bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Shi Zhao[1,2,*], Jinjun Ran[3], Salihu S Musa[4], Guangpu Yang[5,6], Yijun Lou[4], Daozhou Gao[7], Lin Yang[8] and Daihai He[4,*]

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 School of Public Health, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China
4 Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, China
5 Department of Orthopaedics and Traumatology, Chinese University of Hong Kong, Hong Kong, China
6 SH Ho Scoliosis Research Lab, Joint Scoliosis Research Center of Chinese University of Hong Kong and Nanjing University, Hong Kong, China
7 Department of Mathematics, Shanghai Normal University, Shanghai, China
8 School of Nursing, Hong Kong Polytechnic University, Hong Kong, China

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

Email address of all authors
SZ: zhaoshi.cmsa@gmail.com
JR: jimran@connect.hku.hk
SSM: salihu-sabiu.musa@connect.polyu.hk
GY: kennethgpy@link.cuhk.edu.hk
DG: dzgao@shnu.edu.cn
YL: yijun.lou@polyu.edu.hk
LY: l.yang@polyu.edu.hk
DH: daihai.he@polyu.edu.hk

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Declarations

Ethics approval and consent to participate

The ethical approval or individual consent was not applicable.

Availability of data and materials

All data and materials used in this work were publicly available.

Consent for publication

Not applicable.

Funding

This work was not funded.

Acknowledgements

The authors would like to acknowledge anonymous colleagues for helpful comments.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Backgrounds: There has been a novel coronavirus (2019-nCoV) pneumonia outbreak in China since December 2019, and which spreads internationally. This is the first study to quantify the basic reproduction number, R_0, of 2019-nCoV in the early phase of the outbreak.

Methods: Accounting for the impact of the variations in disease reporting rate, we modelled the epidemic curve of 2019-nCoV cases time series, in mainland China from January 10 to January 21, 2020, through the exponential growth. With the estimated intrinsic growth rate (γ), we estimated R_0 by using the serial intervals (SI) of two other well-known coronavirus diseases, MERS and SARS, as approximations for the true unknown SI.

Findings: The early outbreak data largely follows the exponential growth. We estimated that the mean R_0 ranges from 3.30 (95%CI: 2.73-3.96) to 5.47 (95%CI: 4.16-7.10) associated with 0-fold to 2-fold increase in the reporting rate. With rising report rate, the mean R_0 is likely to be below 5 but above 3.

Conclusion: The mean estimate of R_0 for the 2019-nCoV ranges from 3.30 (95%CI: 2.73-3.96) to 5.47 (95%CI: 4.16-7.10), and significantly larger than 1. Our findings indicate the potential of 2019-nCoV to cause outbreaks.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The atypical pneumonia case, caused by a novel coronavirus (2019-nCoV), was first suspected and reported in Wuhan, China in December 31, 2019 [1]. By the end of January 22, 2020, there have been 571 confirmed cases of 2019-nCoV infections in mainland China, including 17 deaths [2]. The 2019-nCoV cases were also reported in Thailand, Japan, Republic of Korea, Hong Kong, Taiwan and the US, and all these cases were exported from China, see WHO news release https://www.who.int/csr/don/en/ from January 14-21. The outbreak is still on-going. A recently published preprint by Imai *et al*. estimated that a total of 1723 (95%CI: 427-4471) cases of nCoV infections in Wuhan had onset of symptoms by January 12, 2020 [3]. The likelihood of travel related risks of disease spreading is suggested by [4], which indicates the potentials of regional and global spread [5].

To the best of our knowledge, there is no existing peer-reviewed literature quantifying the transmissibility of 2019-nCoV as of January 22, 2020. In this study, we estimated the transmissibility of 2019-nCoV via the basic reproduction number, R_0, based on the limited data in the early phase of the outbreak.

Methods

We obtained the number of 2019-nCoV cases time series data in mainland China released by Wuhan Municipal Health Commission, China and National Health Commission of China from January 10 to January 22, 2020 from [6]. All cases were laboratory confirmed following the case definition by national health commission of China [7].

Although there were cases confirmed on or before January 16, the official diagnostic protocol was released by WHO on January 17 [8]. To adjust the impact of this event, we considered a time-varying reporting rate that follows a linear increasing trend, motivated by the previous study [9]. We assumed that the reporting rate, $r(t)$, started increasing since January 17, and stopped at the maximal level at January 20. Denoting the daily reported number of new cases by $c(t)$ for the t-th day, then the adjusted cumulative number of cases, $C(t)$, is $C(t) = \sum_{\tau=0}^{t} c(\tau)/r(\tau)$. Instead of finding the exact value of $r(t)$, we calculated the fold change in $r(t)$ that is defined by the ratio of r on January 10 over that on January 22 minus 1. We considered four scenarios for further analysis with 0-fold (no change), 0.5-fold, 1-fold and 2-fold increase in reporting rate, see Fig 1(a), (c), (e) and (g).

Following previous studies [10, 11], we modelled the epidemic curve obeying the exponential growth. The nonlinear least square (NLS) framework is adopted for data fitting and parameter estimation. The intrinsic growth rate (γ) of the exponential growth was estimated, and the basic reproduction number could be obtained by $R_0 = 1/M(-\gamma)$ with 100% susceptibility for 2019-nCoV at this early stage. The function $M(\cdot)$ is the Laplace transform, i.e., the moment generating function, of the probability distribution for the serial interval (SI) of the disease [10, 12], denoted by $h(k)$ and k is the mean SI. Since the transmission chain of 2019-nCoV remains unclear, we adopted the SI information from other well-known coronavirus diseases, i.e., the middle east respiratory syndrome (MERS), and the severe acute respiratory syndrome (SARS), which share the similar pathogen as 2019-nCoV. We modelled $h(k)$ as Gamma distributions with mean of 7.6 days and standard deviation (SD) of 3.4 days for MERS [13], and mean of 8.4 days and SD of 3.8 days for SARS [14] as well as their average, see the row heads in Table 1 for each scenario.

Results and discussion

The exponential growth fitting results are shown in Fig 1(b), (d), (f) and (h). The coefficient of determination, R-squared, ranges from 0.86 to 0.87 for all reporting rate changing scenarios, which implies that the early outbreak data were largely following the exponential growth. In Table1, we estimated that the R_0 ranges from 3.30 (95%CI: 2.73-3.96) to 5.47 (95%CI: 4.16-7.10) associated with 0-fold to 2-fold increase in the reporting

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rate. All R_0 estimates are significantly larger than 1, which indicates the potential of 2019-nCoV to cause outbreaks. Since the official diagnostic protocol was released by WHO on January 17 [8], an increase in the diagnosis and reporting of 2019-nCoV infections probably occurred. Thereafter, the daily number of newly reported cases started increasing around January 17, see Fig 1, which implies that more infections were likely being diagnosed and recorded. We suggested that changing in reporting might exist thus should be considered in estimation. Although four scenarios about the reporting rate were explored in this study, the real situation is difficult to determine given limited data and (almost) equivalent model fitting performance in terms of R-squared. However, with the rising in the reporting rate, we found the mean R_0 is likely to be below 5 but above 3.

Our analysis and estimation of R_0 rely on the accuracy of the SI of 2019-nCoV, which remains unknown as of January 21. In this work, we employed the SIs of SARS and MERS as approximations to that of 2019-nCoV. The determination of SI requires the knowledge of the chain of disease transmission that needs sufficient number of patient samples and periods of time for follow-up [15], and thus this is unlikely to be achieved shortly. However, using SIs of SARS and MERS as approximation could provide an insight to the transmission potential of 2019-nCoV at the early stage of the outbreak. We reported that the mean R_0 of 2019-nCoV is likely to be from 3.3 to 5.5, and it appeared (slightly) higher than those of SARS, i.e., 2-5 [14, 16], and MERS, i.e., 2.7-3.9 [17]. Furthermore, measures on 'entry, exit and travel' screening and restriction started today (on January 23) in Wuhan is likely to reduce the R_0, which is benefit for the 2019-nCoV control and prevention.

Conclusion

We estimated the mean R_0 of 2019-nCoV ranging from 3.30 (95%CI: 2.73-3.96) to 5.47 (95%CI: 4.16-7.10) and being 3.98 (95%CI: 3.19-4.92) if the reporting efforts doubled after the diagnostic protocol released on January 17, 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. **'Pneumonia of unknown cause – China', Emergencies preparedness, response, Disease outbreak news, World Health Organization (WHO)** [https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/]
2. **'Situation report of the pneumonia cases caused by the novel coronavirus', released on January 22, 2020 by the National Health Commission of the People's Republic of China (in Chinese).** [http://www.nhc.gov.cn/yjb/s3578/202001/a3c8b5144067417889d8760254b1a7ca.shtml]
3. Imai N, Dorigatti I, Cori A, Riley S, Ferguson NM: **Estimating the potential total number of novel Coronavirus (2019-nCoV) cases in Wuhan City, China**. *Preprint published by the Imperial College London* 2020:https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news--wuhan-coronavirus/.
4. Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K: **Pneumonia of unknown etiology in Wuhan, China: potential for international spread via commercial air travel**. *Journal of Travel Medicine* 2020:doi:10.1093/jtm/taaa1008.
5. Leung K, Wu JT, Leung GM: **Nowcasting and forecasting the Wuhan 2019-nCoV outbreak** *Preprint published by the School of Public Health of the University of Hong Kong* 2020:https://files.sph.hku.hk/download/wuhan_exportation_preprint.pdf.
6. **'News press and situation reports of the pneumonia caused by novel coronavirus', from December 31, 2019 to Janunary 21, 2020 released by the Wuhan municipal health commission, China** [http://wjw.wuhan.gov.cn/front/web/list2nd/no/710]
7. **'Definition of suspected cases of unexplained pneumonia', the National Health Commission of the People's Republic of China (in Chinese).** [http://www.nhc.gov.cn/]
8. **Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases, World Health Organization (WHO)** [https://www.who.int/health-topics/coronavirus/laboratory-diagnostics-for-novel-coronavirus]
9. Wu JT, Cowling BJ, Lau EH, Ip DK, Ho L-M, Tsang T, Chuang S-K, Leung P-Y, Lo S-V, Liu S-H: **School closure and mitigation of pandemic (H1N1) 2009, Hong Kong**. *Emerging infectious diseases* 2010, **16**(3):538.
10. Zhao S, Musa SS, Fu H, He D, Qin J: **Simple framework for real-time forecast in a data-limited situation: the Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example**. *Parasites and Vectors* 2019, **12**(1):344.
11. de Silva U, Warachit J, Waicharoen S, Chittaganpitch M: **A preliminary analysis of the epidemiology of influenza A (H1N1) v virus infection in Thailand from early outbreak data, June-July 2009**. *Eurosurveillance* 2009, **14**(31):19292.
12. Wallinga J, Lipsitch M: **How generation intervals shape the relationship between growth rates and reproductive numbers**. *Proceedings of the Royal Society B: Biological Sciences* 2007, **274**(1609):599-604.
13. Assiri A, McGeer A, Perl TM, Price CS, Al Rabeeah AA, Cummings DA, Alabdullatif ZN, Assad M, Almulhim A, Makhdoom H: **Hospital outbreak of Middle East respiratory syndrome coronavirus**. *New England Journal of Medicine* 2013, **369**(5):407-416.
14. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, Gopalakrishna G, Chew SK, Tan CC, Samore MH: **Transmission dynamics and control of severe acute respiratory syndrome**. *Science* 2003, **300**(5627):1966-1970.
15. Cowling BJ, Fang VJ, Riley S, Peiris JM, Leung GM: **Estimation of the serial interval of influenza**. *Epidemiology* 2009, **20**(3):344.
16. Wallinga J, Teunis P: **Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures**. *American Journal of epidemiology* 2004, **160**(6):509-516.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

17. Lin Q, Chiu AP, Zhao S, He D: **Modeling the spread of Middle East respiratory syndrome coronavirus in Saudi Arabia**. *Statistical methods in medical research* 2018, **27**(7):1968-1978.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure

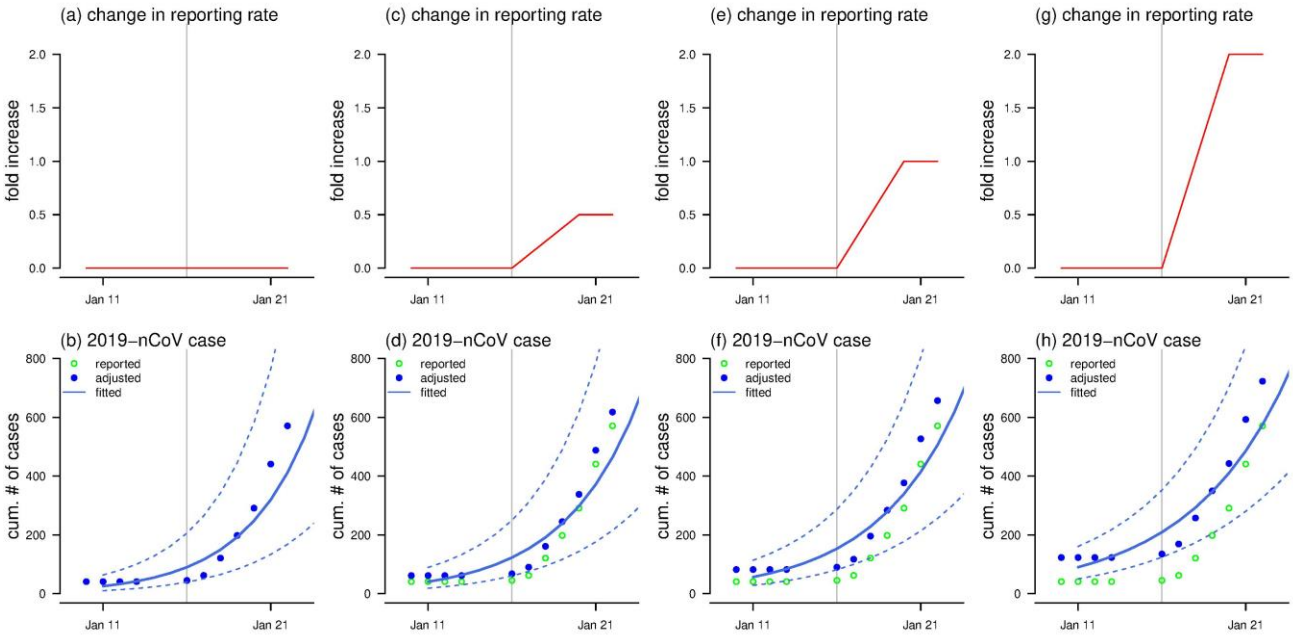

Figure 1

The scenarios of the change in the reporting rate (top panels) and the exponential growth fitting (bottom panels). The top panels, i.e., (a), (c), (e) and (g), show the assumed change in the reporting rate. The bottom panels, i.e., (b), (d), (f) and (h), show the reported (or observed, green circles), adjusted (blue dots) and fitted (blue curve) number of 2019-nCoV infections, and the blue dashed lines are the 95%CI. The vertical grey line represents the date of January 16, 2020, after which the official diagnostic protocol was released by WHO [8]. Panels (a) and (b) show the scenarios that the reporting rate was unchanged. Panels (c) and (d) show the scenarios that the reporting rate increased by 0.5-fold. Panels (e) and (f) show the scenarios that the reporting rate increased by 1-fold. Panels (g) and (h) show the scenarios that the reporting rate increased by 2-fold.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.916395; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table

Table 1

The summary table of the estimated basic reproduction number, R_0, under different scenarios. The estimated R_0 is shown as in the 'median (95%CI)' format. The 'reporting rate increased' indicated the number of fold increase in the reporting rate from January 17, when WHO released the official diagnostic protocol [8], to January 20, 2020.

Reporting rate increased	Estimated R_0		
	same as MERS SI 7.6 ±3.4	SI in average 8.0 ±3.6	same as SARS SI 8.4 ±3.8
(unchanged)	5.10 (3.92-6.57)	5.47 (4.16-7.10)	5.86 (4.41-7.66)
0.5-fold	4.27 (3.38-5.35)	4.55 (3.57-5.74)	4.84 (3.76-6.15)
1-fold	3.75 (3.03-4.61)	3.98 (3.19-4.92)	4.22 (3.36-5.25)
2-fold	3.13 (2.61-3.73)	3.30 (2.73-3.96)	3.47 (2.85-4.19)

Note: The 'SI' is serial interval. The 'MERS' is the middle east respiratory syndrome, and the 'SARS' is the severe acute respiratory syndrome.