Title: Modelling the coronavirus disease (COVID-19) outbreak on the Diamond Princess ship using the public surveillance data from January 20 to February 20, 2020

Running title: Modelling COVID-19 on Diamond Princess

Shi Zhao\textsuperscript{1,2,*}, Peihua Cao\textsuperscript{3}, Daozhou Gao\textsuperscript{4}, Zian Zhuang\textsuperscript{5}, Marc KC Chong\textsuperscript{1,2}, Yongli Cai\textsuperscript{6}, Jinjun Ran\textsuperscript{7}, Kai Wang\textsuperscript{8}, Yijun Lou\textsuperscript{5}, Weiming Wang\textsuperscript{6,*}, Lin Yang\textsuperscript{9}, Daihai He\textsuperscript{5,*}, and Maggie H Wang\textsuperscript{1,2}

1 JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China

2 Shenzhen Research Institute of Chinese University of Hong Kong, Shenzhen, China

3 Clinical Research Centre, Zhujiang Hospital, Southern Medical University, Guangzhou, Guangdong, China

4 Department of Mathematics, Shanghai Normal University, Shanghai, China

5 Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China

6 School of Mathematics and Statistics, Huaiyin Normal University, Huaian, China

7 School of Public Health, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China

8 Department of Medical Engineering and Technology, Xinjiang Medical University, Urumqi, 830011, China

9 School of Nursing, Hong Kong Polytechnic University, Hong Kong, China

* Correspondence to: zhaoshi.cmsa@gmail.com (SZ), weimingwang2003@163.com (WW), and daihai.he@polyu.edu.hk (DH)

Email address of all authors

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
SZ: zhaoshi.cmsa@gmail.com
PC: cphcc@smu.edu.cn
DG: dzgao@shnu.edu.cn
ZZ: larry.zhuang@polyu.edu.hk
MKCC: marc@cuhk.edu.hk
YC: yonglicai@hytc.edu.cn
JR: jimran@connect.hku.hk
KW: wangkaimath@sina.com
YL: yijun.lou@polyu.edu.hk
WW: weimingwang2003@163.com
LY: l.yang@polyu.edu.hk
DH: daihai.he@polyu.edu.hk
MHW: maggiew@cuhk.edu.hk
Abstract – 98 words

The novel coronavirus disease 2019 (COVID-19) outbreak on the Diamond Princess ship has caused over 634 cases as of February 20, 2020. We model the transmission process on the ship with a stochastic model and estimate the basic reproduction number at 2.2 (95% CI: 2.1–2.4). Through estimating the dispersion parameter which quantifies the superspreading events, we find that stochastic bursts were unlikely to occur if the reproduction number is less than 1. The epidemic doubling time is at 4.6 days (95% CI: 3.0–9.3), thus timely action were crucial. The lesson learn on the ship is generally applicable in other settings.

Keywords: coronavirus disease 2019; reproduction number; dispersion; superspreading; serial interval; Diamond Princess ship.
Backgrounds

The coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in Wuhan, China in the end of 2019, and spread to over 40 foreign countries in a short period of time [1-3]. The World Health Organization (WHO) declared the outbreak to be a public health emergency of international concern on January 30, 2020 [4]. The COVID-19 outbreak on the Diamond Princess, a British cruise ship that contains 3711 tourists and crew members, caused 634 confirmed cases as of February 20, 2020 and became the largest outbreak outside China [5].

In this study, we model the COVID-19 outbreak on the Diamond Princess ship, estimate the key epidemiological parameters of this outbreak and explore the effects of several different control measures.

Methods

We consider the population on the Diamond Princess ship as a close cohort with $N = 3711$ individuals. Following [6], we simulate the transmission process stochastically. For a given COVID-19 case with symptom onset on the $t$-th day, the number of secondary cases follows a negative binomial (NB) distribution with a mean at $R_{\text{eff}}(t)$ and a dispersion parameter at $k$ [7]. The $R_{\text{eff}}(t)$ is the effective reproduction number, and $R_{\text{eff}}(t) = R_0[N - C(t)]/N$, where $C(t)$ is the cumulative number of cases at the $t$-th day and $R_0$ is the basic reproduction number of COVID-19 to be estimated. The time delay between symptom onset dates of a primary case and any of its associated secondary cases is the serial interval (SI). Hence, for a given primary case with symptom onset on the $t$-th day, the symptom onset time of each associated secondary case is expected to be the summation of $t$ and SI of
COVID-19. It is conventional to assume the SI to follow a Gamma distribution with mean at 4.5 days and standard deviation (SD) at 3.1 days [8, 9].

We collect the confirmed cases time series released in the situation reports of COVID-19 infections in Japan [5]. We simulate the courses of the COVID-19 outbreaks stochastically starting from one infectious index case on January 20, 2020. We calculate the maximum likelihood estimates of $R_0$ and $k$ by fitting model to the number of confirmed cases with Poisson-distributed likelihood framework. The 95% confidence intervals (95%CI) of $R_0$ and $k$ are calculated by using the profile likelihood estimation framework with a cutoff threshold determined using a Chi-square quantile [10].

We repeat the fitting and estimation procedures above under an alternative scenario with mean SI at 7.5 days and SD at 3.4 days, which was estimated in [2]. We evaluate the modelling performance with mean SI at 4.5 or 7.5 days by the Akaike information criterion (AIC). Under the later scenario with mean SI at 7.5 days, we find that the $R_0$ is larger than 5 which is not in line with WHO estimates and hard to justify in a ‘quarantine’ setting. In particular, the later scenario yields a higher AIC by 10 units than the former scenario with mean SI at 4.5 days. Thus, we argue the former scenario is most likely the reality, and consider it as the main results.

Furthermore, we consider four ‘what-if’ scenarios and they were

- scenario (0): the best-fit model, i.e., baseline scenario;
- scenario (1): 1000 susceptibles were removed on February 11, 2020;
- scenario (2): the $R_0$ was reduced to 1.5; and
- scenario (3): combining scenarios (1) and (2).

We simulate the model under four different scenarios and estimate the key epidemiological parameters. We estimate the cumulative number of cases on February 20, the outbreak final size, the epidemic doubling time and the peaking time to capture the patterns of outbreak.
Ethics approval and consent to participate: The data were collected via public domain [5], and thus neither ethical approval nor individual consent was not applicable.

Results

With a mean SI at 4.5 days, we estimate the $R_0$ to be 2.2 (95%CI: 2.1–2.4), which is consistent with previous estimates [7, 8, 11]. The fitting results are shown in Fig 1(a) and (b), which match the observed data well. We estimate the dispersion term ($k$) to be 44 (95%CI: 6–88), which is significantly larger than 1 and consistent with [7]. The simulation results with or without dispersion term ($k$) were largely consistent, which suggests absent of superspreading events (note that low $k$ implies high chance of superspreading events).

Without control measures, we estimate that the COVID-19 outbreak is likely to cause 3066 cases (95%CI: 2046–3441) on the ship, and the epidemic curve is likely to peak around February 28, 2020 with a doubling time at 4.6 days (95%CI: 3.0–9.3), see Fig 1(b) and Table 1. Under scenario (1), timely reduction in the susceptible population could lower the final size and sustainably reduce the daily incidences. Under scenario (2), the lower the $R_0$, decreasing from 2.2 to 1.5, the lower the final size and the later the peaking time will be. If the susceptible pools and $R_0$ are reduced simultaneously, the COVID-19 outbreak on the ship will be likely mitigated and postponed largely, see Fig 1(d).

Discussion

The $R_0$ of COVID-19 on the Diamond Princess ship is likely to be lower than that of the severe acute respiratory syndrome (SARS) which ranges from 2.2 to 3.6 [12]. The $k$ quantifies the chance of the superspreading events and a large $k$ (> 1 significantly) implies that the stochastic bursts on the ship was less likely to occur if the reproduction number is less than 1 [6, 7, 13]. On the other hand, our finding highlights the different characteristics between COVID-19 (less superspreading) and SARS (more superspreading), which should be taken into consideration into mitigation. As such, the self-
protection actions including wearing facemasks, reducing outdoor activities or gathering and maintain hygiene and sterilized, were thus recommended.

The situation on the Diamond Princess cruise implies that the virus could spread rapidly, most likely with a short SI than previously estimated [2]. Note that a recent study on the 1099 patients found that the median of incubation period is only 3 days, thus a short effective SI is also possible [14]. With a shorter SI, a relatively lower $R_0$ could also result in a rapid growth of the epidemic size [15], and with a shorter epidemic doubling time as we show in Table 1. Therefore, timely contact tracing and effectively quarantine were crucial to shutoff the transmission.

The simulation results under scenarios (1)-(3) indicate that the public health control measures could effectively mitigate the COVID-19 outbreak on the ship in terms of the final size, see Table 1. Decreasing the disease transmissibility in terms of $R_0$ could postpone the peak, which may gain valuable time to prepare and allocate the resources in response to incoming patients. Relocating the population at risk (if possible) could sustainably decrease the daily incidences, and thus relieve the tensive demands in the healthcare, and improve in the treatment outcome.

We suggest maintaining the enhancement in both populational level public health control as well as the individual level self-protection actions in combating the COVID-19 outbreak. With detailed information on public intervention, the analytical framework in this study can be extended to a complex context and used for evaluating the effects of certain control measures.
Declarations

Availability of materials
All data used in this work were publicly available via [5].

Consent for publication
Not applicable.

Funding
DH was supported by General Research Fund (Grant Number 15205119) of the Research Grants Council (RGC) of Hong Kong, China. WW was supported by National Natural Science Foundation of China (Grant Number 61672013) and Huaian Key Laboratory for Infectious Diseases Control and Prevention (Grant Number HAP201704), Huaian, Jiangsu, China.

Acknowledgements
None.

Disclaimer
The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Conflict of Interests
The authors declared no competing interests.

Authors’ Contributions
SZ and DH conceived the study, carried out the analysis, and drafted the first manuscript. All authors discussed the results, critically read and revised the manuscript, and gave final approval for publication.
References


Figure

The observed (dots) and fitted (curves) number of COVID-19 cases on the Diamond Princess ship. Panel (a) shows the cumulative number of cases, and panel (b) shows the daily number of cases. In panels (a) and (b), the black dots are the observed number of cases time series, the blue curves are the simulation median (bold) and 95% CI (dashed), and the grey curves are 1000 simulation samples. In panel (b), the vertical black dashed line is February 20, 2020. Panel (c) shows the estimated epidemic curves under different scenarios. The bold curves are the simulation median, and the dashed curves are the 95% CIs. In panels (b) and (c), the blue curves show the baseline scenario, and they are the same.
# Table

## Table 1.

Summary of the key epidemiological estimates of the outbreak under different scenarios.

<table>
<thead>
<tr>
<th>scenario</th>
<th>Description</th>
<th>cum. # on Feb 20</th>
<th>final size</th>
<th>doubling time (day)</th>
<th>peaking time</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0)</td>
<td>Baseline</td>
<td>657 (296, 1298)</td>
<td>3066 (2046, 3441)</td>
<td>4.6 (3.0, 9.3)</td>
<td>Feb 28 (Feb 20, Mar 6)</td>
</tr>
<tr>
<td>(1)</td>
<td>remove 1000 susceptibles on Feb 11</td>
<td>557 (248, 1076)</td>
<td>2238 (1486, 2548)</td>
<td>5.2 (3.1, 20.4)</td>
<td>Feb 26 (Feb 19, Mar 5)</td>
</tr>
<tr>
<td>(2)</td>
<td>$R_0$ reduced to 1.5</td>
<td>39 (10, 133)</td>
<td>1702 (639, 3249)</td>
<td>3.8 (1.7, 8.1)</td>
<td>Apr 8 (Mar 17, Apr 27)</td>
</tr>
<tr>
<td>(3)</td>
<td>both (1) and (2)</td>
<td>39 (8, 138)</td>
<td>1332 (481, 2506)</td>
<td>3.1 (1.0, 10.7)</td>
<td>Apr 6 (Mar 13, Apr 27)</td>
</tr>
</tbody>
</table>