

# **Epidemiological parameters of coronavirus disease 2019: a case series study**

Shujuan Ma, Jiayue Zhang, Minyan Zeng, Qingping Yun, Wei Guo, Yixiang Zheng,  
Shi Zhao, Maggie H Wang, Zuyao Yang

Submitted to: Journal of Medical Internet Research  
on: May 08, 2020

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 32

    Figures ..... 33

        Figure 1..... 34

        Figure 2..... 35

    Multimedia Appendixes ..... 36

        Multimedia Appendix 4..... 37

        Multimedia Appendix 1..... 37

        Multimedia Appendix 2..... 37

        Multimedia Appendix 3..... 37

        Multimedia Appendix 5..... 37



# Epidemiological parameters of coronavirus disease 2019: a case series study

Shujuan Ma<sup>1\*</sup> PhD; Jiayue Zhang<sup>1\*</sup> MPhil; Minyan Zeng<sup>2\*</sup> MSc; Qingping Yun<sup>3\*</sup> BSc; Wei Guo<sup>1\*</sup> MPhil; Yixiang Zheng<sup>4\*</sup> MD; Shi Zhao<sup>5</sup> MPhil; Maggie H Wang<sup>5</sup> PhD; Zuyao Yang<sup>5</sup> PhD

<sup>1</sup>Department of Epidemiology and Health Statistics Xiangya School of Public Health Central South University Changsha CN

<sup>2</sup>Peking University Shenzhen Hospital Shenzhen CN

<sup>3</sup>Department of Social Medicine and Health Education School of Public Health Peking University Beijing CN

<sup>4</sup>Xiangya Hospital Central South University Changsha CN

<sup>5</sup>JC School of Public Health and Primary Care The Chinese University of Hong Kong Hong Kong HK

\*these authors contributed equally

## Corresponding Author:

Zuyao Yang PhD

JC School of Public Health and Primary Care

The Chinese University of Hong Kong

Shatin, New Territories

Hong Kong

HK

## Abstract

**Background:** The outbreak of coronavirus disease 2019 (COVID-19) has been declared a pandemic by the World Health Organization, while the estimates of several key epidemiological parameters of the disease by many studies were based on small sample sizes or inaccurate due to various issues.

**Objective:** This study aimed to obtain more robust estimates of the incubation period, serial interval, frequency of pre-symptomatic transmission, and basic reproduction number ( $R_0$ ) of COVID-19.

**Methods:** We systematically retrieved and screened 20909 reports of laboratory confirmed COVID-19 cases released by the health authorities of China, Japan, Singapore and Vietnam and 9942 publications retrieved from PubMed and CNKI through April 8, 2020. To be eligible, a report had to contain individual data that allowed for accurate estimation of at least one parameter. Widely used models such as lognormal distributions were fitted to the datasets and the results with best-fitting values were presented.

**Results:** In total, 1605 cases were included for the final analysis. The mean and standard deviation were estimated to be 7.04 days and 4.27 days, respectively, for incubation period ( $n=693$ ), and 6.49 days and 4.94 days, respectively, for serial interval ( $n=1024$  pairs). In 40 (5.77%) cases, the incubation period was longer than 14 days. In 32 (3.13%) infector-infectee pairs, infectees' symptom onsets occurred before those of infectors. Pre-symptomatic transmission occurred in 131 (43.67%) infector-infectee pairs ( $n=300$  pairs).  $R_0$  was estimated to be 1.75 (95% CI: 1.69-1.82).

**Conclusions:** This study provides robust estimates of several epidemiological parameters of COVID-19. The findings support current practice of 14-day quarantine of persons with potential exposure, but also suggest the need for additional measures. Pre-symptomatic transmission together with the asymptomatic transmission reported by previous studies highlight the importance of adequate testing, strict quarantine and social distancing.

(JMIR Preprints 08/05/2020:19994)

DOI: <https://doi.org/10.2196/preprints.19994>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

✓ **No, I do not wish to publish my submitted manuscript as a preprint.**

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

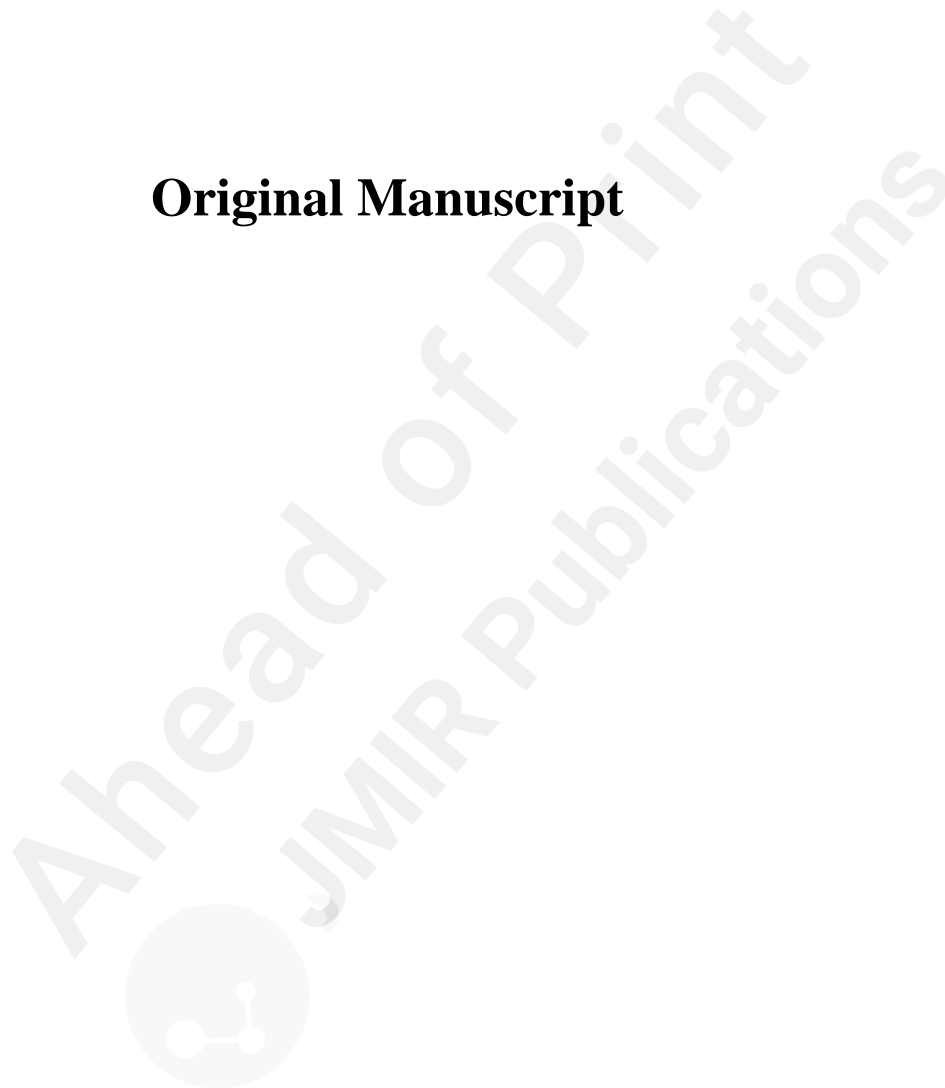
✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to the public.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>, I will be able to make my manuscript PDF available to the public.



## Original Manuscript



## Epidemiological parameters of coronavirus disease 2019: a case series study

Shujuan Ma, PhD<sup>1\*</sup>, Jiayue Zhang, MPH<sup>1\*</sup>, Minyan Zeng, MSc<sup>2\*</sup>, Qingping Yun, BMSc<sup>3\*</sup>, Wei Guo, MPH<sup>1\*</sup>, Yixiang Zheng, PhD<sup>4\*</sup>, Shi Zhao, MPhil<sup>5</sup>, Maggie H. Wang, PhD<sup>5</sup>, Zuyao Yang, PhD<sup>6#</sup>

<sup>1</sup> Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha 410078, China

<sup>2</sup> Department of Neurology, Peking University Shenzhen Hospital, Shenzhen 518036, China

<sup>3</sup> Department of Social Medicine and Health Education, School of Public Health, Peking University, Beijing 100191, China

<sup>4</sup> Department of Infectious Diseases, Key Laboratory of Viral Hepatitis of Hunan, Xiangya Hospital, Central South University, Changsha 410008, China

<sup>5</sup> Division of Biostatistics, JC School of Public Health and Primary Care, the Chinese University of Hong Kong, Hong Kong SAR, China

<sup>6</sup> Division of Epidemiology, JC School of Public Health and Primary Care, the Chinese University of Hong Kong, Hong Kong SAR, China

\* These authors contributed equally to this work as first authors.

# Corresponding author: Zuyao Yang, 4/F, School of Public Health building, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China. Email: [zyang@cuhk.edu.hk](mailto:zyang@cuhk.edu.hk). Phone: +852 2252 8739. Fax: +852 2606 3500.

**Word count: 3411**

## Abstract

**Background:** The outbreak of coronavirus disease 2019 (COVID-19) has been declared a pandemic by the World Health Organization, while the estimates of several key epidemiological parameters of the disease by many studies were based on small sample sizes or inaccurate due to various issues.

**Objectives:** This study aimed to obtain more robust estimates of the incubation period, serial interval, frequency of pre-symptomatic transmission, and basic reproduction number ( $R_0$ ) of COVID-19.

**Methods:** We systematically retrieved and screened 20658 reports of laboratory confirmed COVID-19 cases released by the health authorities of China, Japan, and Singapore and 9942 publications retrieved from PubMed and CNKI through April 8, 2020. To be eligible, a report had to contain individual data that allowed for accurate estimation of at least one parameter. Widely used models such as gamma distributions were fitted to the data sets and the results with best-fitting values were presented.

**Results:** In total, 1591 cases were included for the final analysis. The mean and standard deviation were estimated to be 7.04 days and 4.27 days, respectively, for incubation period ( $n=687$ ), and 6.49 days and 4.90 days, respectively, for serial interval ( $n=1015$  pairs). In 40 (5.82%) cases, the incubation period was longer than 14 days. In 32 (3.15%) infector-infectee pairs, infectees' symptom onsets occurred before those of infectors. Pre-symptomatic transmission occurred in 129 (43.58%) infector-infectee pairs ( $n=296$  pairs).  $R_0$  was estimated

to be 1.85 (95% CI: 1.37, 2.60).

**Conclusions:** This study provides robust estimates of several epidemiological parameters of COVID-19. The findings support current practice of 14-day quarantine of persons with potential exposure, but also suggest the need for additional measures. Pre-symptomatic transmission together with the asymptomatic transmission reported by previous studies highlight the importance of adequate testing, strict quarantine, and social distancing.

**Keywords:** coronavirus disease 2019; incubation period; serial interval; basic reproduction number; pre-symptomatic transmission.

## Introduction

In December 2019, a novel enveloped RNA beta-coronavirus, which was later named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first reported in Wuhan, the capital city of Hubei province of China [1]. The disease caused by SARS-CoV-2, i.e., coronavirus disease 2019 (COVID-19), spread across and outside China rapidly, and was declared a pandemic by the World Health Organization on 11 March 2020.

Despite the explosive growth of the number of studies on COVID-19 [2-4], several key epidemiological parameters of the disease remain to be clarified, among which are incubation period and serial interval. The mean or median incubation period and serial interval estimated by previous studies were mostly 4 to 5 days [5-16]. However, many of the studies included a



limited number of cases (around or less than 100) [5, 7, 9, 12-16]. For example, in the study of Zhang et al [16] which included 8579 cases, only 49 cases can be used for estimating incubation period and 39 pairs for serial interval. On the other hand, some studies might have suffered from inaccuracy of original data. For example, the interval of exposure was long or unclear, making it difficult to determine the exact exposure date [6, 8, 17] and giving rise to error.

Another important parameter of the transmission dynamics is the basic reproduction number ( $R_0$ ), which is defined as the average number of secondary cases caused by a single infectious individual in a totally susceptible population [18]. As  $R_0$  is often estimated based on serial interval [19], the abovementioned issues affecting previous estimates of serial interval might have affected the estimates of  $R_0$  as well. On the other hand, some studies estimated  $R_0$  based on latent period and infectious period [19, 20], with latent period approximated by incubation period. This may not be appropriate for COVID-19 as pre-symptomatic transmission might occur [21, 22]. However, few published studies [23] estimated how often and approximately when the disease could be transmitted prior to the symptom onset.

This study made use of the large amount of data reported by the health authorities of and outside China and those from published studies as of April 8, 2020 to address the above issues. Specifically, it was aimed to obtain more robust estimates of the following

epidemiological parameters of COVID-19: 1) incubation period, defined as the time interval between exposure and onset of disease symptoms; 2) serial interval, defined as the duration between symptom onset of an infector (e.g., a primary case) and that of an infectee (e.g., a secondary case) in a transmission chain, with a negative value meaning that the infectee's symptoms occurred before the infector's symptoms; 3) frequency of pre-symptomatic transmission; and 4)  $R_0$ .

## Methods

### Data sources

The details of data sources can be found in [Multimedia](#) Appendix 1. For China, all provinces, autonomous regions and municipalities (including mainland, Hong Kong, Macau and Taiwan) that had reported cases of COVID-19 were identified according to the daily updates by the National Health Commission of China [24]. Then, the official websites and WeChat accounts (if any) of local governments and health authorities (e.g., Municipal Health Commission, Center for Disease Control and Prevention, Department of Health) were checked manually through April 8, 2020 to identify and download the reports on laboratory-confirmed cases of COVID-19. Google and Baidu were searched to identify public media reports written based on official press releases. Chinese words for the following terms were used to do the search: ('family' OR 'household') AND ('cluster' OR 'dinner' OR 'party') AND 'infection'.

Other countries reporting COVID-19 cases were identified according to the 'coronavirus

disease (COVID-19) situation reports' by World Health Organization [25] and the data were searched through April 8, 2020. For Japan and Singapore, the information of confirmed cases was retrieved from their respective Ministry of Health. We also searched for individual cases from relevant departments and public media of the United States, the United Kingdom, Canada and Australia, but failed to find any with details allowing for parameters estimation in this study. Typically, the dates of exposure and symptom onset were lacking, e.g., see reference [26]. Owing to language barrier, we did not do a comprehensive search for other countries.

PubMed was searched to identify relevant publications by using the following terms: 'coronavirus', '2019-nCov', 'SARS-CoV-2', and 'COVID-19'. The China National Knowledge Infrastructure (CNKI) was searched to identify publications in Chinese journals using 'novel coronavirus' ('Xin Xing Guan Zhuang Bing Du' in Chinese pinyin), which was the official Chinese name for SARS-CoV-2. Both databases were searched from December 2019 to through April 8, 2020. The reference lists of eligible publications were also checked to see if there were other eligible studies not found by previous searches. We also kept an eye on the COVID-19 studies disseminated by public media and the official WeChat accounts of various academic entities in China and those recommended by experts in this field to the authors of the present study.

## Definitions and inclusion criteria

To be eligible, a report had to contain individual data that allowed for estimation of at least one of the following parameters of laboratory-confirmed cases of COVID-19: incubation period, serial interval and symptoms-to-transmission time, which was defined as the day of infectee's contact with infector relative to the latter's symptom onset date, with a negative value meaning that the transmission occurred before the infector developed symptoms (i.e. pre-symptomatic transmission).

To obtain an accurate estimate of incubation period, only the cases with an exposure period spanning 3 days or less were included in the analysis. For those exposed for three continuous days and those exposed on two dates with one day apart (i.e., exposed on the first and third days), the second day was uniformly used as the exposure date in estimation. For those exposed for two continuous days, the first day was uniformly used as the exposure date in estimation. This approach ensured the upper limit of error in the estimated incubation period be smaller than 1 day for the cases with a 2- or 3-day exposure, regardless of when exactly (i.e., first, second, or third day) the transmission actually occurred. The actual overall error was bound to be much smaller than 1 day, as most included cases were exposed for only 1 day which would dilute the error caused by 2- or 3-day exposures.

Serial interval was estimated based on the symptom onset dates of infector-infectee pairs, which were typically from cluster infections. For two cases to qualify as an infector-infectee pair and be included in this study, the following two criteria must both be fulfilled.

First, there was evidence that the presumed infector had been exposed outside the cluster (e.g., close contact with a confirmed case, travel history to Hubei, exposure to a person who returned from Hubei) before he/she **attended** the group activities such as family gathering and business conference that **led** to cluster infections. Second, **in the 14 days prior to symptom onset**, the presumed infectee was exposed to the presumed infector only, without other exposure histories.

The estimation of symptoms-to-transmission time also involved determination of exposure date and judgement about transmission chain, hence the **above** principles applied in estimating incubation period and serial interval were followed as well.

## **Screening, data extraction, and quality control**

Six researchers were involved in data collection. The reports retrieved through the above searches were scrutinized one by one according to the inclusion criteria specified above. The following data were extracted from eligible reports by using a standard extraction form which was pilot-tested with the reports from Liaoning province of China: the geographical location concerned, age, sex, type of exposure, first date and period (if applicable) of exposure, date of symptom onset, initial symptoms, and whether the case was from a cluster. For a clustering case, the generation he/she belonged to, the exposure date, and symptom onset date were also recorded. The retrieved reports were split into six parts, with each researcher responsible for one part. For each part, the eligibility of and data extracted from reports were firstly

determined by one researcher and then cross-checked by another three researchers. All uncertainty and disagreements were discussed among the researchers. If no consensus could be reached after discussion, the concerned cases would be excluded to ensure the correctness and accuracy of data. For example, if it could not be determined who was infected through attending a group activity (i.e., infectee) and who had been infected before he/she attended the activity (i.e., infector), then all clustering cases related to the activity had to be excluded.

## Data analysis

The basic characteristics of included cases were summarized descriptively. Lognormal, Weibull, and gamma distributions were fitted to the data sets of incubation period and serial interval and the one with the smallest Akaike information criterion (AIC) score was used for final analysis. For serial interval, which had negative values, shifted distributions were fitted, with the best-fitting values determined by maximum likelihood. The key parameters were estimated by using the maximum likelihood approach. The data on symptoms-to-transmission time was roughly symmetrical according to visual inspection and thus fitted by normal distribution. For each parameter, the range, median, selected percentiles, mean, and standard deviation were estimated. The 95% confidence intervals (CI) of mean and standard deviation were estimated by using the bootstrap technique.  $R_0$  was estimated by the well-studied Euler-Lotka equation [19]:  $R_0 = 1/M[-r|h(\cdot)]$ , where  $r$  is the exponential growth rate,  $h(\cdot)$  is the estimated distribution of the generation time, and the function  $M[\cdot]$  is the Laplace transform,

i.e., moment generating function known for statisticians, for  $h$ . As generation time is hard to observe directly, it was approximated by serial interval in this study. Compared with generation time, serial interval is expected to have the same mean but a larger variance due to possibly differing incubation periods of infectors and infectees. The exponential growth rate was obtained directly from a previous study reporting incidence data of the early stage of epidemic [1], while the other parameters came from the present study.

Sensitivity analyses were conducted to examine the robustness of the two parameters involving exposure date. Specifically, the third day for those whose exposure period spanned three days and the second day for those exposed for two continuous days were used as their exposure dates in sensitivity analyses. The difference between the estimates from sensitivity analysis and those from primary analysis represents the largest possible error in the latter. All statistical analyses were conducted with SAS software, version 9.4.

## Ethical statement

No ethical approval was needed as the data used in this study was publicly available, either released by health authorities or retrieved from electronic databases.

## Results

Figure 1 shows the process of data collection. We systematically retrieved and screened - 20658 case reports and 9942 publications. The 43 publications included for our analysis can be found in Multimedia Appendix 2. In total, 1591 cases were included for the final analysis,

including 1211 (76.12%) from China, 301 (18.92%) from Japan, 60 (3.77%) from Singapore, and 19 from five other countries. The cases from China covered 156 cities of 31 provinces out of 34 in total.

## Characteristics of included cases

The age of included cases ranged from 5 days to 95 years (Table 1). There were 133 cases (10.74%) with a travel history to Hubei province where Wuhan is the capital city. The source of exposure of 1302 (81.99%) cases was known confirmed cases. The exposure period was exactly 1 day in 508 (53.08%) cases, spanned 2-3 days in 227 (23.72%) cases, and spanned more than 3 days in the other cases (who were not included in estimation of parameters involving exposure date). The top five initial symptoms were fever (730/1009, 72.35%), cough (257/1009, 25.47%), fatigue (69/1009, 6.84%), sore throat (57/1009, 5.65%) and malaise (54/1009, 5.35%). Clustering cases accounted for 84.92% (1351/1591) of all, and most of them were the second generation.

Table 1. Characteristics of included cases

| Characteristics                  | No. (%) <sup>a</sup> |
|----------------------------------|----------------------|
| <b>Age (n=1437)</b>              |                      |
| Range                            | 5 days, 95 years     |
| Mean (standard deviation) (year) | 45.90 (18.02)        |
| 0 - 18                           | 89 (6.19)            |
| 19 - 64                          | 1121 (78.01)         |
| ≥65                              | 227 (15.80)          |
| <b>Sex (n=1559)</b>              |                      |
| Male                             | 768 (49.26)          |



|   |              |
|---|--------------|
| Female  | 791 (50.74)  |
| <b>Travel history to Hubei (n=1238)</b>             |              |
| Yes   | 133 (10.74)  |
| No  | 1105 (89.26) |
| <b>Infecter was a known confirmed case (n=1588)</b> |              |
| Yes   | 1302 (81.99) |
| No  | 286 (18.01)  |
| <b>Span of exposure period (n=957)</b>              |              |
| 1 day   | 508 (53.08)  |
| 2-3 days  | 227 (23.72)  |
| >3 days   | 222 (23.20)  |
| <b>Top 5 initial symptoms (n=1009)</b>              |              |
| Fever   | 730 (72.35)  |
| Cough   | 257 (25.47)  |
| Fatigue   | 69 (6.84)    |
| Sore throat   | 57 (5.65)    |
| Malaise   | 54 (5.35)    |
| <b>Clustering cases (n=1591)</b>                    |              |
| Yes   | 1351 (84.92) |
| No  | 240 (15.08)  |
| <b>Family cluster (n=1238)</b>                      |              |
| Yes   | 514 (41.52)  |
| No  | 724 (58.48)  |
| <b>Generation of clustering cases (n=1350)</b>      |              |
| 1 <sup>st</sup>                                     | 61 (4.52)    |
| 2 <sup>nd</sup>                                     | 1069 (79.19) |
| 3 <sup>rd</sup>                                     | 188 (13.93)  |
| 4 <sup>th</sup> or higher                           | 32 (2.37)    |

<sup>a</sup> Unless other specified

## Epidemiological parameters

Incubation period were estimated from 687 cases, of whom 482 (69.55%) were exposed

for 1 day and the others for two or three days. The incubation periods of individual cases

ranged from 0 to 23 days (Table 2), with a median of 6 days and 5.82% (n=40) of them longer than 14 days. The incubation period is best described by a gamma distribution (Multimedia Appendix 3, Figure 2(A)), with a mean of 7.04 days (95% CI: 6.74, 7.33) and standard deviation of 4.27 days (95% CI: 3.92, 4.43).

The serial interval was estimated for 1015 pairs of infector-infectee and ranged from -5 to 29 days (Table 2), with a median of 6 days. For 32 (3.15%) pairs it was smaller than 0; for these 32 pairs the mean was -2.50 days, meaning that the infectee showed symptoms 2.50 days earlier than the infector. The serial interval is best described by a shifted gamma distribution (Multimedia Appendix 3, Figure 2(B)), with a mean of 6.49 days (95% CI: 6.15, 6.80) and standard deviation of 4.90 days (95% CI: 4.68, 5.25). Based on an exponential growth rate of 0.10 per day[1] and the mean and standard deviation of serial interval,  $R_0$  was estimated at 1.85 (95% CI: 1.37, 2.60). The country specific  $R_0$  for China, Japan, and Singapore was 1.83 (95% CI: 1.37, 2.54), 1.83 (95% CI: 1.37, 2.53), and 1.98 (95% CI: 1.41, 2.97), respectively.

The symptoms-to-transmission time was estimated from 296 pairs of infector-infectee and ranged from -9 to 14 days (Table 2), with a median of 0 day. Pre-symptomatic transmission occurred in 129 infector-infectee pairs (43.58%, 95% CI 37.93%-49.23%); for these 129 pairs the mean symptoms-to-transmission time was -2.88 days, meaning that the transmission occurred 2.88 days before the infectors showed symptoms. Fitted by a normal

distribution (Figure 2(C)), the mean of this parameter was  $-0.07$  days (95% CI:  $-0.43, 0.31$ ) and standard deviation 3.31 days (95% CI: 2.96, 3.66).

The symptoms-to-transmission time can also be inferred by mean serial interval minus mean incubation time ( $-0.55$  day), which is consistent with the direct estimate ( $-0.07$  day) as both suggest the mean symptoms-to-transmission time to be around the day before primary cases' symptom onset. Based on the above estimates, the timeline of infection for an 'average' pair of infector-infectee in a transmission chain is demonstrated in **Multimedia Appendix 4**.

## Sensitivity and stratified and analyses

In sensitivity analysis, the estimates of incubation period and symptoms-to-transmission time remained stable, with a difference of 0.24-0.3 day from those in primary analysis (Table 2), which represents the largest possible error caused by inclusion of cases with an exposure period spanning 2 or 3 days. The results of stratified analysis are summarized in **Multimedia Appendix 5**. The difference between strata was statistically significant in **four** of the analyses (see footnote of **Multimedia Appendix 5**). Briefly, incubation period was longer for those exposed for only 1 day (as compared with those exposed for 2-3 days), while serial interval and symptoms-to-transmission time were longer for those at lower generations of clustered infection. **Serial interval was also longer for those infected through non-household contact.**

Table 2. The estimates of incubation period, serial interval and symptoms-to-transmission time (unit: day)

| Metrics | Incubation period | Serial interval | Symptoms-to-transmission |
|---------|-------------------|-----------------|--------------------------|
|---------|-------------------|-----------------|--------------------------|

|   | time               |                   |                     |
|---|--------------------|-------------------|---------------------|
| Range                                       | 0, 23.00           | −5, 29.00         | −9, 14.00           |
| P <sub>2.5</sub> , P <sub>97.5</sub>        | 1.00, 17.00        | −1.00, 18.00      | −7.00, 7.00         |
| Median (P <sub>25</sub> , P <sub>75</sub> ) | 6.00 (4.00, 10.00) | 6.00 (3.00, 9.00) | 0 (−2.00, 2.00)     |
| Mean (95%CI)                                | 7.04 (6.74, 7.33)  | 6.49 (6.15, 6.80) | −0.07 (−0.43, 0.31) |
| SD (95%CI)                                  | 4.27 (3.92, 4.43)  | 4.90 (4.68, 5.25) | 3.31 (2.96, 3.66)   |
| Mean (95%CI) <sup>a</sup>                   | 6.74 (6.43, 7.03)  | --                | 0.17 (−0.19, 0.55)  |
| SD (95%CI) <sup>a</sup>                     | 4.34 (3.98, 4.50)  | --                | 3.32 (2.96, 3.68)   |

<sup>a</sup> Results of sensitivity analyses. For incubation period, the third day for those whose exposure period spanned three days and the second day for those exposed for two continuous days were used as their exposure dates in sensitivity analyses. For symptoms-to-transmission time, exposure dates were taken in a different way in sensitivity analyses (like those of incubation period). Abbreviations: CI, confidence interval; SD, standard deviation.

## Discussion

By pooling the individual data of 1591 cases, we estimated the mean incubation period of COVID-19 to be 7.04 days, serial interval 6.49 days, and symptoms-to-transmission time -0.07 day. The incubation period was longer than 14 days in 5.82% of the cases. Infectees' symptom onsets occurred before those of infectors in 3.15% of the infector-infectee pairs. Pre-symptomatic transmission occurred in 43.58% of the infector-infectee pairs.  $R_0$  was estimated to be 1.85 (95% CI: 1.37, 2.60).

To our knowledge, this study represents the largest case series with accurate data on incubation period, serial interval, and frequency of pre-symptomatic transmission. Our estimates of incubation period and serial interval are longer than most of the previous estimates [5-10]. There are several possible reasons for the difference. First, the sample size is generally small in previous studies [5, 7, 9, 12-15], but much larger in the present one. Second, most cases in previous studies had a long or unclear interval of exposure, making it difficult to determine the exact exposure date [6, 8, 17] and giving rise to error. By contrast, this study applied strict inclusion criteria regarding exposure period and a simple method to determine exposure date to ensure the potential error in the estimates be small ( $< 0.3$  day, according to the sensitivity analysis). Third, the order of transmission (i.e., who is infector and who is infectee) in clustering cases, which is crucial to estimation of both parameters, is easy to be mistaken given the possibility of pre-symptomatic and asymptomatic transmission

[21, 22, 27, 28]. According to our experience in screening the publicly available data, for many clusters, there was no clear evidence (e.g., who got infected outside and brought the virus to the cluster?) to establish the transmission order. To ensure the accuracy of data, we excluded such clusters in this study, while previous studies rarely described how this issue was handled [11].

Our finding that the incubation period was within 14 days for 94% of the cases lends support to current practice of 14-day quarantine of persons with potential exposure to SARS-CoV-2. In line with other studies [17], we also found some cases who developed symptoms 14 days after exposure, indicating that longer quarantine periods might be justified for some people. However, as it is hard to know beforehand who will develop symptoms beyond 14 days of exposure, the cost of extending mandatory quarantine of many people and the potential consequence of failure to identify a few symptomatic cases must be weighed carefully. Repeated nucleic acid test at the end of quarantine and social distancing afterwards may help reduce the risk associated with this phenomenon.

The negative values of serial interval and symptoms-to-transmission time provide evidence of pre-symptomatic transmission. Specifically, 43.58% of the transmission events in this study occurred before infectors' symptom onsets. Asymptomatic transmission was also reported by the publications included in this study (Multimedia Appendix 2). These phenomena constitute a big challenge to the control of epidemic and highlight the importance

of adequate testing, strict quarantine, and social distancing to reduce the transmission caused by 'hidden' cases.

The  $R_0$  we estimated is smaller than those from previous estimates which were mostly between 2 and 4 [20, 29, 30]. The difference may be either due to the methodological issues in obtaining parameters as discussed above or due to the estimating method itself. In estimating  $R_0$ , serial interval was used to approximate generation time in the Euler-Lotka equation, which may lead to underestimation because of the possibly differing incubation periods of infector and infectee. Assuming the incubation periods of infector and infectee following the same distribution, we found that the  $R_0$  was 1.90, almost the same with the original one, suggesting a small error in our estimation. In any case, a smaller  $R_0$  should not be interpreted as low risk of transmission. Slow response of government, pre-symptomatic and asymptomatic transmission, and insufficient protection measures taken by the public together could lead to an out-of-control epidemic, as with the current situation in many countries.

This study has some limitations. First, it was based on publicly reported cases. Previous studies suggested that such cases may overrepresent the severe ones. However, as the publicly available information on severity was incomplete and inaccurate, we were unable to assess the magnitude and potential impact of this issue. Another problem with publicly reported cases is that their epidemiology history might be reported in more detail at the early stage of

epidemic than when the number of cases exploded. We found that this was especially true for Japan and Singapore. Thus, in general, the results of this study reflected more the situation of early stage. Second, to ensure accuracy in estimating incubation period, the cases exposed for a long (>3 days) or unclear period were excluded (which was inevitable and justifiable). This might have contributed partly to the longer incubation period in this study than others, as there was a trend towards a longer incubation period in the cases exposed for 1 day only (a relatively shorter period) as compared with those exposed for 2-3 days (Multimedia Appendix 5). Third,  $R_0$  might have been underestimated as Euler-Lotka equation was used; however, as explained in last paragraph, the magnitude of underestimation, if any, was very small.

In conclusion, this study obtained robust estimates of several key epidemiological parameters of COVID-19. It provides additional evidence on the mean incubation period of COVID-19, which supports the current practice of 14-day quarantine of persons with potential exposure but also suggests the need for additional measures. Pre-symptomatic transmission together with the asymptomatic transmission reported by previous studies highlight the importance of adequate testing, strict quarantine, and social distancing.



**Acknowledgement**

We thank all the people who worked hard, risked their health, and even sacrificed their lives to fight against COVID-19. They are real heroes. We also thank the numerous individuals who reduced their social activities to minimum to help contain the spread of SARS-CoV-2.

**Contributors**

YZ conceived the idea and designed the study, has full access to all data in the study, and takes responsibility for the integrity of data and the accuracy of data analysis. MS, ZJ, ZM, YQ, GW, ZY and YZ contributed to data acquisition, statistical analysis, and writing of the report. WMH and ZS contributed to statistical analysis. All authors contributed to data interpretation and critical revision and reviewed and approved the final version.

**Conflicts of interest**

All authors declare no competing interests.

**Funding**

This work received no funding from any source.

**Figure legends**

Figure 1. The flowchart of cases selection from the official reports and published studies.

Figure 2. The distribution of (A) incubation period (n=687), (B) serial interval (n=1015 pairs), and (C) symptoms-to-transmission time (n=296 pairs).

### **Multimedia Appendices**

**Multimedia** Appendix 1. A full list of the data sources (except academic studies)

**Multimedia** Appendix 2. A list of eligible studies included for analysis

**Multimedia** Appendix 3. The AIC scores and shifted terms for lognormal, Weibull and Gamma distributions

**Multimedia** Appendix 4. The timeline of events for an ‘average’ pair of infector-infectee in a transmission chain, according to the estimates from this study.  $E_1$ , exposure of infector;  $O_1$ , symptom onset of infector;  $E_2$ , exposure of infectee;  $O_2$ , symptom onset of infectee

**Multimedia Appendix 5.** Stratified analyses of incubation period, serial interval, and symptoms-to-transmission time according to selected characteristic

## References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *The New England journal of medicine*. 2020 Mar 26;382(13):1199-207. PMID: 31995857. doi: 10.1056/NEJMoa2001316.
2. Li J, Chen Z. Identification of Symptoms Prognostic of COVID-19 Severity: Multivariate Data Analysis of a Case Series in Henan Province. *Journal of medical Internet research*. 2020 Jun 30;22(6):e19636. PMID: 32544071. doi: 10.2196/19636.
3. Luo H, Lie Y. Surveillance of COVID-19 in the General Population Using an Online Questionnaire: Report From 18,161 Respondents in China. *JMIR public health and surveillance*. 2020 Apr 27;6(2):e18576. PMID: 32319956. doi: 10.2196/18576.
4. Timmers T, Janssen L. Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study. *JMIR mHealth and uHealth*. 2020 Jun 23;8(6):e19822. PMID: 32516750. doi: 10.2196/19822.
5. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro surveillance*. 2020 Feb;25(5). PMID: 32046819. doi: 10.2807/1560-7917.es.2020.25.5.2000062.
6. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England journal of medicine*. 2020 Apr 30;382(18):1708-20. PMID: 32109013. doi: 10.1056/NEJMoa2002032.
7. Jiang X, Rayner S. Does SARS-CoV-2 has a longer incubation period than SARS and MERS? *Journal of medical virology*. 2020 May;92(5):476-8. PMID: 32056235. doi: 10.1002/jmv.25708.

8. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of clinical medicine*. 2020 Feb 17;9(2):538. PMID: 32079150. doi: 10.3390/jcm9020538.
9. Xu XW, Wu XX, Jiang XG, Xu KJ, Ying LJ, Ma CL, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ (Clinical research ed)*. 2020 Feb 19;368:m606. PMID: 32075786. doi: 10.1136/bmj.m606.
10. Yang Y, Lu QB, Liu MJ, Wang YX, Zhang AR, Jalali N, et al. Epidemiological and clinical features of the 2019 novel coronavirus outbreak in China. <https://doi.org/10.1101/2020021020021675>. 2020.
11. Du ZW, Wang L, Xu XK, Wu Y, Cowling BJ, Meyers LA. Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerging infectious diseases*. 2020;26(6):1341-3. doi: 10.3201/eid2607.200357.
12. Tindale LCS, J. E., Coombe M, Garlock ES, Lau WYY, Saraswat M, Colijn C. Evidence for transmission of COVID-19 prior to symptom onset. *Elife*. 2020 2020 Jun 22;9(e57149). doi: 10.7554/eLife.57149.
13. You C, Deng YH, Hu WJ, Sun JR, Lin QS, Zhou F, et al. Estimation of the Time-Varying Reproduction Number of COVID-19 Outbreak in China. *Int J Hyg Environ Health*. 2020 2020 Jul;228:113555. doi: 10.1016/j.ijheh.2020.113555.
14. Zhao S, Gao DZ, Zhuang Z, Chong MK, Cai YL, Ran JJ, 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.

<https://doi.org/10.1101/2020022120026559>. 2020.

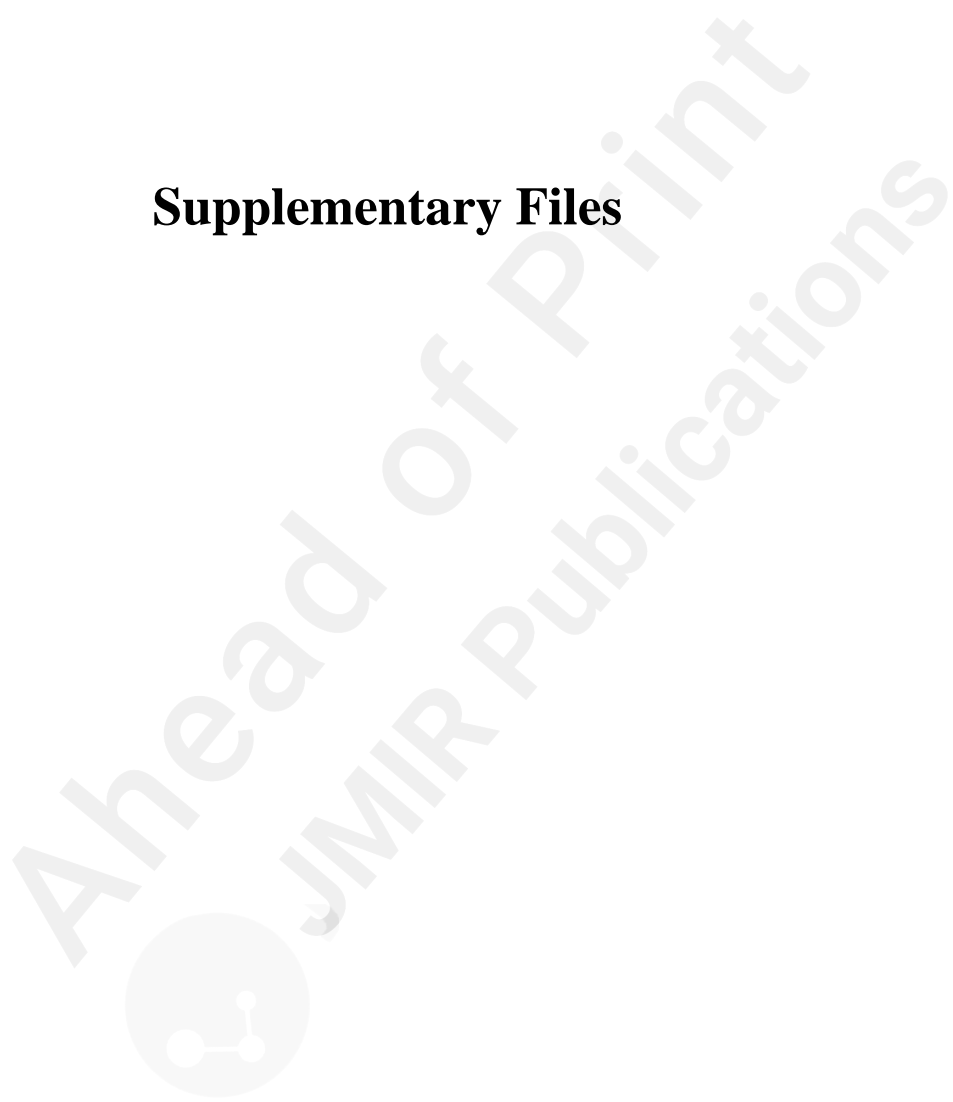
15. Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *International journal of infectious diseases*. 2020 Mar 4;93:284-6. PMID: 32145466. doi: 10.1016/j.ijid.2020.02.060.
16. Zhang J, Litvinova M, Wang W, Wang Y, Deng X, Chen X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious diseases*. 2020 Jul;20(7):793-802. PMID: 32247326. doi: 10.1016/s1473-3099(20)30230-9.
17. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of internal medicine*. 2020 2020 May 5;172(9):577-82. PMID: 32150748. doi: 10.7326/m20-0504.
18. Milwid R, Steriu A, Arino J, Heffernan J, Hyder A, Schanzer D, et al. Toward Standardizing a Lexicon of Infectious Disease Modeling Terms. *Frontiers in public health*. 2016;4:213. PMID: 27734014. doi: 10.3389/fpubh.2016.00213.
19. Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc Biol Sci*. 2007;274(1609):599-604. doi: 10.1098/rspb.2006.3754.
20. Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *Journal of evidence-based medicine*. 2020 Feb;13(1):3-7. PMID: 32048815. doi: 10.1111/jebm.12376.
21. Tong ZD, Tang A, Li KF, Li P, Wang HL, Yi JP, et al. Potential Presymptomatic Transmission of SARS-CoV-2, Zhejiang Province, China, 2020. *Emerging infectious*

- diseases. 2020 May 17;26(5):1052-4. PMID: 32091386. doi: 10.3201/eid2605.200198.
22. Yu P, Zhu J, Zhang Z, Han Y, Huang L. A Familial Cluster of Infection Associated With the 2019 Novel Coronavirus Indicating Possible Person-to-Person Transmission During the Incubation Period. *The Journal of infectious diseases*. 2020 Feb 18;221(11):1757-61. PMID: 32067043. doi: 10.1093/infdis/jiaa077.
23. Ferretti L, Wymant C. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*. 2020 May 8;368(6491):eabb6936. PMID: 32234805. doi: 10.1126/science.abb6936.
24. people.cn. <http://healthpeoplecomcn/GB/26466/431463/431576/indexhtml>.
25. WHO. Coronavirus disease (COVID-2019) situation reports. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>.
26. CBCnews. <https://www.cbcc.ca/amp/15482310>.
27. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. *Jama*. 2020 Feb 21;323(14):1406-7. PMID: 32083643. doi: 10.1001/jama.2020.2565.
28. Huang R, Xia J, Chen Y, Shan C, Wu C. A family cluster of SARS-CoV-2 infection involving 11 patients in Nanjing, China. *The Lancet Infectious diseases*. 2020 Feb 28;20(5):534-5. PMID: 32119823. doi: 10.1016/s1473-3099(20)30147-x.
29. Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. 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. *International journal of infectious diseases*. 2020 Mar;92:214-7. PMID: 32007643. doi: 10.1016/j.ijid.2020.01.050.
30. Distant C, Piscitelli P, Miani A. Covid-19 Outbreak Progression in Italian Regions:

Approaching the Peak by the End of March in Northern Italy and First Week of April in Southern Italy. International journal of environmental research and public health. 2020 Apr 27;17(9). PMID: 32349259. doi: 10.3390/ijerph17093025.



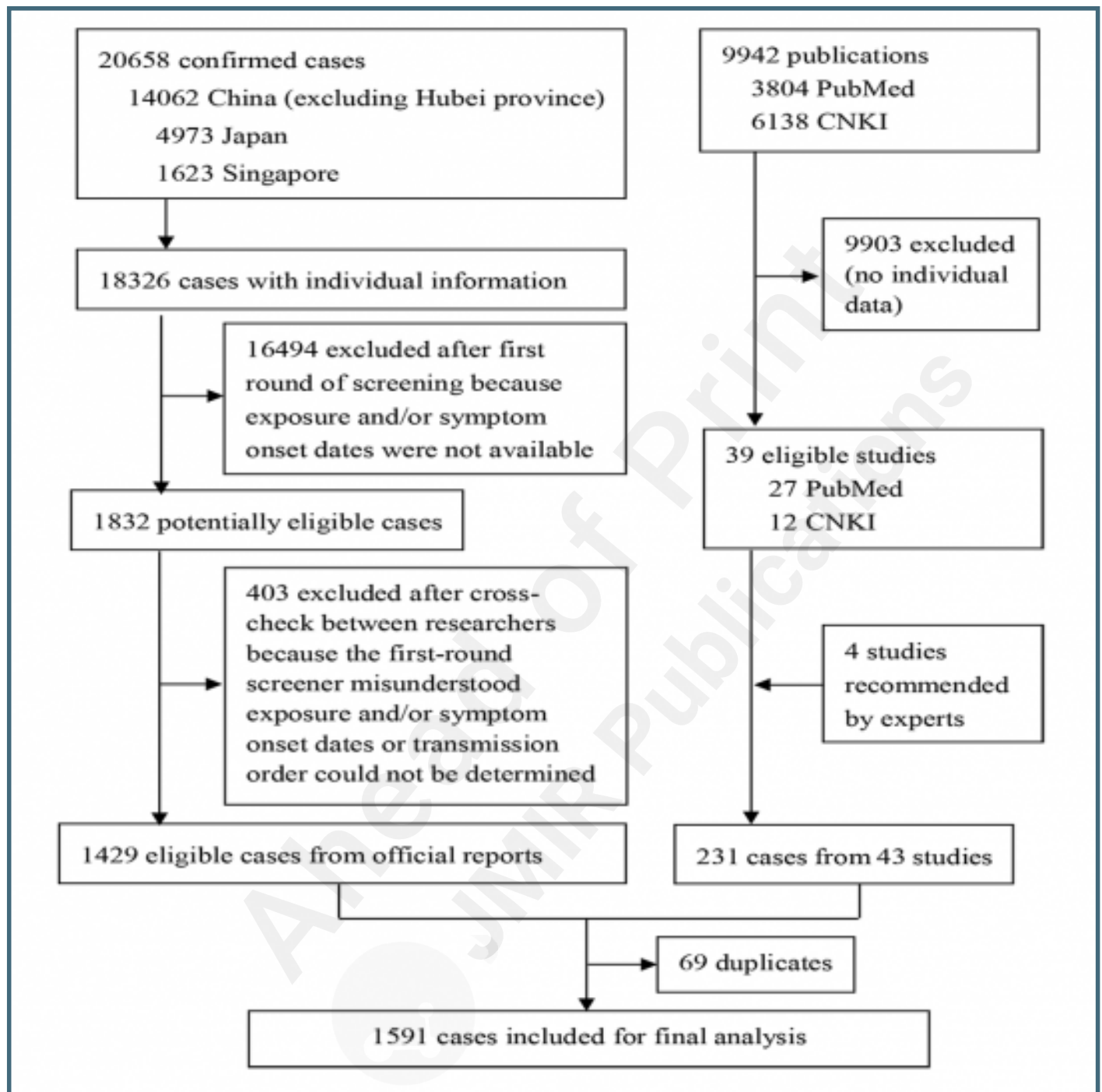
## Supplementary Files



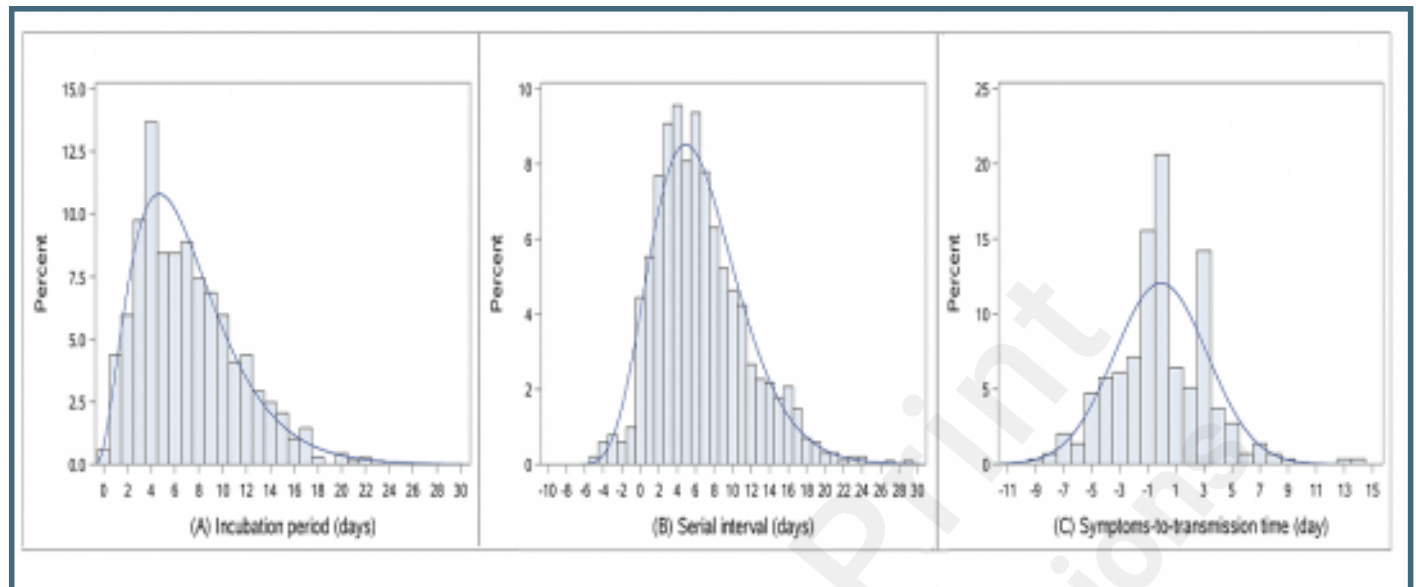


## Figures

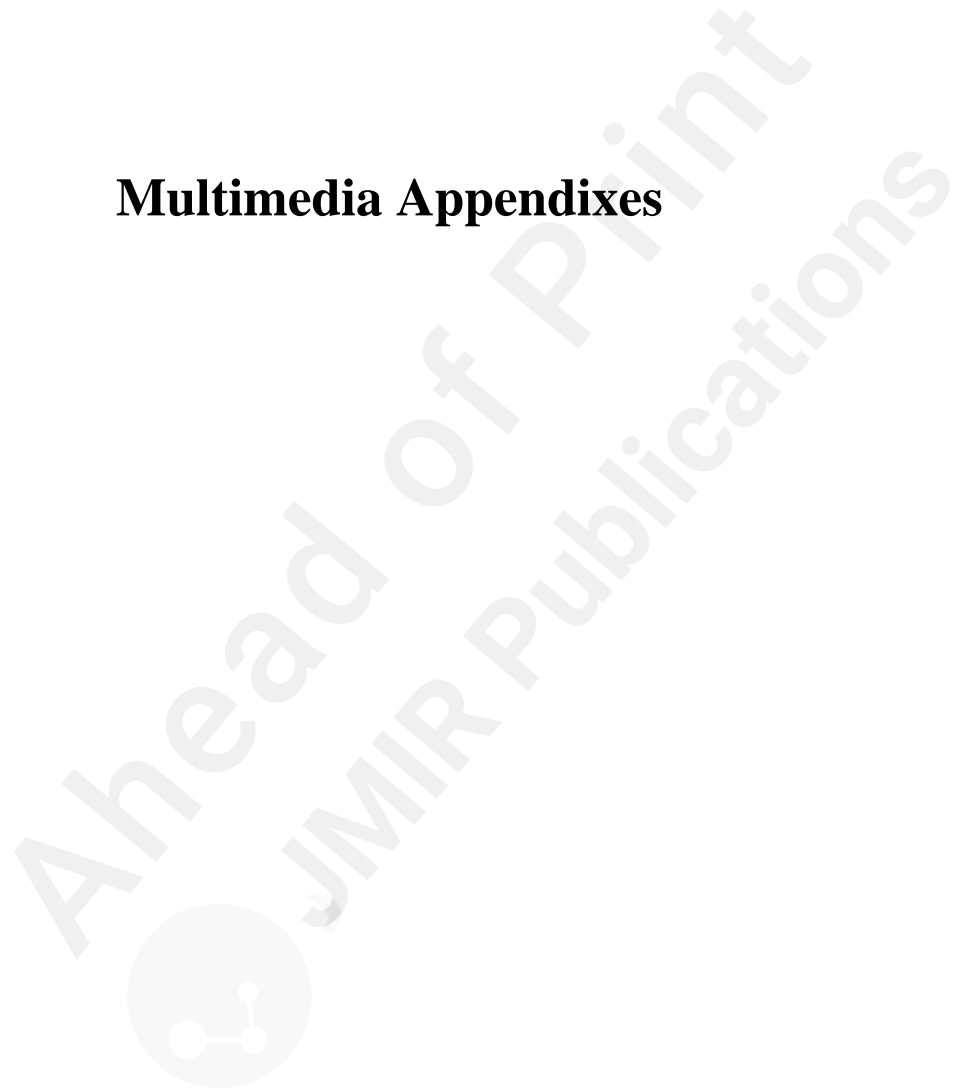
The flowchart of cases selection from the official reports and published studies.



The distribution of (A) incubation period (n=687), (B) serial interval (n=1015 pairs), and (C) symptoms-to-transmission time (n=296 pairs).



## **Multimedia Appendixes**



The timeline of events for an ‘average’ pair of infector-infectee in a transmission chain, according to the estimates from this study. E1, exposure of infector; O1, symptom onset of infector; E2, exposure of infectee; O2, symptom onset of infectee.

URL: <https://asset.jmir.pub/assets/8261ec40a78faedf6d0f625dd779815a.png>

A full list of the data sources (except academic studies).

URL: <https://asset.jmir.pub/assets/9d9670858af38d744bde7a97ce68768e.xlsx>

A list of eligible studies included for analysis.

URL: <https://asset.jmir.pub/assets/41a4e86fb5edde907231dd524f7bd752.docx>

The AIC scores and shifted terms for lognormal, Weibull and Gamma distributions.

URL: <https://asset.jmir.pub/assets/228327a732759c4b641405677936495e.xlsx>

Stratified analyses of incubation period, serial interval, and symptoms-to-transmission time according to selected characteristic.

URL: <https://asset.jmir.pub/assets/a9ebfe2ec3347771bb65319c3bd98c95.docx>

