

Evaluation of the representativeness of data in the COVID-19 Registry Japan during the first six waves of the epidemic

Kayoko Hayakawa^{1,2,*}, Yusuke Asai², Nobuaki Matsunaga², Shinya Tsuzuki^{2,3}, Mari Terada^{1,4}, Setsuko Suzuki¹, Koji Kitajima⁴, Sho Saito¹, Norio Ohmagari^{1,2}

¹ Disease Control and Prevention Center, National Center for Global Health and Medicine, Tokyo, Japan;

² AMR Clinical Reference Center, National Center for Global Health and Medicine, Tokyo, Japan;

³ Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium;

⁴ Center for Clinical Sciences, National Center for Global Health and Medicine, Tokyo, Japan.

Abstract: The COVID-19 Registry Japan (COVIREGI-JP), a registry of patients hospitalized with coronavirus disease (COVID-19), contains the largest national COVID-19 inpatient population. Since COVIREGI-JP invites voluntary participation by facilities, selection bias is inevitable. The current study examined the representativeness of COVIREGI-JP data in comparison to open-source national data. The number of infections and deaths among hospitalized COVID-19 patients in COVIREGI-JP were compared to those in national data recorded during the six waves of the COVID-19 epidemic until March 6, 2022. During the period studied, patients in COVIREGI-JP represented 1% of the total COVID-19 cases according to national data; the proportion was high during the first wave (32.7%) and tended to decrease, especially after the fourth wave. The overall proportion of patients from each region varied from 0.8% to 2.5%, but case fatality rates in COVIREGI-JP tended to be higher than those in the national data, with the exception of a few waves, in several regions. The difference was smallest during the first wave. Although COVIREGI-JP consistently registered cases from all regions of the country, the proportion tended to decline after the beginning of the epidemic. Given the epidemiological persistence and the ever-changing epidemiology of COVID-19, continued case registration and data utilization in COVIREGI-JP is desirable, although selection bias in COVIREGI-JP registration of cases should be carefully interpreted.

Keywords: COVIREGI-JP, epidemiology of COVID-19, pandemic, selection bias

Introduction

The number of cumulative cases and hospitalizations due to coronavirus disease (COVID-19) continues to increase in Japan (1,2). The COVID-19 Registry Japan (COVIREGI-JP), a registry of hospitalized COVID-19 patients, was started on March 2, 2020 and has, with the cooperation of various facilities, accumulated detailed information on hospitalized COVID-19 patients in Japan since the beginning of the COVID-19 pandemic (3). Detailed clinical information (clinical epidemiology, comorbidities, vital signs, pharmacotherapy and supportive care, complications, detailed prognosis, laboratory findings, *etc.*) that is unavailable in publicly accessible government-reported data is available in COVIREGI-JP and has been utilized in various studies as well as for public policy formulation and implementation (4-6). Moreover, some of the published data have been used to formulate the national COVID-19 treatment guidelines (7-9).

The fact that COVIREGI-JP contains the largest

COVID-19 inpatient population in the country is a major advantage. However, various limitations of COVIREGI-JP data have previously been discussed (10), and especially the inevitable selection bias, because COVIREGI-JP is a voluntary registry that involves the participation of many facilities. One needs to understand the biases that affect COVIREGI-JP data before interpreting and utilizing the data from this registry. This review has examined the representativeness of COVIREGI-JP data in comparison to open-source national data.

Study design

Data sets

This retrospective study used data aggregated in COVIREGI-JP and the national open database (2). To determine the extent to which the COVIREGI data are nationally representative of Japan, the number of infections and deaths among hospitalized patients

registered in COVIREGI-JP were compared to those in Japan (*i.e.*, national data). Moreover, the total number of hospitals (11) and the facilities participating in COVIREGI-JP were summarized. For region-wise comparisons, 10 categories were obtained from the existing classifications. In this study, Tokyo and Osaka were not included in each regional category, but rather, were separately accounted for in 12 regions (12).

The details of COVIREGI-JP have been described previously (3,10). To ascertain the number of patients and deaths, data on patients for whom major items, including outcome at discharge, had been entered and finalized (*i.e.*, these items were finalized and the facility was unable to make any further modifications) by March 6, 2022 were used. National data were the data released by the Ministry of Health, Labour, and Welfare (MHLW) of Japan prior to March 7, 2022.

Waves of the COVID-19 epidemic

The waves of the COVID-19 epidemic were defined as follows using the admission date: *i*) first wave (Wave 1), January 1–May 31, 2020; *ii*) second wave (Wave 2), June 1–October 31, 2020; *iii*) third wave (Wave 3), November 1, 2020–March 31, 2021; *iv*) fourth wave (Wave 4), April 1–June 30, 2021; *v*) fifth wave (Wave 5), July 1–November 30, 2021; and *vi*) sixth wave (Wave 6), November 1, 2021–March 6, 2022.

For cases where the date of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) test was later than the date of admission (*e.g.*, in hospital-acquired COVID-19 cases), the test date was used instead of the admission date.

Statistical analysis

Data were analyzed descriptively and calculated for proportions, differences, and ratios. Differences in the mortality between the national and regional data were determined in two data sets (national data on total infections and COVIREGI-JP data). In the national data, the total value for each area did not match the national value due to a problem with the original data source. All statistical analyses were performed using R, version 3.5.1 (R Core Team, Vienna, Austria) and Microsoft Excel (2019).

Statistical analysis

Ethical review was considered unnecessary for this study because only aggregate data were used.

Characteristics, regional divisions, and the region-wise proportion of patients in COVIREGI-JP

Figure 1 separately shows the epi-curves of the number of infections reported nationwide and those of

COVIREGI-JP registrations. Data in COVIREGI-JP were consistently registered although fewer data tended to be registered after the fourth wave, and the number of registrations during the sixth wave of the epidemic was lower than that during other waves.

Figure 2 shows a map of the regional divisions and the region-wise proportion of patients (the number of COVIREGI-JP registrants/number of persons infected with SARS-CoV-2 according to national data). Patients registered in COVIREGI-JP accounted for 1% of the number of total SARS-CoV-2 infections according to national data during the period studied; this figure was initially high during the first wave (32.7%) and then tended to decrease.

The overall proportion of patients in each region varied from 0.8% (Tokyo, Kanto, Osaka) to 2.5% (Tohoku) but followed almost the same trend as the waves of the epidemic progressed. In Shikoku, an increase in the proportion of patients was noted during the third wave compared to the second wave.

Number of COVID-19 cases and deaths according to national data and COVIREGI-JP

Table 1a (Online Data, <https://www.globalhealthmedicine.com/site/supplementaldata.html?ID=50>) shows the number of cases and deaths according to national data and COVIREGI-JP data. The national data included 5,273,350 infections and 24,927 deaths; COVIREGI-JP included 54,350 cases and 2,512 deaths. In the national data, the largest number of cases was in Kanto ($n = 1,354,178$; 25.7% of all cases), followed by Tokyo ($n = 1,056,983$; 20.0%) and Osaka ($n = 677,489$; 12.8%). The Kanto region accounted for the largest number of deaths (5,398, or 21.7% of the total), followed by Osaka with 4,105 deaths (16.5%) and Tokyo with 3,809 deaths (15.3%).

In COVIREGI-JP data, the largest number of registrations were from the Kanto region ($n = 10,941$; 20.1% of all COVIREGI-JPs), followed by Tokyo ($n = 8,659$; 15.9%) and Kinki ($n = 6,714$; 12.4%). Analysis of COVIREGI-JP data indicated that the largest number of deaths occurred in Kanto ($n = 575$; 22.9%), followed by Tokyo ($n = 395$; 15.7%) and Osaka ($n = 361$; 14.4%).

Table 1b (Online Data, <https://www.globalhealthmedicine.com/site/supplementaldata.html?ID=50>) shows the ratio of case fatality rates for COVIREGI-JP data compared to that in national data. Overall, case fatality rates in COVIREGI-JP tended to be higher (overall, 1.4 times higher) than those in the national data, and the difference was smallest during the first wave. However, the case fatality rate in the national data was higher than that in COVIREGI-JP in Hokkaido, Tokai, and Osaka during the first wave and in Hokuriku during the fifth wave.

Table 1c (Online Data, <https://www.globalhealthmedicine.com/site/supplementaldata.html?ID=50>) shows the ratio of case fatality rates for COVIREGI-JP data compared to that in national data.

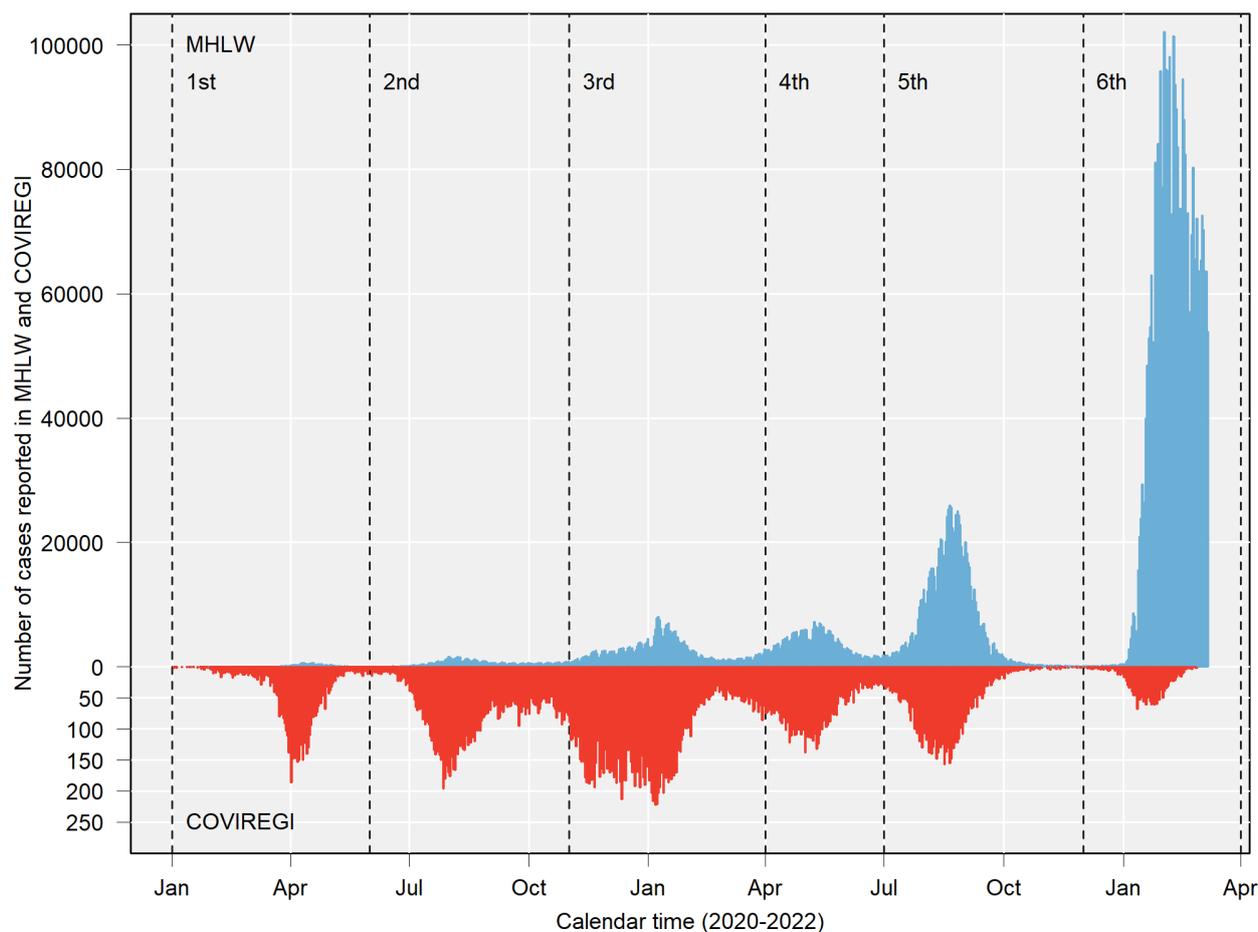


Figure 1. Epi-curves of the number of infections nationwide and the number of COVIREGI registrations.

[globalhealthmedicine.com/site/supplementaldata.html?ID=50](https://www.globalhealthmedicine.com/site/supplementaldata.html?ID=50)) shows the difference in the national and regional case fatality rates between the national data and COVIREGI-JP data. Positive values indicate that the case fatality rate in the region is higher than the national rate, whereas negative values indicate a lower rate. In most regions, the direction of the positive and negative values was consistent for both national and COVIREGI-JP data; however, there was a divergence in the overall rates in the Kanto, Kinki, and Kyushu regions when all waves were evaluated together. Moreover, there are some areas that diverged during each wave, and, there were five regions that diverged from each other during Wave 3.

Number of facilities that registered cases during the six waves and the number of patients from facilities that continued to register cases

Table 2 (Online Data, <https://www.globalhealthmedicine.com/site/supplementaldata.html?ID=50>) shows the number of facilities that registered cases during each wave and the number of patients from facilities that continued to register cases throughout the six waves. However, the number of facilities decreased significantly after the fourth wave. The percentage of patients

registered by the 33 facilities that continuously registered cases during the six waves was minimal during the first wave and gradually increased, especially after the fourth wave.

The representativeness of COVIREGI-JP in regions and the six waves of the epidemic

This review summarizes the representativeness of COVIREGI-JP from two perspectives: the region and each wave of the epidemic. Although the representativeness of COVIREGI-JP up to the third wave has previously been examined (13), this is the first time that the representativeness of COVIREGI-JP up to Wave 6 has been examined.

Patients were consistently registered in COVIREGI-JP, although a decrease was noted after the fourth wave. In waves with a large number of cases, such as waves 5 or 6, the proportion of patients decreased significantly. There were few COVIREGI-JP registrations during the sixth wave because the epidemic was still in progress, and the number of cases itself increased steeply, leading to a large decline in the proportion of patients. COVIREGI-JP commenced registrations on March 2, 2020, and permitted retrospective data entry. One

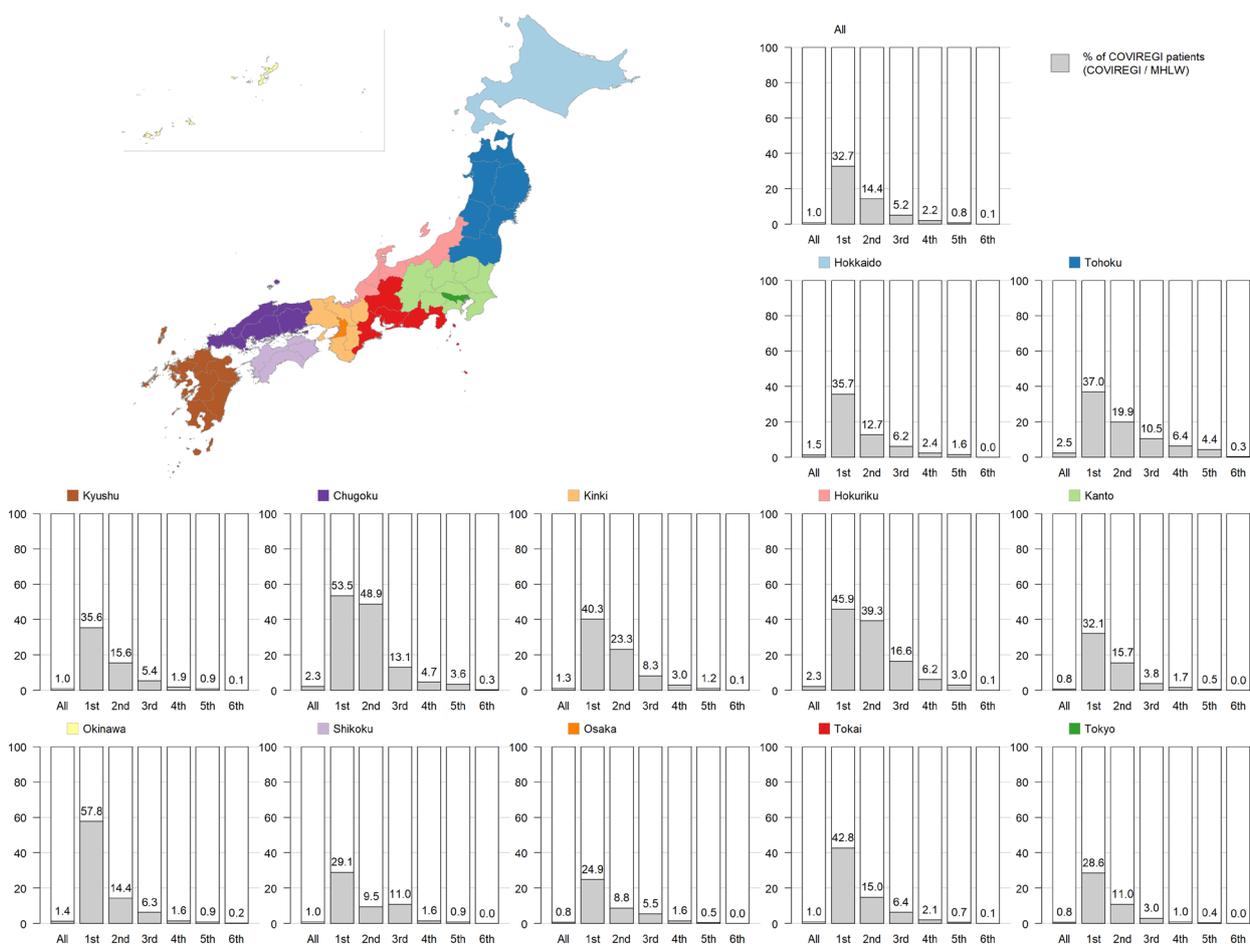


Figure 2. Map of the regional divisions and the region-wise proportion of patients. Proportion of patients = Number of COVIREGI-JP registrants / Number of COVID-19 cases according to national data).

factor that could have affected the number of cases was the reduction in the funding to each participating facility for each patient registered after November 28, 2020 as well as the termination of funding on April 1, 2021. The aforementioned factors were likely to have reduced the number of cases registered after the fourth wave. In addition, medical facilities were encouraged to participate in COVIREGI-JP when they contacted the MHLW to requisition favipiravir, and this may have facilitated case registration. The MHLW provision for the use of favipiravir for COVID-19 ended on December 28, 2021. The current benefits of case registration in COVIREGI-JP by participating facilities include the return of institutional data, the ability to propose and conduct research using multicenter data (if a certain number of patients are registered by the facility), and the contribution to COVID-19 research and to public health through the use as public data. Nonetheless, registration of cases is a manual process and places a heavy burden on facilities.

The importance of and challenges with continuous data registration

Although a registry system for emerging infectious

diseases is essential for the accumulation of real-world data and clinico-epidemiological national evaluation outside of fully controlled conditions, such as in interventional studies, the role of such registries is likely to change over time after the beginning of an epidemic. For example, studies to understand the clinico-epidemiological characteristics and to prioritize the measures of an epidemic should be conducted as soon as possible in the early stages of the epidemic. In contrast, more detailed studies along the lines of research questions are often considered in the later stages as the epidemic continues. As in the case of COVID-19, continuous data registration is important in a situation where various conditions are continually changing, such as vaccine coverage, the prevalence of variants, advances in treatment, and policy changes. Case report forms need to be revised to ascertain these changes while maintaining a balance with data continuity. Some facilities have continued to register cases after the termination of funding for case registration, and such data are very valuable to reducing selection bias during each wave in terms of continuity at the same facility. For the continuity of the registry, financial support commensurate with the registration effort, automated entry, at least partially, to reduce

registration effort, and resources for quality control are important not just in the early stages of the epidemic.

One of the reasons for evaluating the representativeness of COVIREGI-JP data using the approaches in this study is that the cumulative number of hospitalized patients is not publicly available in Japan. Although the number of cases requiring hospitalization is publicly available, this number actually includes patients who were treated at home. Moreover, the number of cases requiring hospitalization includes diverse factors besides disease severity and the patient's underlying condition (*e.g.*, occupancy of beds by patients with COVID-19 and municipal policies) and is not uniform. Therefore, this study used the number of infected patients and the number of deaths, which are less susceptible to such factors, for comparison.

COVIREGI-JP registered cases from all regions, but the regions with a low proportion of patients (Tokyo, Kanto, and Osaka) had the highest number of infected cases. An inversion in the ratio of the case fatality rate (the case fatality rate of the entire infected population was higher than the case fatality rate in COVIREGI-JP) was noted, and especially during Wave 1, when hospitalization was basically recommended for all cases, including mild ones (14). This may have contributed, at least partially, to the fact that registration by participating facilities was biased toward milder cases within each region. The overall case fatality rate ratio of COVIREGI-JP to that of the infected patients nationwide was lowest during this period.

In the analysis of the differences in case fatality rates between the national and regional data in the national data and COVIREGI-JP data sources, there were several regions where the results from COVIREGI-JP and national data diverged. The direction of the discrepancies varied and was not consistent, possibly reflecting the selection bias of COVIREGI-JP in these regions. Despite the decrease in the proportion of patients after Wave 4, divergence did not tend to increase. In addition, although Tokyo, Osaka, and other areas had a large number of infected cases, no divergence was noted. In the regions where the results diverged, that phenomenon was thought to reflect the lack of parallelism in terms of disease severity among the overall infected population and among patients registered in COVIREGI-JP.

A limitation of this study is that the data from Wave 6 were included to provide the most up-to-date data. However, because the epidemic is ongoing and the fact that data entry was incomplete, the analysis in this study may have missed more than a few cases from Wave 6. Therefore, the results of this analysis do not adequately represent the trends during Wave 6. Moreover, the prefecture in the national data and the prefecture in COVIREGI-JP may not match in some cases (*e.g.*, when the patient was admitted to a medical facility far from his or her place of residence). The number of hospitals is based on published data (11),

but it may not indicate the current number. Due to differences in the timing of data download and regional classifications, there are minor discrepancies in the data from a previous study (13) up to the third wave that were included in the current study.

Conclusion

The current study reviewed the representativeness of COVIREGI-JP data; although COVIREGI-JP consistently registered cases from all regions of the country, the proportion of patients tended to decline compared to the beginning of the epidemic. In light of the epidemic's persistence and the ever-changing epidemiology of COVID-19, continued case registration and data utilization in COVIREGI-JP is desirable, although selection bias in COVIREGI-JP registration of cases should be carefully interpreted.

Acknowledgements

The authors wish to thank all the participating facilities for their care of patients with COVID-19 and their cooperation with data entry in the registry.

Funding: This research was supported by the NCGM COVID-19 Gift Fund.

Conflict of Interest: There is no conflict of interest to disclose.

References

1. Kokudo N, Sugiyama H. Hospital capacity during the COVID-19 pandemic. *Glob Health Med.* 2021; 3:56-59.
2. Ministry of Health, Labour, and Welfare of Japan. Information on COVID-19. Ministry of Health, Labour, and Welfare. <https://covid19.mhlw.go.jp/extensions/public/index.html> (accessed April 1, 2022). (in Japanese)
3. National Center for Global Health and Medicine of Japan. COVID-19 Registry Japan. <https://covid-registry.ncgm.go.jp> (accessed April 1, 2022). (in Japanese)
4. Ministry of Health, Labour, and Welfare of Japan. 41st Meeting of the Subcommittee on Immunization and Vaccine of the Health and Welfare Science Council (November/9/2020). <https://covid-registry.ncgm.go.jp/achievements/documents/41.pdf> (accessed April 1, 2022). (in Japanese)
5. National Center for Global Health and Medicine of Japan. COVIREGI-JP Research Results. <https://covid-registry.ncgm.go.jp/achievements/#section03> (accessed April 1, 2022). (in Japanese)
6. Tokyo iCDC. Tokyo iCDC Expert Board Infectious Disease Treatment Team (December/02/2020). <https://covid-registry.ncgm.go.jp/achievements/documents/iCDC.pdf> (accessed April 1, 2022). (in Japanese)
7. Terada M, Ohtsu H, Saito S, Hayakawa K, Tsuzuki S, Asai Y, Matsunaga N, Kutsuna S, Sugiura W, Ohmagari N. Risk factors for severity on admission and the disease progression during hospitalisation in a large cohort of

- patients with COVID-19 in Japan. *BMJ Open*. 2021; 11:e047007.
8. Yamada G, Hayakawa K, Matsunaga N, Terada M, Suzuki S, Asai Y, Ohtsu H, Toyoda A, Kitajima K, Tsuzuki S. Predicting respiratory failure for COVID-19 patients in Japan: A simple clinical score for evaluating the need for hospitalisation. *Epidemiology and Infection*. 2021; 149:e175.
 9. Medical Practice Guide Review Committee. COVID-19 Clinical Practice Guideline ver. 6.2. <https://www.mhlw.go.jp/content/000888608.pdf> (accessed April 1, 2022). (in Japanese)
 10. Matsunaga N, Hayakawa K, Terada M, *et al*. Clinical epidemiology of hospitalized patients with coronavirus disease 2019 (COVID-19) in Japan: Report of the COVID-19 Registry Japan. *Clin Infect Dis*. 2021; 73:e3677-e3689.
 11. e-Stat. Survey of Medical Facilities, 2022 Survey of Medical Facilities (Dynamic), Prefectural ed., Table T1. <https://www.e-stat.go.jp/dbview?sid=0003420870> (accessed April 1, 2022). (in Japanese)
 12. Ministry of Internal Affairs and Communications of Japan. Guidelines on Regional Depictions, 2019. https://www.soumu.go.jp/main_content/000611949.pdf (accessed April 1, 2022). (in Japanese)
 13. Matsunaga N, Hayakawa K, Asai Y, *et al*. Clinical characteristics of the first three waves of hospitalised patients with COVID-19 in Japan prior to the widespread use of vaccination: A nationwide observational study. *Lancet Reg Health West Pac*. 2022; 22:100421.
 14. Ministry of Health, Labour, and Welfare of Japan. Acceptance of Inpatients with COVID-19, February, 10, 2020. <https://www.mhlw.go.jp/content/000600290.pdf> (accessed April 1, 2022). (in Japanese)
-
- Received April 7, 2022; Revised May 20, 2022; Accepted May 25, 2022.
- Released online in J-STAGE as advance publication May 29, 2022.
- *Address correspondence to:*
 Kayoko Hayakawa, Disease Control and Prevention Center, National Center for Global Health and Medicine, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan.
 E-mail: khayakawa@hosp.ncgm.go.jp