

1 **Title: Ascertainment rate of novel coronavirus disease (COVID-**
2 **19) in Japan**

3 **Running title:** Ascertainment in Japan

4 **Authors:**

5 Ryosuke Omori¹, Kenji Mizumoto^{2,3}, Hiroshi Nishiura^{4,5}

6 **Affiliations:**

7 ¹ Research Center for Zoonosis Control, Hokkaido University, Kita-ku, Sapporo-
8 shi, Hokkaido 001-0020, Japan;

9 ² Graduate School of Advanced Integrated Studies in Human Survivability,
10 Kyoto University Yoshida-Nakaadachi-cho, Sakyo-ku, Kyoto, Japan

11 ³ Hakubi Center for Advanced Research, Kyoto University, Yoshidahonmachi,
12 Sakyo-ku, Kyoto, Japan;

13 ⁴ Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome,
14 Kita-ku, Sapporo-shi, Hokkaido 060-8638, Japan;

15 ⁵ Core Research for Evolutional Science and Technology (CREST), Japan
16 Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama 332-0012,
17 Japan

18 **Correspondence to:**

19 H. Nishiura, Graduate School of Medicine, Hokkaido University, Kia 15 Jo Nishi
20 7 Chome, Kitaku, Sapporo, 060-8638, Japan

21 Tel: +81 11 706 5066 Fax: +81 11 706 7819

22 E-mail: nishiurah@med.hokudai.ac.jp

23 **Article type:**

24 Short communications

25

1 **Abstract**

2 **Objective:** To estimate the ascertainment rate of novel coronavirus (COVID-19).

3 **Methods:** We analyzed the epidemiological dataset of confirmed cases with
4 COVID-19 in Japan as of 28 February 2020. A statistical model was constructed
5 to describe the heterogeneity of reporting rate by age and severity. We estimated
6 the number of severe and non-severe cases, accounting for under-ascertainment.

7 **Results:** The ascertainment rate of non-severe cases was estimated at 0.44 (95%
8 confidence interval: 0.37, 0.50), indicating that unbiased number of non-cases
9 would be more than twice the reported count.

10 **Conclusions:** Severe cases are twice more likely diagnosed and reported than
11 other cases. Considering that reported cases are usually dominated by non-severe
12 cases, the adjusted total number of cases is also about a double of observed count.
13 Our finding is critical in interpreting the reported data, and it is advised to
14 interpret mild case data of COVID-19 as always under-ascertained.

15 **Keywords:** coronavirus; outbreak; diagnosis; reporting; statistical model;
16 epidemiology; viruses

17

18 **Highlights**

- 19 - Epidemiological dataset of COVID-19 in Japan was analyzed.
20 - The ascertainment rate of non-severe cases was estimated at 0.44 (95%
21 confidence interval: 0.37, 0.50).
22 - Severe cases are twice more likely diagnosed and reported than other cases.
23 - Mild cases of COVID-19 are under-ascertained.

24

1 **Introduction**

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

7 The majority of COVID-19 cases exhibit limited severity; 81% of
8 reported cases in China has been mild and only 16% are severe (Guan et al.,
9 2020). It is natural that the ascertainment rate would be different between severe
10 and non-severe cases. The present study aims to estimate the ascertainment rate
11 of non-severe cases, employing a statistical model.

12 **Methods**

13 We analyzed the epidemiological dataset of confirmed cases with
14 COVID-19 in Japan as of 28 February 2020. The confirmatory diagnosis was
15 made by means of reverse transcriptase polymerase chain reaction (RT-PCR).
16 The present study specifically analyzed cases by (i) prefecture, (ii) age, and (iii)
17 severity. Severe case was defined as (i) severe dyspnea that required oxygen
18 support plus pneumonia or intubation or (ii) case that required management in
19 intensive care unit.

20 We estimated the number of severe and non-severe cases using the ratio
21 of non-severe to severe reported cases (Guan et al., 2020, Novel, 2020). We
22 estimated the ascertainment rate among non-severe cases by $1/k$, describing data
23 generating process of both severe and non-severe generated from Poisson process
24 with probabilities $p_{x,a}$ for severe cases and $kf_a p_{x,a}$ for non-severe cases in age

1 group a and prefecture x , respectively. Here f_a denotes the ratio of non-severe to
2 severe reported case of age group a , as estimated from age-specific severity and
3 incidence rate ratio in China (Guan et al., 2020, Novel, 2020). We estimate k and
4 $p_{x,a}$ using the loglikelihood function:

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

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

9 **Results**

10 The ascertainment rate of non-severe cases, k , was estimated at 0.44 (95%
11 confidence interval (CI): 0.37, 0.50). Resulting estimate of non-severe cases is
12 shown in Figure 1A, showing along with reasonably good fit to severe case data
13 in Figure 1B. Age-specific pattern of estimated non-severe cases was similar to
14 that among severe cases. The largest estimated number of non-severe cases was
15 80 cases (95% CI: 63, 98) among those aged 50-59 years and 78 (95% CI: 61,
16 95) among cases aged 60-69 years, respectively. Such adjustment gives adjusted
17 estimate of the total cases by age group.

18 **Discussion**

19 The present study estimated the ascertainment-adjusted number of cases in
20 Japan, using age-specific severe fraction of cases. We assumed that the ratio of
21 severe to non-severe cases in a given age group is a constant and that the age-
22 independent gap is explained by the under-diagnosis and under-reporting,
23 estimating the ascertainment rate among non-severe cases to be 0.44.

1 As a take home, it must be remembered that severe cases are twice more
2 likely diagnosed and reported than other cases. Reported cases are usually
3 dominated by non-severe cases, and the adjusted total number of cases is about a
4 double of observed count. Our finding is critical in interpreting the reported data,
5 and it is advised to regard the mild case data as always under-ascertained.

6 In addition to the proposed adjustment, it should be noted that the
7 ascertainment rate of severe cases needs to be additionally estimated, and such
8 estimation requires direct measurement of the total number of cases or infected
9 individuals by means of seroepidemiological study or other testing methods of all
10 samples (Nishiura et al., 2020). That is, the actual total number of cases is greater
11 than what it was adjusted in the present study. Using seroepidemiological
12 datasets, we plan to address relevant issues in the future. Other limitations
13 include that (i) we did not explore detailed natural history, e.g. dynamically
14 changing symptoms over the course of infection, and underlying comorbidities,
15 (ii) we ignored right-censored data, e.g. the time delay from illness onset to
16 severe manifestations, for simplicity. The latter led us to underestimate the
17 ascertainment rate. (iii) it is worth noting that the data of age dependent severity
18 employed in our analysis is only based on the observed data in China.
19 Considering the possibility of underreporting or biased age distribution, the
20 nature of this age distribution may lead to underestimation.

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

25

1 **Conflict of interest**

2 The authors declare no conflicts of interest.

3 **Funding source**

4 H.N. received funding support from Japan Agency for Medical Research and
5 Development [grant number: JP18fk0108050] the Japan Society for the
6 Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI
7 in Japanese abbreviation) grant nos. 17H04701, 17H05808, 18H04895 and
8 19H01074, and the Japan Science and Technology Agency (JST) Core Research
9 for Evolutional Science and Technology (CREST) program [grant number:
10 JPMJCR1413]. The funders had no role in study design, data collection and
11 analysis, decision to publish, or preparation of the manuscript.

12 **Ethical approval**

13 This study was based on publicly available data and did not require ethical
14 approval.

15

16 **References**

- 17 1. Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, et al. Clinical
18 Characteristics of Coronavirus Disease 2019 in China. *New England Journal*
19 of Medicine 2020.
20 2. Nishiura H, Kobayashi T, Yang Y, Hayashi K, Miyama T, Kinoshita R, et al.
21 The rate of underascertainment of novel coronavirus (2019-nCoV) infection:
22 Estimation using Japanese passengers data on evacuation flights.
23 Multidisciplinary Digital Publishing Institute; 2020.
24 3. Novel CPERE. The epidemiological characteristics of an outbreak of 2019
25 novel coronavirus diseases (COVID-19) in China. *Zhonghua liu xing bing xue*
26 za zhi= *Zhonghua liuxingbingxue zazhi* 2020;41(2):145.
27 4. WHO. Coronavirus disease 2019 (COVID-19) Situation Report – 41; 2020.
28 Available from: [https://www.who.int/docs/default-](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200301-sitrep-41-covid-19.pdf?sfvrsn=6768306d_2)
29 [source/coronaviruse/situation-reports/20200301-sitrep-41-covid-](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200301-sitrep-41-covid-19.pdf?sfvrsn=6768306d_2)
30 [19.pdf?sfvrsn=6768306d_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200301-sitrep-41-covid-19.pdf?sfvrsn=6768306d_2). [Accessed 2nd March 2020].

1

2 **Figure Legends**

3 **Figure 1. Age-specific number of novel coronavirus disease (COVID-19)**

4 **cases by age group and severity**

5 Top: non severe cases, middle: severe cases, and bottom: total cases. x-marks

6 represent observed counts, while unfilled circles show estimated cases. Whiskers

7 extend to lower and upper 95% confidence intervals, derived from profile

8 likelihood.

9

