Research paper

An assessment of self-reported COVID-19 related symptoms of 227,898 users of a social networking service in Japan: Has the regional risk changed after the declaration of the state of emergency?

Shuhei Nomura [a,b,1], Daisuke Yoneoka [a,b,c,1], Shoi Shi [d,e,1], Yuta Tanoue [f,1], Takayuki Kawashima [g,1], Akifumi Eguchi [h,1], Kentaro Matsuura [i,j], Koji Makiyama [j,k], Keisuke Ejima [l], Toshibumi Taniguchi [m], Haruka Sakamoto [a,b], Hiroyuki Kunishima [n], Stuart Gilmour [c], Hiroshi Nishiura [o], Hiroaki Miyata [a,*]

[a] Department of Health Policy and Management, School of Medicine, Keio University, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
[b] Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[c] Graduate School of Public Health, St. Luke's International University, Tokyo, Japan
[d] Department of Systems Pharmacology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
[e] Laboratory for Synthetic Biology, RIKEN Center for Biosystems Dynamics Research, Osaka, Japan
[f] Institute for Business and Finance, Waseda University, Tokyo, Japan
[g] Department of Mathematical and Computing Science, Tokyo Institute of Technology, Tokyo, Japan
[h] Center for Preventive Medical Sciences, Chiba University, Chiba, Japan
[i] Department of Management Science, Graduate School of Engineering, Tokyo University of Science, Tokyo, Japan
[j] HOXO-M Inc., Tokyo, Japan
[k] Yahoo Japan Corporation, Tokyo, Japan
[l] Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, USA
[m] Department of Infectious Diseases, Chiba University, Chiba, Japan
[n] Department of Infectious Diseases, St. Marianna University School of Medicine, Kanagawa, Japan
[o] Graduate School of Medicine, Hokkaido University, Hokkaido, Japan

ARTICLE INFO

ABSTRACT

Background: In the absence of widespread testing, symptomatic monitoring efforts may allow for understanding the epidemiological situation of the spread of coronavirus disease 2019 (COVID-19) in Japan. We obtained data from a social networking service (SNS) messaging application that monitors self-reported COVID-19 related symptoms in real time in Fukuoka Prefecture, Japan. We aimed at not only understanding the epidemiological situation of COVID-19 in the prefecture, but also highlighting the usefulness of symptomatic monitoring approaches that rely on self-reporting using SNS during a pandemic, and informing the assessment of Japan's emergency declaration over COVID-19.

Methods: We analysed symptoms data (fever over 37.5° and a strong feeling of weariness or shortness of breath), reported voluntarily via SNS chatbot by 227,898 residents of Fukuoka Prefecture during March 27 to May 3, 2020, including April 7, when a state of emergency was declared. We estimated the spatial correlation coefficient between the number of the self-reported cases of COVID-19 related symptoms and the number of PCR confirmed COVID-19 cases in the period (obtained from the prefecture website); and estimated the empirical Bayes age- and sex-standardised incidence ratio (EBSIR) of the symptoms in the period, compared before and after the declaration. The number of symptom cases was weighted by age and sex to reflect the regional population distribution according to the 2015 national census.

Findings: Of the participants, 3.47% reported symptoms. There was a strong spatial correlation of 0.847 (p < 0.001) at municipality level between the weighted number of self-reported symptoms and the number of COVID-19 cases for both symptoms. The EBSIR at post-code level was not likely to change remarkably before and after the declaration of the emergency, but the gap in EBSIR between high-risk and low-risk areas appeared to have increased after the declaration.

---

* Corresponding author.
  E-mail address: hiroaki.miyata@gmail.com (H. Miyata).
[1] Shared co-first authorship.

Interpretation: While caution is necessary as the data was limited to SNS users, the self-reported COVID-19 related symptoms considered in the study had high epidemiological evaluation ability. In addition, though based on visual assessment, after the declaration of the emergency, regional containment of the infection risk might have strengthened to some extent. SNS, which can provide a high level of real-time, voluntary symptom data collection, can be used to assess the epidemiology of a pandemic, as well as to assist in policy assessments such as emergency declarations.

Research in context

Evidence before this study

The extent of the coronavirus disease 2019 (COVID-19) outbreak in Japan is unknown due to incomplete reporting of mild cases and delayed case confirmation. Prime Minister Shinzo Abe declared a state of emergency on the evening of April 7, which asked residents living in the areas covered by the declaration to refrain from going out when it was not necessary or urgent to reduce contact with people. However, the measures taken in accordance with Japan's declaration of a state of emergency are not legally binding and are very different from so-called 'lockdowns' with penalties that have been implemented in other countries. At the same time, Abe stressed that Japan would not be locked down and will maintain as much economic activity as possible. This state of emergency relies heavily on voluntary behaviour change, and the extent to which the declaration has been effective in preventing the spread of infection has not been fully examined. Meanwhile, many countries are developing contact tracing apps and the use of digital big data in evaluating what COVID-19 measures are effective is expected to increase globally.

Added value of this study

This study captures the current epidemiology of COVID-19 using self-reported data from a social networking service (SNS)'s chatbot-based monitoring system (named CO-OPERA) with more than 200,000 voluntary participants between March 27 and May 3, 2020 in Fukuoka Prefecture. We found that the spatial correlation between the weighted number of COVID-19 related symptoms observed in CO-OPERA and the number of PCR-confirmed cases of COVID-19 was highly significant. This suggests that COOPERA has had some success in tracking the COVID-19 pandemic situation in Fukuoka in real time. In addition, although based on the visual assessment, the empirical Bayesian estimates of age- and sex-standardised incidence ratio (EBSIR) was not likely to change significantly before and after the declaration of the emergency, but the gap in EBSIR between high-risk and low-risk areas appeared to have increased after the declaration. This may indicate that self-restraint of movement across local communities has been advanced since the declaration, which resulted in greater regional spatial heterogeneity of COVID-19's transmission intensity in each community. These findings also may imply that regional containment of the infection risk might have strengthened to some extent after the declaration of a state of emergency.

Implications of all the available evidence

COVID-19 is prevalent nationwide in major cities in Japan, and medical resources are in short supply. By using the real-time epidemiological situation monitoring using SNS like COOPERA, it is possible to gather epidemiological information which supports decision-making for managing infection spread. Considering that Japan does not have a strong lockdown policy to counteract the spread of infection, the find-ing of this study could be useful in preparing for, for example, a second wave of COVID-19 as a pandemic measure while maintaining economic activity to some extent..

1. Introduction

The worldwide spread of coronavirus disease 2019 (COVID-19), first reported in Wuhan City, Hubei Province, China, illustrates the difficulties in preventing the spread of infectious diseases in a global society [1,2]. There are many challenges to understanding the spread of COVID-19 in Japan. Incomplete reporting of mild cases and delayed case confirmation could mask the epidemic's growth rate [3]. While several small, population-limited COVID-19 antibody testing studies have been performed [4], due in part to the poor quality of the antibody test, large-scale testing programs at the representative population level have not yet been conducted in the country. In addition, the assessment of 'excess death', which is used as an indicator of the scale of, for example, influenza epidemics [5], has not yet been conducted because timely mortality statistics have not been published [6]. This situation is similar in many other countries.

In Japan, Prime Minister Shinzo Abe declared a state of emergency on the evening of April 7, 2020. Because of the increasing number of COVID-19 cases, seven out of 47 prefectures were designated as areas subject to the declaration of emergency, and expanded to all prefectures on April 16 [7]. Under the state of emergency, prefectural governors can ask residents to refrain from going out when it is not necessary or urgent to reduce opportunities for contact with people. However, the measures taken under the Japanese emergency declaration have limited legal force and are very different from the so-called "lockdown" with penalties already in place in countries such as the United States, the United Kingdom, France, Germany, Italy and India [8]. Japan's emergency declaration is a measure while maintaining economic activities to some extent. However, it relies heavily on voluntary changes in behaviour, and the extent to which the declaration led to changes in the spread of CIVID-19 has not been well examined.

It is highly expected that the use of digital big data will be useful for the epidemiological understanding to guide infectious disease outbreak as well as in informing policy and epidemic management decisions in real-time outbreak response [9]. For example, mobility and location information on smartphone users held by telecommunications and internet companies is beginning to be used in some countries to detect the possibility of close contact with an infected person and to evaluate social-distancing strategies [10–13]. Our research team uses the LINE application (Japan's most popular messaging social networking service [SNS]) in each prefecture to support the response of prefectures and residents to COVID-19 in real time. LINE users can reply to questions about their symptoms (i.e. self-report) related to COVID-19 sent by the

prefecture via a chatbot, and in accordance with the guidelines of the Ministry of Health, Labor and Welfare, users who meet certain criteria will receive feedback information such as contact information of the COVID-19 support center operated by the local government in which they reside. Because of the nature of this service, it also functions as symptomatic monitoring.

Because COVID-19 can cause influenza-like illness [14,15], symptomatic monitoring approaches may help understand potential undocumented COVID-19 when laboratory confirmation is difficult or impossible. In fact, syndromic monitoring efforts have proven useful in the 2003 outbreak of severe acute respiratory syndrome (SARS) [16]. In addition, symptomatic monitoring efforts may allow for understanding the epidemiological situation of the spread of COVID-19, including the extent of local transmission and outbreaks, in the absence of widespread testing [17,18].

In this study, we obtained the service usage data of Fukuoka Prefecture (located on the northern side of the island of Kyushu, with population of 5.11 million), which is one of the first targets of the emergency declaration [7], has been operating the above LINE service since before the declaration, and has a sufficient analysable population. We analysed it for the following three objectives: (1) to estimate the temporal trends in the prevalence of self-reported COVID-19 related symptoms (fever and strong feeling of weariness or shortness of breath) in the prefecture; (2) to compare the regional distribution of cases of self-reported COVID-19 related symptoms with those of the number of conformed COVID-19 cases (obtained from the prefecture website) using spatial correlation coefficients; and (3) to estimate an empirical Bayes standardised incidence ratio of the symptoms, and compare them before and after the declaration.

Our analysis employed a disease mapping approach, which is an epidemiological technique widely used to highlight geographic areas (also known as spatial clusters) with high and low risk of a specific disease and to map how such risks vary over the study region [19,20]. This paper not only contributes to the understanding of the epidemiological situation of COVID-19 in the prefecture and highlights the usefulness of symptomatic monitoring approaches that rely on self-reporting using SNS during a pandemic, but also informs the assessment of the emergency declaration.

2. Methods

2.1. Participants

The LINE service we support is called COOPERA (COvid-19: Operation for Personalized Empowerment to Render smart prevention And care seeking), which uses a chatbot to ask users to report COVID-19 related symptoms if any (see below) as well as personal characteristics such as gender, age, residential location information at post-code level, occupation (single answer from choices), medical condition (multiple answers from choices), and infection prevention action taken (multiple answers from choices). Note that all questions are required to be answered, and the questionnaire cannot be completed and the user will not receive feedback if there are unanswered questions. The details of COOPERA can be found elsewhere [21]. In this study, 15 years of age and older were included in the analysis. Geographical scopes of the municipalities in Fukuoka Prefecture and their populations (the latest national census in 2015[22]) are presented in the appendix Fig. 1.

To use the COOPERA, LINE users need to add the Fukuoka's LINE account for COOPERA as a friend, [23,24], which is run by each prefecture (in other words, a voluntary participant). Users can register from the QR code page on the prefectural website or from the banner link that appears at the top of the LINE app. The banners were displayed on March 28. In this study, we obtained data on 227,898 COOPERA users (our study participants) from March

27, 2020, when the service was launched in Fukuoka Prefecture, to May 3, 2020. In addition, after responding to a national survey related to COVID-19 using LINE by the Ministry of Health, Labour and Welfare, which is conducted irregularly, there is an invitation to register for a COOPERA LINE account for each prefecture. So far, that induction has come after surveys on April 5, 12, and May 1.

2.2. Outcomes of interest

The main outcome of this study was COVID-19 related symptoms [14,15]. In COOPERA, LINE users are asked about two symptoms related to COVID-19, separately for the date of response and the past month: whether they have and/or had a fever of 37.5° or higher (hereafter, condition A); and whether they have and/or had a strong feeling of weariness or shortness of breath (condition B). In addition, they are also asked about the durations of these symptoms – how long the symptoms have/had lasted (for symptoms in the past month, users select the most recent date when symptoms were observed from the pull-down menu and indicate how many days it had lasted (choose up to 31 days). In this study, we also considered the following conditions as a combination of conditions A and B: (condition C) having both conditions A and B; and (condition D): having either condition A or B. For those who used the COOPERA multiple times, only the first usage data was used in the analysis.

2.3. Statistical and geographical analysis

Three major analyses were carried out in this study. First, the prevalence of the condition A–D among the participants was calculated daily from March 1 to April 30 in the prefecture. The day-by-day prevalence here is calculated as follows: Since symptoms were asked back to the past month and since users could choose up to 31 days for the duration of the symptoms, one respondent was counted back two months from the date of their COOPERA use in the denominator of prevalence. For example, a user on May 1 could respond that he/she had symptoms on April 1 and that they had been present for 31 days (from March 2). If a user had any of the symptoms between, for example, April 20 and 23, they would also be counted in the numerator of prevalence for these date. Because of this definition, the more recent the date, the smaller the denominator, and the more likely to bias the user (in Discussion for more details). Therefore, the period for presenting results was not until the end of the study period (May 3) but until the end of April.

Next, we compared the regional distribution of the number of participants with conditions A–D during March 27 to May 3 with that of the number of COVID-19 cases confirmed by PCR testing in the period. We used a modified test of spatial correlation coefficients developed by Clifford et al. (1989) to identify spatial correlations between them [25]. The test is based on the modified variance and degrees of freedom of the standard t-test. In this (and following) analyses, the number of participants with conditions was weighted by age and sex to reflect the regional population distribution according to the 2015 national census [22]. The number of PCR-confirmed cases in Fukuoka was extracted from the prefecture's website [26]. Since the geographical information on PCR-confirmed cases was only available at the municipal level (n=72) rather than the post-code (n = 3,199), it was tested at municipal level.

In order to minimize the influences of unstable incidence of COVID-19 related symptoms from areas with small number of the participants, we employed the spatial smoothing technique based on the empirical Bayes methods to 'borrow strength' from neighboring areas and minimize problems associated with small populations at risk [27]. Empirical bayes methods have been widely
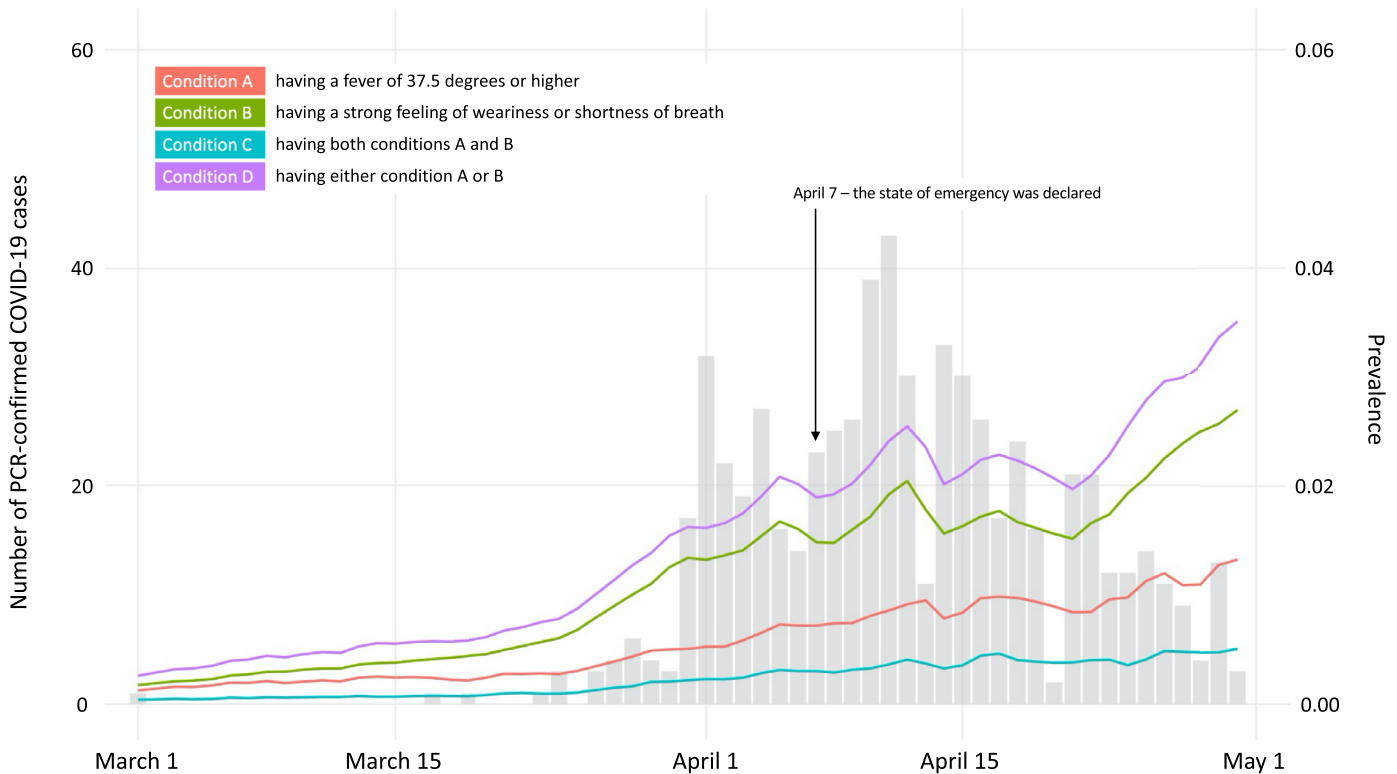
Fig. 1. Daily trend in prevalence of conditions A–D in Fukuoka Prefecture among study participants from March 1 to April 30, 2020. Red, green, orange, and purple lines indicate the prevalence of participants with condition A–D, respectively. Gray bars indicate the number of confirmed PCR cases in Fukuoka [25]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

used in variety of fields including small area estimation and disease mapping [28,29]. Then, an empirical Bayes age- and sex-standardised incidence ratio (EBSIR) of the conditions A–D during March 27 to May 3 was estimated at the post-code level [29,30]. Unless noted otherwise, we present the 95% uncertainty intervals along with them. The spatial neighbourhood and its associated local adjacency matrix were defined based on the k-nearest neighbourhood method with k = 60 [31]. As a sensitivity check of the EBSIR due to the change in the definition of neighbourhood, we re-estimated it in the case of the queen contiguity, where neighbours share at least a common boundary, and in the case of local adjacency matrices based on the k-nearest neighbourhood method with k = 20, 80, and 100. In addition, EBSIR of the conditions A–D were calculated before (March 28–April 6) and after the emergency declaration (April 7–May 3) at the post-code level. In view of the reduction in the number of samples analysed by dividing the data before and after the declaration, the local adjacency matrix were defined based on the k-nearest neighbourhood method with k = 100. In the sense of robustness check, we also estimated EBSIR at municipal level for all conditions with k = 10 and, as a sensitivity check, with queen contiguity or k = 5, 15, or 20.

Statistical analysis was performed using R (version 3.6.0) software. The type I error rate was fixed at 0.05 (2-tailed).

3. Results

The demographic characteristics of 227,898 participants are shown in Table 1 by condition type. The range in the number of participants per postal code was 0 to 1468 with median of 23. Of these respondents, 96.52% reported no COVID-19 related symptoms at the date of COOPERA use, while 1.40%, 2.73%, 0.66%, and 3.47% reported conditions A–D, respectively. The mean and standard deviation (SD) of age were 43.07 (12.46), 36.44 (13.13), 36.67

(11.67), 35.53 (12.96), 36.8 (12.03), for participants who reported no COVID-19 related conditions, and condition A–D, respectively. There were approximately twice as many women as men. About 40% of respondents were office workers, followed by part-time workers and unemployed workers at about 13–16%, depending on conditions. Except for dialysis, participants without COVID-19 related symptoms had the lowest proportion of having a chronic disease. The implementation rate of infection preventions was also highest among those without symptoms for all actions, except for gargling with Isodine and taking time off from school or work when having a fever or other symptoms.

Fig. 1 shows the prevalence of each condition among the participants on each day from March 1 to April 30, 2020, in a time series. Gray bars indicate the number of confirmed PCR cases in Fukuoka. Changes in the number of PCR-confirmed COVID-19 cases and the prevalence of each condition were roughly linked except for the last week of the study period. The daily number of participants in the denominator in the prevalence calculation are shown in appendix Fig. 2.

Fig. 2 shows the regional distribution of the weighted number of cases of each condition in Fukuoka during March 27 to May 3, 2020, at post-code level. It was found that the weighted number of cases with conditions was higher in prefectural capitals and core cities with large populations, such as Fukuoka City. Appendix figure 3 shows the regional distribution of PCR-confirmed COVID-19 cases by municipality in the same period. There was a statistically significant spatial correlation with the weighted number of cases with each condition at municipality level, with correlation coefficients of 0.847 (p < 0.001), 0.847 (p < 0.001), 0.846 (p < 0.001), and 0.850 (p < 0.001) for conditions A–D, respectively.

Fig. 3 shows the EBSIR for each condition at post-code level with k = 60 in Fukuoka. The EBSIR can be interpreted as a measure of relative risk, with a value of 1.0 representing similar risk to that

Table 1
Basic characteristics of participants from Fukuoka Prefecture at the date of COOPERA use

	No COVID-19 related symptom (n = 219,983)	Condition A (n = 3,195)	Condition B (n = 6,227)	Condition C (n = 1,507)	Condition D (n = 7,915)
Age, years					
Mean (standard deviation)	43.07 (12.46)	36.44 (13.13)	36.67 (11.67)	35.53 (12.96)	36.8 (12.03)
Range (min–max*)	15–101	15–101	15–101	15–101	15–101
15–19	4521 (2.06)	161 (5.04)	209 (3.36)	73 (4.84)	297 (3.75)
20–29	25,663 (11.67)	893 (27.95)	1644 (26.40)	475 (31.52)	2062 (26.05)
30–39	57,323 (26.06)	1017 (31.83)	2031 (32.62)	474 (31.45)	2574 (32.52)
40–49	66,874 (30.40)	672 (21.03)	1488 (23.90)	295 (19.58)	1865 (23.56)
50–59	42,950 (19.52)	244 (7.64)	619 (9.94)	103 (6.83)	760 (9.60)
60–69	18,040 (8.20)	111 (3.47)	164 (2.63)	43 (2.85)	232 (2.93)
70–79	4205 (1.91)	77 (2.41)	55 (0.88)	34 (2.26)	98 (1.24)
80–	407 (0.19)	20 (0.63)	17 (0.27)	10 (0.66)	27 (0.34)
Sex					
Female	147,436 (67.02)	1998 (62.54)	4028 (64.69)	882 (58.53)	5144 (64.99)
Male	72,227 (32.83)	1187 (37.15)	2185 (35.09)	619 (41.07)	2753 (34.78)
Other	320 (0.15)	10 (0.31)	14 (0.22)	6 (0.40)	18 (0.23)
Pregnant	3214 (1.46)	34 (1.06)	104 (1.67)	21 (1.39)	117 (1.48)
Occupation					
Self-employed	20,623 (9.37)	295 (9.23)	621 (9.97)	161 (10.68)	755 (9.54)
Company employees	91,080 (41.40)	1278 (40.00)	2612 (41.95)	612 (40.61)	3278 (41.42)
Government worker	11,680 (5.31)	126 (3.94)	232 (3.73)	53 (3.52)	305 (3.85)
Student	7841 (3.56)	263 (8.23)	352 (5.65)	123 (8.16)	492 (6.22)
Part-time job	35,174 (15.99)	408 (12.77)	840 (13.49)	153 (10.15)	1095 (13.83)
Unemployed	29,938 (13.61)	445 (13.93)	839 (13.47)	213 (14.13)	1071 (13.53)
Others	23,647 (10.75)	380 (11.89)	731 (11.74)	192 (12.74)	919 (11.61)
Taking antifebrile medications					
Current	3768 (1.71)	1272 (39.81)	1688 (27.11)	693 (45.99)	2267 (28.64)
Past one month	10,874 (4.94)	490 (15.34)	1188 (19.08)	286 (18.98)	1392 (17.59)
Medical conditions					
Malignant tumor with anticancer drugs	1097 (0.50)	25 (0.78)	38 (0.61)	13 (0.86)	50 (0.63)
Malignant tumor without anticancer drugs	1998 (0.91)	38 (1.19)	61 (0.98)	26 (1.73)	73 (0.92)
Cardiovascular diseases	3922 (1.78)	61 (1.91)	159 (2.55)	37 (2.46)	183 (2.31)
Kidney diseases	1553 (0.71)	29 (0.91)	60 (0.96)	18 (1.19)	71 (0.90)
Diabetes mellitus	7111 (3.23)	112 (3.51)	236 (3.79)	63 (4.18)	285 (3.60)
In dialysis treatment	250 (0.11)	7 (0.22)	7 (0.11)	6 (0.40)	8 (0.10)
Chronic obstructive pulmonary disease	685 (0.31)	28 (0.88)	76 (1.22)	19 (1.26)	85 (1.07)
Treatment with immunosuppressant	2302 (1.05)	46 (1.44)	102 (1.64)	24 (1.59)	124 (1.57)
Preventive action					
Washing hands in running water (multiple times a day)	131,263 (59.67)	1811 (56.68)	3642 (58.49)	829 (55.01)	4624 (58.42)
Washing hands with soap and water (multiple times a day)	196,166 (89.17)	2634 (82.44)	5115 (82.14)	1191 (79.03)	6558 (82.86)
Washing hands with alcohol (multiple times a day)	155,589 (70.73)	1969 (61.63)	3946 (63.37)	912 (60.52)	5003 (63.21)
Etiquette (masks, handkerchiefs, etc.) in case of coughing or sneezing	194,554 (88.44)	2695 (84.35)	5335 (85.68)	1252 (83.08)	6778 (85.63)
Take time off from school or work when you have a fever or other symptoms	88,911 (40.42)	1589 (49.73)	2394 (38.45)	696 (46.18)	3287 (41.53)
Gargling with water	108,326 (49.24)	1359 (42.54)	2745 (44.08)	614 (40.74)	3490 (44.09)
Gargling with Isozine	29,617 (13.46)	403 (12.61)	873 (14.02)	201 (13.34)	1075 (13.58)
Regular ventilation	125,942 (57.25)	1535 (48.04)	3013 (48.39)	678 (44.99)	3870 (48.89)
Maintaining humidity	54,043 (24.57)	589 (18.44)	1147 (18.42)	261 (17.32)	1475 (18.64)
A well-balanced diet	98,618 (44.83)	943 (29.51)	1756 (28.20)	403 (26.74)	2296 (29.01)
Regular exercise	54,243 (24.66)	459 (14.37)	847 (13.60)	229 (15.20)	1077 (13.61)
Plenty of rest	105,878 (48.13)	1132 (35.43)	2163 (34.74)	513 (34.04)	2782 (35.15)
Telework	16,168 (7.35)	219 (6.85)	428 (6.87)	96 (6.37)	551 (6.96)
Staggered commuting hour	13,949 (6.34)	155 (4.85)	328 (5.27)	69 (4.58)	414 (5.23)
Avoidance of crowds other than staggered commuting hour	50,477 (22.95)	502 (15.71)	1033 (16.59)	207 (13.74)	1328 (16.78)
Obtain up-to-date coronavirus information	131,923 (59.97)	1483 (46.42)	3059 (49.12)	656 (43.53)	3886 (49.10)
Other preventive action	3366 (1.53)	40 (1.25)	107 (1.72)	18 (1.19)	129 (1.63)
No preventive action	867 (0.39)	40 (1.25)	74 (1.19)	29 (1.92)	85 (1.07)

Condition A: having a fever of 37.5 degrees or higher; Condition B: having a strong feeling of weariness or shortness of breath; Condition C: having both conditions A and B; and Condition D: having either condition A or B.

* In COOPERA, due to the specification of the format, all people aged 101 years or older are registered as being 101 years old.

in the population of. Our estimates suggest that there was regional variation in the risk of each condition within the whole prefecture. The 95% uncertainty lower and upper bounds of the EBSIR are presented in the appendix figures 4 and 5, respectively. The EBSIR and its 95% uncertainty lower and upper bounds for each condition at municipal level with k = 10 are presented in the appendix Figs. 6–

8, respectively. These also demonstrated the regional variation in the risk of each condition in Fukuoka Prefecture.

As a sensitivity analysis, the EBSIR for conditions A and C using the queen contiguity or k = 20, 80, or 100 at post-code level are presented in appendix Fig. 9. In addition, those using the queen contiguity or k = 5, 15, or 20 at municipal level are presented in
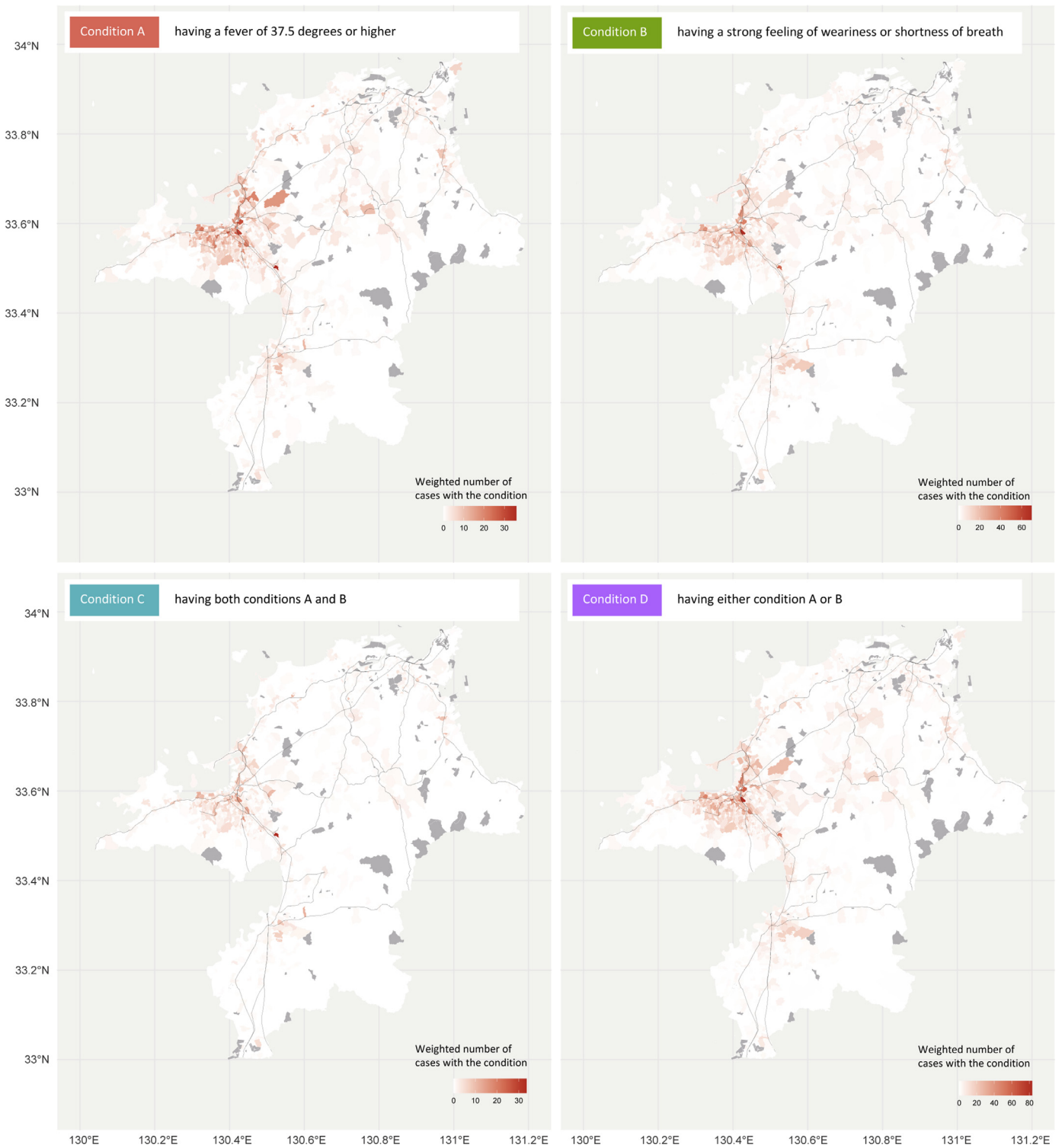
Fig. 2. Maps plotting weighted number of cases with conditions A–D at post-code level in Fukuoka Prefecture during March 27 to May 3, 2020. The black line represents a railroad track. The gray areas represent that there were no participants. The number of participants with conditions was weighted by age and sex to reflect the regional population distribution according to the 2015 national census [22]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

appendix Fig. 10. Some relatively high or low EBSIR areas were unevenly observed in the queen contiguity model, but similar trends are seen in the k-nearest neighbourhood models.

Fig. 4 (and the appendix Figs. 11 and 12) shows the EBSIR (and its 95% uncertainty lower and upper bounds, respectively) at postcode level with k = 100 for condition A–D before and after the emergency declaration. Compared to the situation before the dec-

laration of an emergency, the trend in the risk plot after the declaration has changed little except in some areas. However, the gap in EBSIR between high-risk areas (red) and low-risk areas (blue) has become stronger since the declaration (have increased colour contrast). Those at municipal level with k = 10 are also presented in the Appendix 13, 14, and 15, respectively.
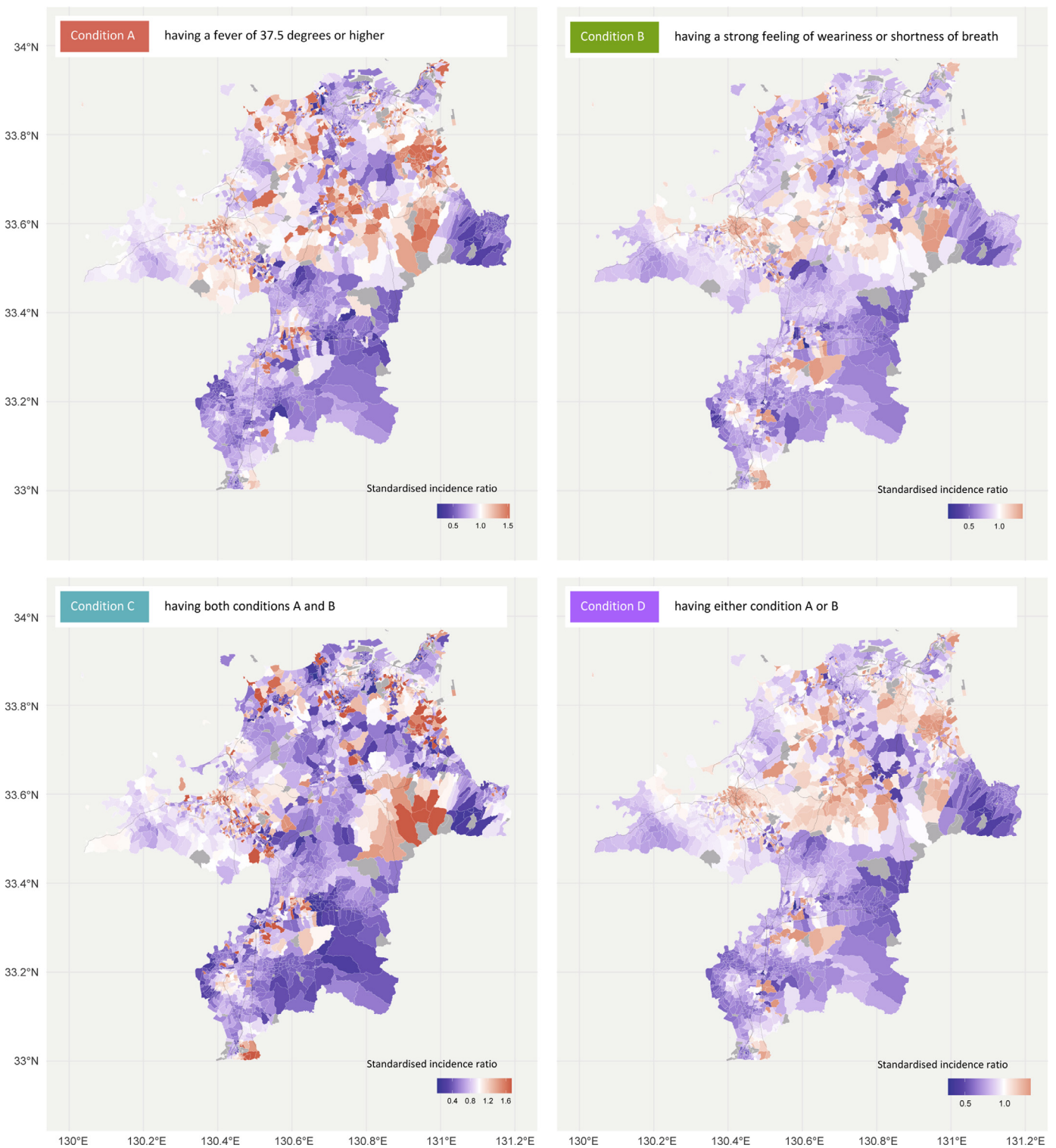
Fig. 3. Maps plotting the empirical Bayesian estimates of age- and sex-standardised incidence ratio of conditions A–D at post-code level in Fukuoka Prefecture during March 27 to May 3, 2020. The black line represents a railroad track. The spatial neighbourhood and its associated local adjacency matrix were defined based on the k-nearest neighbourhood method with $k = 60$. The gray areas represent that there were no participants or an estimation was not possible. The number of participants with conditions was weighted by age and sex to reflect the regional population distribution according to the 2015 national census [22].

4. Discussion

Information on the potential for undocumented infections in COVID-19 is important for implementing well-targeted preventive interventions, as noted in the COVID-19 outbreak in Italy [32]. Attempts to understand the epidemiological situation of infection risks using a syndromic monitoring approach may identify communities that are more vulnerable to COVID-19 and allow appropriate policy-making and resource allocation decisions, particularly when a large-scale laboratory confirmation is difficult or impossible [18]. Here, we analysed data on self-reported COVID-19 related symptoms of more than 220,000 LINE users using the LINE personal
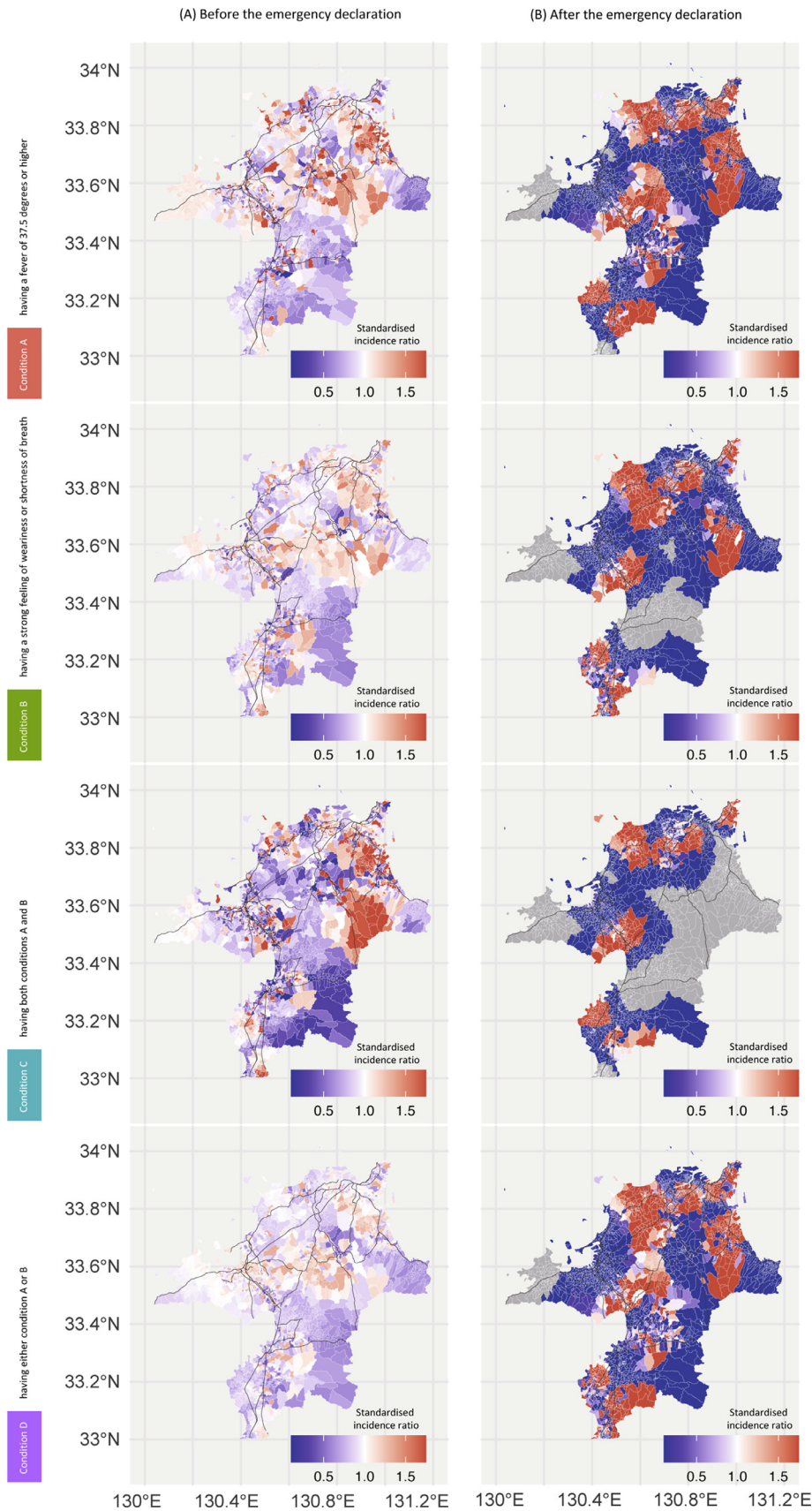
Fig. 4. Maps plotting the empirical Bayesian estimates of age- and sex-standardised incidence ratio of conditions A–D at post-code level in Fukuoka Prefecture before (A) and after (B) the declaration of state of emergency. The black line represents a railroad track. The spatial neighbourhood and its associated local adjacency matrix were defined based on the k-nearest neighbourhood method with *k* = 100. The gray areas represent that there were no participants or an estimation was not possible. The number of participants with conditions was weighted by age and sex to reflect the regional population distribution according to the 2015 national census [22]. Before: during March 27 to April 6, 2020; and After: during April 7 to May 3, 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

support service 'COOPERA', which Fukuoka Prefecture launched as part of its COVID-19 response.

One of the great advantages of using SNS such as LINE to understand the epidemiological status of infectious diseases is its immediacy [33]. With COOPERA, symptoms can be reported immediately via LINE, making data collection extremely quick. Compared to prescription-based pharmacy surveillance, which is more common in understanding the timing and scale of influenza epidemics [34], or methods for identifying epidemic areas based on school closures [35], it is possible to identify signs and high-risk locations before an outbreak or in the early stages of an epidemic. In addition, SNS makes it easy to track changes in an individual's symptoms over time, such as when and for how long symptoms appeared. When the central government, local governments, and/or academic societies launch guidelines to determine whether an infection is suspected or not, or to seek medical consultations, it would be possible to identify individuals who fall under these guidelines in a timely manner, leading to the necessary support. With COOPERA, in fact, LINE users who are found to have a fever of 37.5° or higher for four days or more (the Ministry of Health, Labour and Welfare's guideline for consultation with health centers [revised on May 8]), will receive contact information and a recommendation for health center consultation via a chatbot.

In the present study, the temporal trends in the number of PCR-confirmed cases of COVID-19 in Fukuoka Prefecture were roughly consistent with the trends in the prevalence of COVID-19 related symptoms reported on COOPERA, suggesting that COOPERA has some ability to understand the epidemiological status of COVID-19. The last approximately one week shows no similarity in the trend, but this is likely due to the participation bias involved in how the prevalence was calculated.

The spatial correlation between the weighted number of COVID-19 related symptoms observed in COOPERA and the number of PCR-confirmed cases of COVID-19 was also demonstrated to be highly significant (with coefficient of >0.8 for all conditions). In addition, we showed that there were several geographic hotspots where EBSIR is relatively high. Especially in urban areas, for example, Fukuoka City and Kitakyushu City, which are the core cities of the prefecture, were likely to form relatively large geographic hotspots locally. The findings may also imply that hotspots have formed along railroads, especially around the stations (where the black lines on the maps intersect). COOPERA can thus help identify areas at risk of infection in sub-regional small areas and may also be useful for further consideration of how to allocate appropriate medical and human resources for COVID-19 control.

Furthermore, although it is important to note that this is based on visual assessments rather than statistical hypothesis tests, the regional distribution of EBSIR did not change remarkably before and after the declaration of the emergency, but the gap in EBSIR between high-risk and low-risk areas obviously increased after the declaration. This may mean that self-restraint of movement across local communities has been advanced since the declaration, which resulted in greater regional spatial heterogeneity of COVID-19′s transmission intensity in each community. In other words, the mobility of people across communities may have decreased, and the correlation of the transmission intensity between communities may have become weaker [36,37]. These findings also may imply that regional containment of the infection risk might have strengthened to some extent after the declaration of a state of emergency, even though it cannot prohibit the mobility/movement of residents with penalties. In fact, according to a report released by Google LLC evaluating human mobility using anonymized location data collected from smartphones (Google COVID-19 Community Mobility Reports), in Fukuoka Prefecture, there was a substantial difference in a decrease in people's mobility at transit stations, such as bus and train stations before and after April 7, compared to those in January (before the COVID-19 pandemic): a two-week average of −28.7% and −51.4%, respectively.[38]

This study has several limitations. The first is the bias of the participants. COOPERA uses LINE, which has the largest number of users in Japan, but may be affected by a selection bias that does not include data on people who do not have access to the Internet or smartphones. Unfortunately, the only available information on total LINE users is that there are 83 million active users, and there is no detailed distribution data on user backgrounds such as gender, age, prefecture of residence, etc. Therefore, it is difficult to verify the selection bias because it is not possible to evaluate the COOPERA utilization rate of LINE users as well as the demographics of COOPERA users and non-users.

Since the use of COOPERA is also based on the autonomy of LINE users, there is a possibility that LINE users who have relatively strong health awareness and anxiety are more likely to use COOPERA. Therefore, attention should be paid to the possibility that the prevalence of conditions A–D in this study was overestimated. It should also be noted that COOPERA data include recall bias. COOPERA includes questions about COVID-19 related symptoms from the current to recent past month, but people may not remember symptoms as accurately if they occurred further back in time. This recall bias may have produced a trend toward higher prevalence for more recent days (appendix Fig. 2) when only new information is considered, as in the recent trend of higher prevalence in Fig. 1.

It should be noted that the COOPERA questionnaire can be responded to multiple times, but this study considered only the first response data. The timing of the second response depends on the content of the first response. Users who report any COVID-19 related symptoms, such as fever, receive a questionnaire every day until they report no symptoms. Users who report no symptoms receive a questionnaire every four days. Whether to respond or not is voluntary. Also, importantly, the people who respond the first time receive feedback and advice about the COVID-19. Therefore, those who respond on subsequent occasions probably have had their behaviour changed by the COOPERA, and thus may not reflect the naïve behaviour of the population they are from, and the degree to which they are not representative of the population will grow with each re-response. For these reasons (e.g. different response opportunities, different response motivations, possible behaviour changes, etc.), we believe that there is a very large response and other biases in the data for the second and subsequent responses. Verification of the nature of the second and subsequent responses is an extremely important area of research for self-administered questionnaires such as COOPERA that can track the symptoms of respondents, and is one of our next research objectives.

As of May 1, 2020, male and female people in Fukuoka Prefecture aged 15 and over (the target age in this study) accounted for 46.5% and 53.5%, respectively [39]. Meanwhile, 67.0% of the participants in this study were women and 32.9% were men, indicating a tendency for women to participate more. This gender difference in study participation has been noted in many previous studies [40]. In addition, the majority of participants in the study were 40–49 years old (30.16%), followed by 30–39 year olds (26.28%). On the other hand, these age groups accounted for 16.4% and 13.6% of the adult population of Fukuoka Prefecture. Also, while people aged 70 or older accounted for 2.08% of the participants in this study, this age group accounted for 24.62% in Fukuoka Prefecture. In other words, the study participants were relatively younger than the prefecture' population. However, it should be noted that since the number of those with conditions was weighted for age and sex based on the latest national population census in 2015 [22], the effects of these participant biases might not be significant.

Third, it should be noted that COOPERA can only evaluate COVID-19 related symptoms, such as fever, and not a definitive

diagnosis by PCR testing. That is, the epidemiological ability of COOPERA to assess COVID-19 situation may be affected by the prevalence of colds and influenza with symptoms similar to those of COVID-19. However, this year's seasonal influenza in Japan has been reported to be less active than in previous years [41].

Fourth, unfortunately, strict verification of the reliability of our self-reported data was not possible because there was no 'correct' data available (such as those measured objectively). Self-report data collection methods are generally applied to large medical or biomedical studies for efficient reasons. However, previous studies have shown that health risk factors, including biochemical measurements such as weight [42,43], smoking status [44,45], and nutritional intake [46,47], tend to be overreported. On the other hand, some studies suggested that the reliability of the self-reported data is relatively high with regard to information on health conditions, such as medical history, although it depends to some extent on the severity of the condition, study design, condition of interest, and target population [48,49]. The primary outcome used in this study was COVID-19 related symptoms, which are the center of worldwide interest today, and therefore we believe that reliability of our self-reported data can be secured to some extent. In addition, COOPERA is a service that returns individual feedback based on information entered by the users. Unlike traditional questionnaire surveys, COOPERA provides an incentive to respond properly, so the data may be reliable.

Lastly, in the analyses, there was no adjustment for other LINE users' information collected by the COOPERA other than age and sex, such as medical conditions, occupation, and preventive action. Since this is the first study to analyse the COOPERA data from Fukuoka Prefecture with the uncertainty surrounding self-reported data during a pandemic, we chose simple methods that are more easily interpreted and communicated [50–52].

5. Conclusion

The COVID-19 response in Japan using SNS based on voluntary reports provides many lessons for future infectious disease pandemic countermeasures. SNS, which provides a high level of real-time data collection, whether it relies on voluntary reporting, can be used to assess the epidemiology of a pandemic in terms of time and geography, as well as to assist in policy assessments such as emergency declarations, if the data is large. The COVID-19 related symptoms evaluated by COOPERA using LINE have high epidemiological evaluation ability both temporally and geographically. In addition, after the declaration of the emergency, regional containment of the infection risk might have strengthened to some extent.

Contributors

All authors took responsibility for the integrity of the data and the accuracy of the data analysis. All the authors made critical revisions to the manuscript for important intellectual content and gave final approval of the manuscript. The opinions, results, and conclusions reported in this paper are those of the authors and are independent from the funding bodies.

Ethics statement

Ethical approval was granted by the ethics committee of Keio University School of Medicine, under authorization number 20190338. We collected data from only those who gave consent for the prefecture that administers the questionnaire to provide their information to a third party for research use. Respondents are required to give consent before they proceed to the questionnaire response page on the LINE chatbot.

Data statement

Data available on request due to privacy/ethical restrictions.

Declaration of Competing Interest

Acknowledgement

Funding

Role of the funding source

The funder of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the paper. The authors had full access to all the data in the study and had final responsibility to submit for publication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.lanwpc.2020.100011.

References

[1] Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020;382(13):1199–207.

[2] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese center for disease control and prevention. JAMA 2020.

[3] Omori R, Mizumoto K, Chowell G. Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate. Int J Infect Dis 2020;94:116–18.

[4] Doi A, Iwata K, Kuroda H, et al. Estimation of seroprevalence of novel coronavirus disease (COVID-19) using preserved serum at an outpatient setting in Kobe, Japan: a cross-sectional study. medRxiv2020: 2020.04.26.20079822.

[5] Vestergaard LS, Nielsen J, Krause TG, et al. Excess all-cause and influenza-attributable mortality in Europe, December 2016 to February 2017. Euro Surveill 2017;22(14).

[6] Leon DA, Shkolnikov VM, Smeeth L, Magnus P, Pechholdova M, Jarvis CI. COVID-19: a need for real-time monitoring of weekly excess deaths. Lancet 2020;395(10234):e81.

[7] Cabinet Secretariat. [COVID-19] Press Conference by the Prime Minister Regarding the Declaration of a State of Emergency. 2020. https://japan.kantei.go.jp/98_abe/statement/202004/_00001.html (Accessed 5 May 2020).

[8] The Asahi Shimbun. Japan's state of emergency is no lockdown. What's in it? 2020. http://www.asahi.com/ajw/articles/13281427 (Accessed 5 May 2020).

[9] George DB, Taylor W, Shaman J, et al. Technology to advance infectious disease forecasting for outbreak management. Nat Commun 2019;10(1):3932.

[10] Google. COVID-19 community mobility reports. 2020. https://www.google.com/covid19/mobility/ (Accessed 5 May 2020).

[11] Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. JAMA 2020.

[12] Kim MJ, Denyer S. A 'travel log' of the times in South Korea: Mapping the movements of coronavirus carriers. 2020. https://www.washingtonpost.com/world/asia_pacific/coronavirus-south-korea-tracking-apps/2020/03/13/2bed568e-5fac-11ea-ac50-18701e14e06d_story.html (Accessed 5 May 2020).

[13] Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020;26(4):463–4.

[14] Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020.

[15] Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020;382(18):1708–20.

[16] Wilder-Smith A, Chiew CJ, Lee VJ. Can we contain the COVID-19 outbreak with the same measures as for SARS? Lancet Infect Dis 2020;20(5):e102–e1e7.

[17] Gong M, Liu L, Sun X, Yang Y, Wang S, Zhu H. Cloud-based system for effective surveillance and control of COVID-19: useful experiences from Hubei, China. J Med Internet Res 2020;22(4):e18948.

[18] Wiemken TL, Shacham E. Identifying potential undocumented COVID-19 using publicly reported influenza-like-illness and laboratory-confirmed influenza disease in the United States: an approach to syndromic surveillance? Am J Infect Control 2020.

[19] Desjardins MR, Whiteman A, Casas I, Delmelle E. Space-time clusters and co-occurrence of chikungunya and dengue fever in Colombia from 2015 to 2016. Acta Trop 2018;185:77–85.

[20] Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. Appl Geogr 2020;118:102202.

[21] Yoneoka D, Kawashima T, Tanoue Y, et al. Early SNS-based monitoring system for the COVID-19 outbreak in Japan: a population-level observational study. J Epidemiol 2020. https://doi.org/10.2188/jea.JE20200150.

[22] Portal site of official statistics of Japan. 2015 Population Census. https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200521&tstat=000001049104&cycle=0&tclass1=000001049105&stat_infid=000031594311 (Accessed 10 June 2020).

[23] LINE Corporation. Fukuoka – new corona countermeasure personal support. 2020. https://page.line.me/014bvrfx (Accessed 12 May 2020).

[24] Fukuoka Prefecture. LINE official account: Fukuoka – new corona countermeasure personal support. https://www.pref.fukuoka.lg.jp/contents/covid19-linepersonalsupport.html (Accessed 12 May 2020).

[25] Clifford P, Richardson S, Hemon D. Assessing the significance of the correlation between two spatial processes. Biometrics 1989;45(1):123–34.

[26] Prefecture F. Occurrence in Fukuoka Prefecture. 2020. https://www.pref.fukuoka.lg.jp/contents/covid19-hassei.html (Accessed 6 May 2020).

[27] Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. J R Stat Soc Ser C 1991;40(2):283–94.

[28] Tsutakawa RK, Shoop GL, Marienfeld CJ. Empirical Bayes estimation of cancer mortality rates. Stat Med 1985;4(2):201–12.

[29] Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987;43(3):671–81.

[30] Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. J R Stat Soc Ser C Appl Stat 1991;40(2):283–94.

[31] Lawson AB. Bayesian disease mapping: hierarchical modeling in spatial epidemiology. CRC Press; 2008.

[32] Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. JAMA 2020.

[33] Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020.

[34] Japan medical association, Japan pharmaceutical association, graduate school of pharmacy, Nihon University, EM Systems Co Ltd. Pharmacy surveillance daily report. http://prescription.orca.med.or.jp/syndromic/kanjyasuikei/index.php (Accessed 12 May 2020).

[35] National Institute of Infectious Diseases. Reports of influenza-like illness (number of absentees). https://www.niid.go.jp/niid/ja/flu-flulike.html (Accessed 12 May 2020).

[36] Keeling MJ, Bjørnstad ON, Grenfell BT. 17 - Metapopulation dynamics of infectious diseases. In: Hanski I, Gaggiotti OE, editors. Ecology, genetics and evolution of metapopulations. Burlington: Academic Press; 2004. p. 415–45.

[37] Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci 2009;106(51):21484.

[38] Google LLC. COVID-19 Community Mobility Reports [Internet]. 2020. https://www.google.com/covid19/mobility/ (Accessed 9 July 2020).

[39] Fukuoka Prefecture. Fukuoka Prefecture – population migration survey by municipality and age as of May 1, 2020. https://ckan.open-governmentdata.org/dataset/401000_jinkouidouchousa-shikuchouson-nenrei/resource/b2e659da-d140-4530-8db5-dfd63c15c638 (Accessed 10 June 2020).

[40] Ekholm O, Gundgaard J, Rasmussen NK, Hansen EH. The effect of health, socio-economic position, and mode of data collection on non-response in health interview surveys. Scand J Public Health 2010;38(7):699–706.

[41] Sakamoto H, Ishikane M, Ueda P. Seasonal influenza activity during the SARS–CoV-2 outbreak in Japan. JAMA 2020.

[42] Bonn SE, Trolle Lagerros Y, Balter K. How valid are Web-based self-reports of weight? J Med Internet Res 2013;15(4):e52.

[43] Lassale C, Peneau S, Touvier M, et al. Validity of web-based self-reported weight and height: results of the Nutrinet-Sante study. J Med Internet Res 2013;15(8):e152.

[44] Brigham J, Lessov-Schlaggar CN, Javitz HS, Krasnow RE, McElroy M, Swan GE. Test-retest reliability of web-based retrospective self-report of tobacco exposure and risk. J Med Internet Res 2009;11(3):e35.

[45] Graham AL, Papandonatos GD, Bock BC, et al. Internet- vs. telephone-administered questionnaires in a randomized trial of smoking cessation. Nicotine Tob Res 2006(1):S49–57.

[46] Labonte ME, Cyr A, Baril-Gravel L, Royer MM, Lamarche B. Validity and reproducibility of a web-based, self-administered food frequency questionnaire. Eur J Clin Nutr 2012;66(2):166–73.

[47] Christensen SE, Moller E, Bonn SE, et al. Two new meal- and web-based interactive food frequency questionnaires: validation of energy and macronutrient intake. J Med Internet Res 2013;15(6):e109.

[48] Landkroon AP, de Weerd S, van Vliet-Lachotzki E, Steegers EA. Validation of an internet questionnaire for risk assessment in preconception care. Public Health Genomics 2010;13(2):89–94.

[49] van Gelder MM, Schouten NP, Merkus PJ, Verhaak CM, Roeleveld N, Roukema J. Using web-based questionnaires and obstetric records to assess general health characteristics among pregnant women: a validation study. J Med Internet Res 2015;17(6):e149.

[50] May RM. Uses and abuses of mathematics in biology. Science 2004;303(5659):790–3.

[51] Viboud C, Sun K, Gaffey R, et al. The RAPIDD ebola forecasting challenge: synthesis and lessons learnt. Epidemics 2018;22:13–21.

[52] Kahn R, Mahmud AS, Schroeder A, et al. Rapid forecasting of cholera risk in Mozambique: translational challenges and opportunities. Prehosp Disaster Med 2019;34(5):557–62.