Ascertainment rate of novel coronavirus disease (COVID-19) in Japan

Ryosuke Omori[a], Kenji Mizumoto[b,c], Hiroshi Nishiura[d,e,*]

[a] Research Center for Zoonosis Control, Hokkaido University, Kita-ku, Sapporo-shi, Hokkaido 001-0020, Japan
[b] Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University, Yoshida-Nakaadachi-cho, Sakyo-ku, Kyoto, Japan
[c] Hakubi Center for Advanced Research, Kyoto University Yoshidahonmachi, Sakyo-ku, Kyoto, Japan
[d] Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome, Kita-ku, Sapporo-shi, Hokkaido 060-8638, Japan
[e] Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama 332-0012, Japan

ABSTRACT

Objective: To estimate the ascertainment rate of novel coronavirus disease (COVID-19).
Methods: The epidemiological dataset of confirmed cases with COVID-19 in Japan as of February 28, 2020 was analyzed. A statistical model was constructed to describe the heterogeneity of the reporting rate by age and severity. We estimated the number of severe and non-severe cases, accounting for under-ascertainment.
Results: The ascertainment rate of non-severe cases was estimated at 0.44 (95% confidence interval 0.37–0.50), indicating that the unbiased number of non-severe cases would be more than twice the reported count.
Conclusions: Severe cases are twice as likely to be diagnosed and reported when compared to other cases. Considering that reported cases are usually dominated by non-severe cases, the adjusted total number of cases is also approximately double the observed count. This finding is critical in interpreting the reported data, and it is advised that the mild case data for COVID-19 should always be interpreted as under-ascertained [Au?1].

1. Introduction

As of March 1, 2020, a total of 58 countries had reported at least one confirmed case of novel coronavirus disease (COVID-19), and the cumulative number of deaths had reached 2977 persons across the world (WHO, 2020). To achieve appropriate countermeasures, it is vital to understand the current epidemiological situations of the COVID-19 epidemic.

The majority of COVID-19 cases exhibit a disease of limited severity; 81% of reported cases in China have been mild and only 16% have been severe (Guan et al., 2020). It is expected that the ascertainment rate will be different between severe and non-severe cases. The aim of this study was to estimate the ascertainment rate of non-severe cases, employing a statistical model.

2. Methods

The epidemiological dataset of confirmed cases with COVID-19 in Japan as of February 28, 2020 was analyzed. A confirmatory diagnosis was made by means of reverse transcriptase polymerase chain reaction (RT-PCR). This study specifically analyzed cases by prefecture, age, and severity. A severe case was defined as (1) a patient with severe dyspnea that required oxygen support plus pneumonia or intubation, or (2) a patient who required management in the intensive care unit.

The number of severe and non-severe cases was estimated using the ratio of non-severe to severe reported cases (Guan et al., 2020; Novel, 2020). We estimated the ascertainment rate among non-severe cases by $1/k$, describing the data of both severe and non-severe cases generated by Poisson process, with probabilities $p_{x,a}$ for severe cases and $kf_a p_{x,a}$ for non-severe cases in age group a and prefecture x [Au?1]. Here f_a denotes the ratio of non-severe to

* Corresponding author. Graduate School of Medicine, Hokkaido University, Kia 15 Jo Nishi 7 Chome, Kitaku, Sapporo, 060-8638, Japan. Tel.: +81 11 706 5066; fax: +81 11 706 7819.
E-mail address: nishiurah@med.hokudai.ac.jp (H. Nishiura).

severe reported cases in age group a, as estimated from the age-specific severity and incidence rate ratio in China (Guan et al., 2020; Novel, 2020). k and $p_{x,a}$ were estimated using the log likelihood function:

$$ll = \sum_{x} \sum_{a} \ln \left[\frac{(N_{x,a}kf_ap_{x,a})^{D_{ns,x,a}}\exp(-N_{x,a}kf_ap_{x,a})}{D_{ns,x,a}!} \frac{(N_{x,a}p_{x,a})^{D_{s,x,a}}\exp(-N_{x,a}p_{x,a})}{D_{s,x,a}!} \right], \quad (1)$$

where $N_{x,a}$, $D_{ns,x,a}$, and $D_{ns,x,a}$ represent the population size and the observed counts of non-severe and severe cases, respectively, in age group a in prefecture x [Au?2]. Maximum likelihood estimates were obtained by maximizing equation (1), and the profile likelihood-based confidence intervals (CI) were computed.

3. Results

The ascertainment rate of non-severe cases, k, was estimated at 0.44 (95% CI 0.37–0.50). The resulting estimate of non-severe cases is shown in Figure 1A. This estimate showed a reasonably good fit to severe case data as shown in Figure 1B [Au?1]. The age-specific pattern of estimated non-severe cases was similar to that among severe cases. The largest estimated number of non-severe cases was 80 cases (95% CI 63–98) among those aged 50–59 years and 78 (95% CI 61–95) among those aged 60–69 years. This adjustment gives an adjusted estimate of the total cases by age group.

4. Discussion

This study estimated the ascertainment-adjusted number of cases in Japan, using age-specific severe fractions of cases. It was assumed that the ratio of severe to non-severe cases in a given age group is a constant and that the age-independent gap is explained by the under-diagnosis and under-reporting. In this way, the ascertainment rate of non-severe cases was estimated to be 0.44 [Au?1].

As a take home, it must be remembered that severe cases are twice as likely to be diagnosed and reported when compared to other cases. Reported cases are usually dominated by non-severe cases, and the adjusted total number of cases is about double the observed count. This finding is critical in interpreting the reported data, and it is advised that the mild case data for COVID-19 should always be interpreted as under-ascertained [Au?1].

In addition to the proposed adjustment, it should be noted that the ascertainment rate of severe cases also needs to be
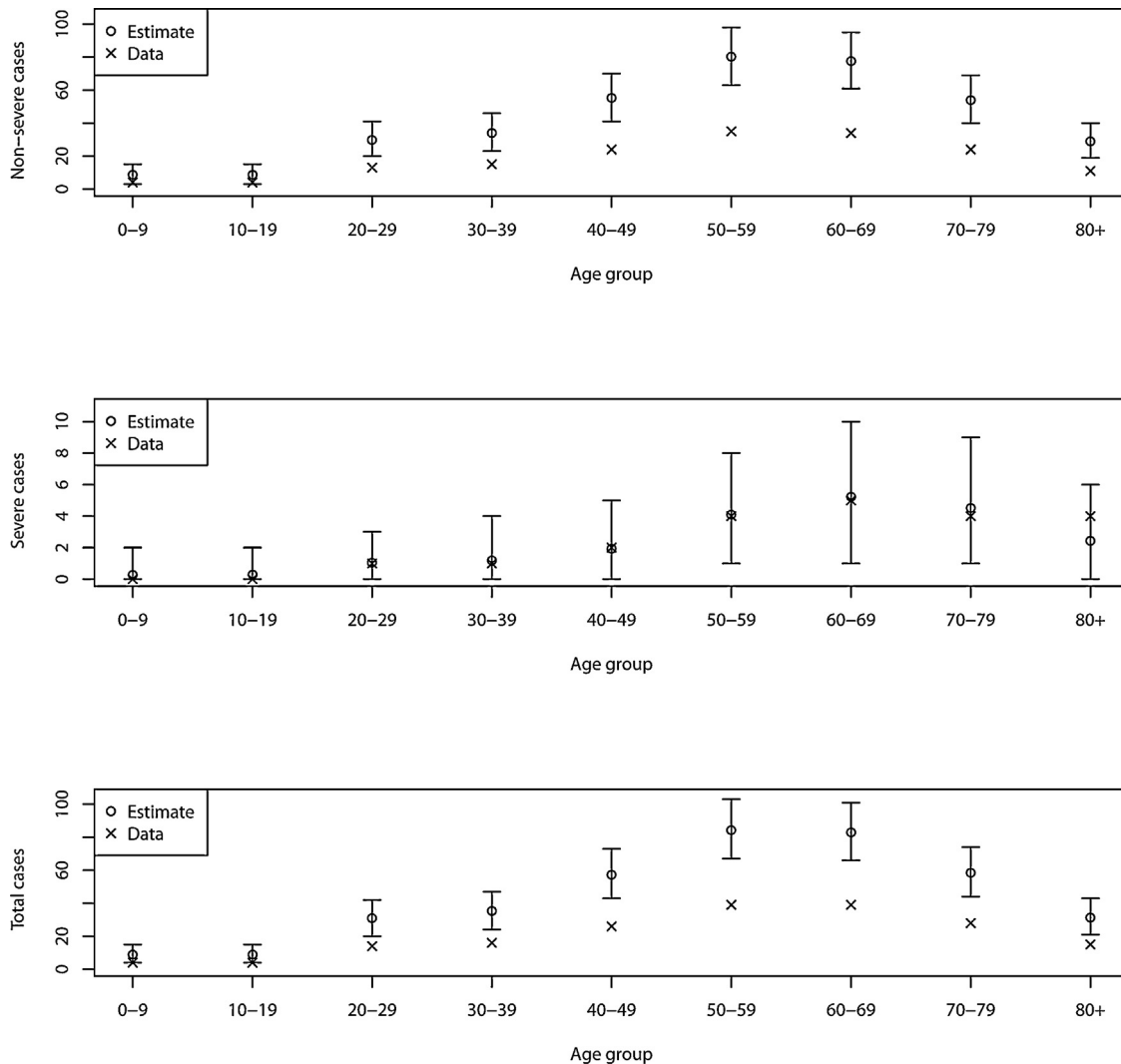

Figure 1. Age-specific number of novel coronavirus disease (COVID-19) cases by age group and severity. [Au?3]
Top: non-severe cases, middle: severe cases, and bottom: total cases. The x-marks represent observed counts, while unfilled circles show estimated cases. Whiskers extend to lower and upper 95% confidence intervals, derived from profile likelihood.

estimated, and this estimation requires direct measurement of the total number of cases or infected individuals by means of a sero-epidemiological study or other testing methods of all samples (Nishiura et al., 2020). In other words, the actual total number of cases is greater than the number that was adjusted in the present study [Au?1]. Using sero-epidemiological data-sets, we plan to address relevant issues in the future. Other limitations of this study include the following: (1) we did not explore the detailed natural history, e.g., dynamically changing symptoms over the course of infection, or underlying comorbidities, (2) we ignored right-censored data, e.g., the time delay from illness onset to severe manifestations, for simplicity. The latter led us to underestimate the ascertainment rate. Furthermore, (3) it is worth noting that the data of age-dependent severity employed in the analysis are based only on the observed data in China. Considering the possibility of under-reporting or biased age distribution, the nature of this age distribution may lead to an underestimation. (4) The dataset included asymptomatic test-positive individuals, while Guan et al. (2020) analyzed only hospitalized cases. (5) The ascertainment rate can be influenced by changes in reporting rate. As both the daily number of tests and the positivity rate during the study period remained constant (Ministry of Health, 2020), it can be considered that reporting bias is minimal.

Despite multiple future tasks, we believe that the study reported here successfully demonstrated that the ascertainment rate can be partly adjusted by examining age-dependent numbers of cases including severe cases. The proposed adjustment should be practiced in other country settings and also for other diseases.

5. Declarations

Ethical approval: This study was based on publicly available data and did not require ethical approval.

Conflict of interest: The authors declare no conflicts of interest.

References

Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine 2020;.

Nishiura H, Kobayashi T, Yang Y, Hayashi K, Miyama T, Kinoshita R, et al. The rate of underascertainment of novel coronavirus (2019-nCoV) infection: Estimation using Japanese passengers data on evacuation flights. Multidisciplinary Digital Publishing Institute;; 2020.

Novel CPERE. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi 2020;41(2):145.

WHO. Coronavirus disease 2019 (COVID-19) Situation Report – 41. 2020 Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200301-sitrep-41-covid-19.pdf?sfvrsn=6768306d_2. [Accessed 2nd March 2020]..

Ministry of Health, Labour and Welfare, Japan. Available from https://www.mhlw.go.jp/stf/houdou/index.html [Accessed 20th April 2020] (in Japanse)