval between origin and the respective start dates and mid dates of the data window used in the study to estimate the serial intervals. Then we use these intervals as the proxy factor variable for start and mid dates and performed the correlation tests.

Location	Outcome Variables	Factor Variables	Pearson's test		Spearman's test		Kendall test	
			$r$	p-value	$\rho$	p-value	$\tau$	p-value
All Locations	Serial Interval Estimates	Sample size*	-0.02 (-0.25, 0.22)	0.900	-0.25 (-0.45, -0.02)	0.040	-0.17 (-0.31, -0.02)	0.048
		Duration	-0.02 (-0.24, 0.22)	0.898	0.02 (-0.21, 0.25)	0.873	0.02 (-0.13, 0.16)	0.856
	Uncertainty for Serial Interval Estimates	Sample size*	-0.22 (-0.43, 0.02)	0.077	-0.72 (-0.82, -0.59)	< 0.001	-0.61 (-0.69, -0.50)	< 0.001
		Duration	0.02 (-0.22, 0.24)	0.896	0.02 (-0.21, 0.25)	0.837	0.02 (-0.13, 0.17)	0.826
Mainland China	Serial Interval Estimates	Sample size*	-0.21 (-0.50, 0.13)	0.220	-0.42 (-0.66, -0.11)	0.011	-0.29 (-0.48, -0.08)	0.015
		Duration	0.04 (-0.29, 0.35)	0.834	0.06 (-0.27, 0.38)	0.718	0.04 (-0.18, 0.26)	0.714
		Start date*	-0.35 (-0.60, -0.03)	0.033	-0.16 (-0.47, 0.18)	0.332	-0.13 (-0.34, 0.09)	0.269
		Mid date*	-0.33 (-0.59, -0.01)	0.041	-0.20 (-0.49, 0.14)	0.237	-0.13 (-0.34, 0.09)	0.261
	Uncertainty for Serial Interval Estimates	Sample size*	-0.24 (-0.53, 0.10)	0.162	-0.85 (-0.92, -0.72)	< 0.001	-0.73 (-0.82, -0.61)	< 0.001
		Duration	0.16 (-0.17, 0.45)	0.351	0.10 (-0.24, 0.41)	0.568	0.07 (-0.15, 0.29)	0.528

Note: \* indicate the factors are statistically significant (i.e., p-value < 0.05).  $\rho$ : Spearman's rank correlation coefficients (rho).  $\tau$ : Kendall rank correlation coefficients (tau).

**Table S8:** Table showing the correlation tests (Pearson's, Spearman rank and Kendall rank correlation tests) results (including the respective correlation coefficients and p-values) for the estimates of outcome variables, isolation delay related interval and corresponding uncertainty for sample size, duration, start date and mid date for the studies for all countries and mainland China. We first considered the earliest start/ mid date as the origin and calculated the interval between origin and the respective start dates and mid dates of the data window used in the study to estimate the isolation delay related intervals. Then we use these intervals as the proxy factor variable for start and mid dates and performed the correlation tests.

Location	Outcome Variables	Factor Variables	Pearson's test		Spearman's test		Kendall test	
			r	p-value	$\rho$	p-value	$\tau$	p-value
All Locations	isolation delay related intervals	Sample size*	-0.38 (-0.67, 0.004)	0.053	-0.52 (-0.75, -0.17)	0.006	-0.36 (-0.60, -0.12)	0.009
		Duration	-0.31 (-0.61, 0.07)	0.107	-0.22 (-0.56, 0.17)	0.252	-0.18 (-0.42, 0.08)	0.185
	Uncertainty for isolation delay related intervals	Sample size	0.29 (-0.11, 0.61)	0.157	-0.04 (-0.42, 0.34)	0.831	-0.13 (-0.38, 0.13)	0.343
		Duration	0.71 (0.44, 0.86)	< 0.001	0.13 (-0.26, 0.49)	0.518	0.08 (-0.18, 0.33)	0.581
Mainland China	isolation delay related intervals	Sample size	-0.32 (-0.69, 0.17)	0.190	-0.44 (-0.74, 0.02)	0.070	-0.31 (-0.57, -0.002)	0.069
		Duration	-0.17 (-0.57, 0.30)	0.476	-0.08 (-0.51, 0.39)	0.739	-0.09 (-0.39, 0.23)	0.580
		Start date*	-0.47 (-0.76, -0.03)	0.037	-0.42 (-0.73, 0.04)	0.066	-0.32 (-0.58, -0.01)	0.051
		Mid date*	-0.64 (-0.84, -0.28)	0.002	-0.58 (-0.82, -0.18)	0.007	-0.42 (-0.65, -0.12)	0.010
	Uncertainty for isolation delay related intervals	Sample size	0.40 (-0.09, 0.73)	0.104	-0.11 (-0.54, 0.36)	0.666	-0.17 (-0.46, 0.15)	0.324
		Duration	0.14 (-0.35, 0.57)	0.582	-0.03 (-0.48, 0.43)	0.906	-0.03 (-0.34, 0.28)	0.849

Note: \* indicate the factors are statistically significant (i.e., p-value < 0.05).  $\rho$ : Spearman's rank correlation coefficients (rho).  $\tau$ : Kendall rank correlation coefficients (tau).

## References:

1. 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**(4).
2. Ali ST, Wang L, Lau EHY, et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* 2020; **369**(6507): 1106-9.

3. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Meyers LA. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg Infect Dis* 2020; **26**(6): 1341-3.