

Measuring Public Concern About COVID-19 Through Search Queries

Zhiwei Gao, Sumio Fujita, Nobuyuki Shimizu, Kongmeng Liew, Taichi Murayama, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki

Submitted to: JMIR Public Health and Surveillance
on: April 23, 2021

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 4

Supplementary Files..... 18

 Figures 19

 Figure 1..... 20

 Figure 2..... 21

 Figure 3..... 22

 Figure 4..... 23

Measuring Public Concern About COVID-19 Through Search Queries

Zhiwei Gao¹ BEng; Sumio Fujita² PhD; Nobuyuki Shimizu² PhD; Kongmeng Liew¹ PhD; Taichi Murayama¹ MSc; Shuntaro Yada¹ PhD; Shoko Wakamiya¹ PhD; Eiji Aramaki¹ PhD

¹Nara Institute of Science and Technology Ikoma JP

²Yahoo Japan Corporation Tokyo JP

Corresponding Author:

Eiji Aramaki PhD

Nara Institute of Science and Technology

8916-5 Takayama-cho

Ikoma

JP

Abstract

Background: COVID-19 has disrupted lives and livelihoods and caused widespread panic across communities and societies. Emerging reports suggest that people living in rural areas in some countries are more susceptible to COVID-19. However, there is a lack of quantitative evidence that can shed light on whether residents of rural areas are more concerned about COVID-19 than residents of urban areas.

Objective: This study investigated attitudes toward COVID-19 in different Japanese prefectures by aggregating and analyzing Yahoo! JAPAN search queries.

Methods: We measured COVID-19 concerns in each Japanese prefecture by aggregating search counts of the COVID-19-related queries of Yahoo! JAPAN users and the data related to COVID-19 cases. We then defined two indices—the localized concern index (LCI) and localized concern index by patient percentage (LCIPP)—to quantitatively represent the degree of concern. To investigate the impact of emergency declarations on people's concerns, we divided our study period into three phases according to the timing of the declaration of the state of emergency in Japan: before, during, and after. In addition, we evaluated the relationship between the LCI and LCIPP in different prefectures by correlating them with prefecture-level indicators of urbanization.

Results: Our results demonstrated that the concerns about COVID-19 in the prefectures changed in accordance with the declaration of the state of emergency. The correlation analyses also indicated that the differentiated types of public concern measured by the LCI and LCIPP reflect the prefectures' level of urbanization to a certain extent (i.e., the LCI appears to be more suitable for quantifying COVID-19 concern in urban areas, while the LCIPP seems to be more appropriate for rural areas).

Conclusions: We quantitatively defined Japanese Yahoo users' concerns about COVID-19 by using the search counts of COVID-19-related search queries. Our results also showed that the LCI and LCIPP have external validity.

(JMIR Preprints 23/04/2021:29865)

DOI: <https://doi.org/10.2196/preprints.29865>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in http://www.jmir.org/preprint/29865

Original Manuscript

Original Paper

Zhiwei Gao[†], Sumio Fujita^{††}, Nobuyuki Shimizu^{††}, Kongmeng Liew[†], Taichi Murayama[†], Shuntaro Yada[†], Shoko Wakamiya[†], Eiji Aramaki[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Japan

^{††} Yahoo Japan Corporation, Tokyo, Japan

*Corresponding author:

Eiji Aramaki, PhD.

Nara Institute of Science and Technology, Japan

8916-5, Takayama-cho, Ikoma, Nara 630-0192, Japan

Telephone: +81-743-72-6053

Email address: aramaki@is.naist.jp

Measuring Public Concern About COVID-19 Through Search Queries

Abstract

Background: COVID-19 has disrupted lives and livelihoods and caused widespread panic across communities and societies. Emerging reports suggest that people living in rural areas in some countries are more susceptible to COVID-19. However, there is a lack of quantitative evidence that can shed light on whether residents of rural areas are more concerned about COVID-19 than residents of urban areas.

Objective: This infodemiology study investigated attitudes toward COVID-19 in different Japanese prefectures by aggregating and analyzing Yahoo! JAPAN search queries.

Methods: We measured COVID-19 concerns in each Japanese prefecture by aggregating search counts of the COVID-19-related queries of Yahoo! JAPAN users and the data related to COVID-19 cases. We then defined two indices—the localized concern index (LCI) and localized concern index by patient percentage (LCIPP)—to quantitatively represent the degree of concern. To investigate the impact of emergency declarations on people's concerns, we divided our study period into three phases according to the timing of the declaration of the state of emergency in Japan: before, during, and after. In addition, we evaluated the relationship between the LCI and LCIPP in different prefectures by correlating them with prefecture-level indicators of urbanization.

Results: Our results demonstrated that the concerns about COVID-19 in the prefectures changed in accordance with the declaration of the state of emergency. The correlation analyses also indicated that the differentiated types of public concern measured by the LCI and LCIPP reflect the prefectures' level of urbanization to a certain extent (i.e., the LCI appears to be more suitable for quantifying COVID-19 concern in urban areas, while the LCIPP seems to be more appropriate for rural areas).

Conclusions: We quantitatively defined Japanese Yahoo users' concerns about COVID-19 by using the search counts of COVID-19-related search queries. Our results also showed that the LCI and LCIPP have external validity.

Keywords: COVID-19; search query; infodemiology; quantitative analysis; concern; rural; urban; Internet; information-seeking behavior; attitude; Japan;

Introduction

The COVID-19 pandemic has been threatening global health since the end of December 2019. The

outbreak has created critical challenges for public health, research, and medical communities [1]. As of December 14, 2020, COVID-19 has already affected 218 countries and territories with over 72 million confirmed cases, and has claimed over 1.6 million lives [2]. COVID-19 has also disrupted many lives and caused psychological trauma [3, 4] on a large scale.

As with any outbreak of an infectious disease, the population's psychological reactions play a critical role in shaping the spread of the disease and the occurrence of emotional distress and social disorder during and after the outbreak [5]. Recently, Ahorsu et al. [6] developed the Fear of COVID-19 Scale (FCV-19S) by conducting qualitative interviews to assess individuals' fear of COVID-19. Gao et al. [7] found that greater concern about COVID-19 (frequent exposure to COVID-19-related social media) was positively associated with adverse mental health outcomes. Furthermore, Su et al.'s [8] Twitter-based analysis revealed that spatial-temporal and socioeconomic disparities shaped the US residents' response to COVID-19.

Infodemiology is the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy [9]. The underlying idea of this field is to measure the pulse of public opinion, attention, behavior, knowledge, and attitudes by tracking what people do and write on the Internet [10], such as analyzing queries from Internet search engines to predict disease outbreaks. Bernardo et al. [11] presented a scoping review, assessed the current state of knowledge regarding the use of search queries and social media for disease surveillance and showing their usability. Daughton et al. [12] presented an infodemiology study using social media data from Twitter to identify human behaviors associated with COVID-19 transmission and the perceived impact of COVID-19 on individuals. Mavragani et al. [13] presented a study using Google Trends, provided a methodological framework for using Google Trends in infodemiology, analyzing the value and validity of using Google Trends data.

Worryingly, use of the Internet and social media has increased dramatically due to the enforcement of “social distancing” and “staying home” in many areas, which has led to more search queries for updates on the local COVID-19 situation [14]. However, there is a lack of quantitative evidence about the relationship between these searches for updates and the populational psychological reactions toward COVID-19. Additionally, one could think that residents of urban areas would show more concern about COVID-19 due to larger crowds and easier access to transportation. However, concern about COVID-19 is also prevalent in rural areas, given that people who live in rural areas may be more vulnerable to COVID-19 than residents of urban areas [15, 16].

Therefore, we aimed to quantitatively reflect the different types of concern in rural and urban areas by analyzing the Japanese public's psychological reactions toward COVID-19 based on the infodemiology (i.e., concern about COVID-19 in search queries). We used Yahoo! JAPAN to get the search queries in this study because it is the largest portal site in Japan with a significant user base: about 70 million and 20 million monthly active users for smartphone and PCs, respectively, covering about 85% of smartphone users and 61% of PC users in Japan. The proportion of female and male smartphone users is 52% and 48%, respectively, with about 32% of users aged 20 to 39, 40% of users aged 40 to 59, and 28% of users aged 60 and above [17]. We first developed two concern indices based on search queries to measure generalized COVID-19 concerns, which we called the localized concern index (LCI) and the localized concern index by patient percentage (LCIPP). We then used these indices to investigate COVID-19 concerns in relation to prefecture urbanization. To evaluate the feasibility of these concern indices, we examined the prefecture-level correlations with several indicators of ruralization and public health outcomes. The Methods section details the process of defining the LCI and LCIPP equations.

Methods

Target Queries

First, we explored people's COVID-19 concerns by analyzing search queries over different time periods. We initially established a baseline of common search queries prior to and during the pandemic by selecting the search queries of Yahoo! JAPAN's users from April to May 2019 and from April to May 2020 (our choice of examining queries during April and May was due to the Japanese government's declaration of a state of emergency in April 2020 [18]). Additionally, since the elderly people (over the age of 65) are at a significantly greater risk of adverse COVID-19 outcomes [19], we speculated that there might be more COVID-19-related search queries from this group. Hence, we started the analysis by targeting the search queries of people over 65 years of age. We extracted 100,000 search queries of this elderly population for the two aforementioned time periods and ranked them in reverse order according to the search counts. A change in time period results in a change in search query ranking; thus, we quantified the difference in ranking by defining the following rank change index:

$$\text{Rank change index}(q) = \frac{\log(\text{No. of } q \in 2020 + 2)}{\log(\text{No. of } q \in 2019 + 2)}$$

The larger the index, the fewer the search counts for this query in 2019 or the greater the search counts in 2020, and vice versa. A constant of two was added to both the numerator and denominator, to avoid instances of zeros (uncountable) in the search counts. Table 1 shows the top five rapidly increasing search queries in April 2020 compared to April 2019. As expected, these terms appeared to be COVID-19-related.

Table 1. Top five rapidly ascending search queries and their rank change index in April 2020 compared to April 2019. (The first one, "novel_coronavirus", is the original data we retrieved from the database.)

Search Query	Rank Change Index
novel_coronavirus	18.87
シャープのマスク (Sharp's face mask)	18.74
新型コロナウイルス (novel coronavirus)	17.99
コロナ 症例 (corona cases)	17.96
東京 新型コロナウイルス (Tokyo coronavirus cases)	17.63

Consistently, of the top 100 queries in the ascending query list, 76 queries contained COVID-19-related keywords (e.g., “コロナ [corona],” “マスク [mask]”), and of these, 33 queries contained the prefecture names with “コロナ 症例 (corona cases).” A query pattern, such as the prefecture name plus “coronavirus cases,” clearly displays the prefecture's information and it also reflects the user's concern about COVID-19 to some extent. Therefore, we chose this query pattern as our target query and sorted the queries by region according to the prefecture names mentioned in the query. Table 2 lists some query samples used in this study.

Table 2. Samples of target queries.

Search Query	Rank Change Index
東京 新型コロナウイルス (Tokyo coronavirus cases)	17.63
神奈川県 コロナ 症例 (Kanagawa corona cases)	16.26

東京都 (Saitama coronavirus cases)	16.18
福岡県 (Fukuoka coronavirus cases)	16.11
埼玉県 (Ibaraki coronavirus cases)	15.84

Baseline Queries

However, when we calculated the search counts of target queries that comprised the prefecture name and “coronavirus cases” from January to September 2020, we found that the search counts in Tokyo were much higher than in the other prefectures, as shown in Figure 1. It is possible that Tokyo’s larger population meant that the frequencies of the general search counts were higher than in other less populated prefectures. We speculated that this excessive disparity would inevitably have an impact on our subsequent calculations. To mitigate this effect, we introduced baseline queries for each prefecture that were frequently searched for in that prefecture and that had a relatively stable search count for a certain period. The baseline queries contained the prefecture name and “X,” where “X” referred to any keywords as long as the sample variance of their monthly search counts was as small as possible. For example, for Tokyo, “東京都 23 区 (Tokyo 23 wards)” and “東京都 天気 (past weather in Tokyo)” were some of the baseline queries, and their sample variances of the search counts in these nine months were 499.6 and 683, respectively. In summary, we compared the search counts of possible queries with the above pattern form from January to September 2020, and then identified the top three queries with the smallest sample variance in the nine months to form the baseline queries. This was done to balance the impact of the excessive disparity by using the quotient of the baseline query and our target query, as the search counts of the target queries were large for urbanized prefectures, such as Tokyo, where the search count frequencies of the baseline queries were also large.

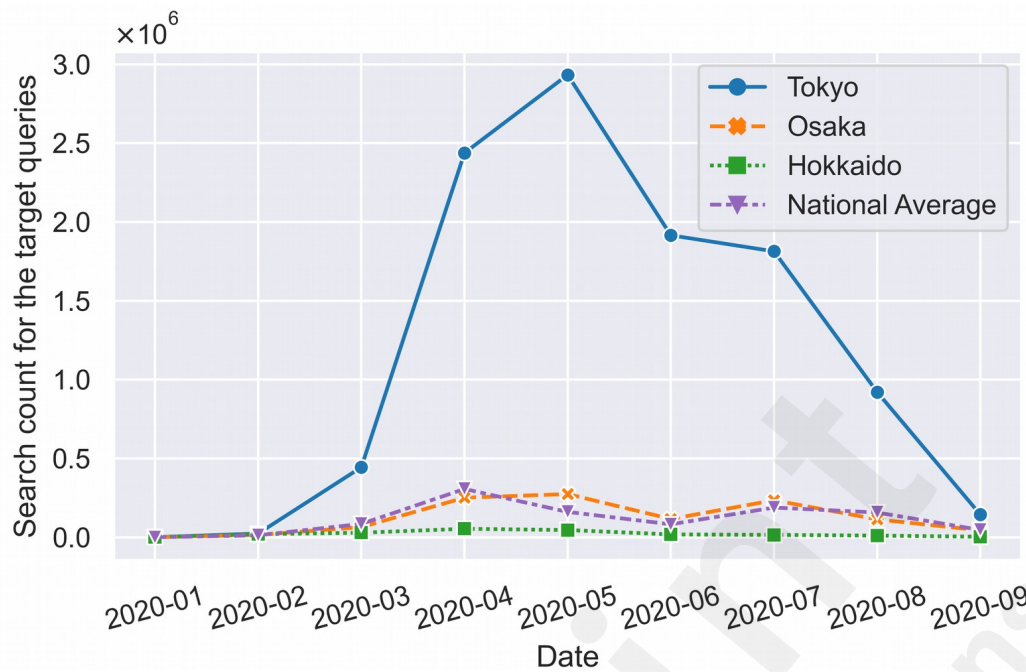
COVID-19 Concern Indices

Next, our query-based equation to quantify the level of concern about COVID-19—the LCI equation—was defined for each prefecture $pref$ as follows:

$$LCI_{pref} = \frac{Count(tq_{pref}) + 1}{Count(bq_{pref})}$$

where tq_{pref} and bq_{pref} are the target query and the baseline query, respectively, and $Count(.)$ is a function that counts the occurrences of a query. We used the logarithmic result for the LCI calculation, which suggests that a higher LCI means a higher frequency of baseline-controlled, prefecture-specified COVID-19-related queries, which in turn reflects a greater level of concern about COVID-19 in that prefecture.

Figure 1. Search counts of targeted queries in Tokyo, Osaka, Hokkaido and the national average from January to September.



However, the LCI appeared to be greatly influenced by the severity of the COVID-19 situation in each prefecture. For example, urbanized areas, such as Tokyo and Osaka, which are the first and second most populous cities in Japan and the hardest hit, naturally had the highest search counts for target queries and a higher LCI. As Gamonal Limcaoco et al. [20] mentioned, the COVID-19 pandemic has raised people's anxiety levels, and the LCI appears to reflect this attitude in hard-hit, largely urban areas. However, heightened concern about COVID-19 in rural areas [21] does not seem to be reflected in the LCI. Rather, given the lower number of cases in rural areas, it is likely that concern about COVID-19 in rural areas may be related to socio-ecological variables that are outside of the risk of infection.

For this reason, we attempted to improve our LCI equation to be able to examine COVID-19 concerns beyond the direct influence of actual COVID-19 cases. We argue that this would quantify the prevalence of concern about COVID-19 beyond the risk of actual infection. To do so, we modified the LCI to account for the number of new cases per month in each prefecture by the population of that prefecture to calculate the percentage of infected patients; that is, *No. of monthly new cases / population*.

In summary, the LCIPP for each prefecture $pref$ was as follows:

$$LCIPP_{pref} = \frac{LCI_{pref}}{\left(\frac{\text{No. of monthly new patients}_{pref} + 1}{\text{population}_{pref}} \right)}$$

Similarly, we used the logarithmic result for the LCIPP calculation.

Correlation Indicators

To examine whether the differentiated patterns of associations for the LCI and LCIPP reflected the COVID-19 concerns in rural and urban areas, we examined prefecture-level correlations with both indices. Accordingly, we included corresponding measures, examined the following aspects, with the corresponding indicators in parentheses: (1) the prevalence of farming (number of farming household, percentage of farmland, rice production), (2) population change (rate of population change), (3) ease of accessibility (reachable areas within one hour, travel time from Tokyo to major

stations in each prefecture), and (4) public health outcomes (reported symptoms, daily outpatients, COVID-19 cases per million residents).

We argue that rural areas can be identified by the higher prevalence of farming households, farming area, and rice production. Conversely, urban areas can be determined by the higher population density, ease of accessibility (the proportion of reachable areas within one hour), and rate of population change (urban areas should exhibit population growth, while rural areas should show a population decline). Finally, prefectural public health was measured as the number of cumulative COVID-19 cases per 1 million residents (as of September 30, 2020), the proportion of the population that reported general symptoms (non-COVID-19-related), the number of ambulance dispatches, and the number of daily outpatients in 2019. Public health outcomes are able to reflect the current state of local healthcare to some extent. Since COVID-19 outbreaks would take over existing health care resources, we aimed to find the relationship between public concern about COVID-19 and the current state of local healthcare through the public health outcomes indicators.

Results

LCI and LCIPP Results by Prefectures

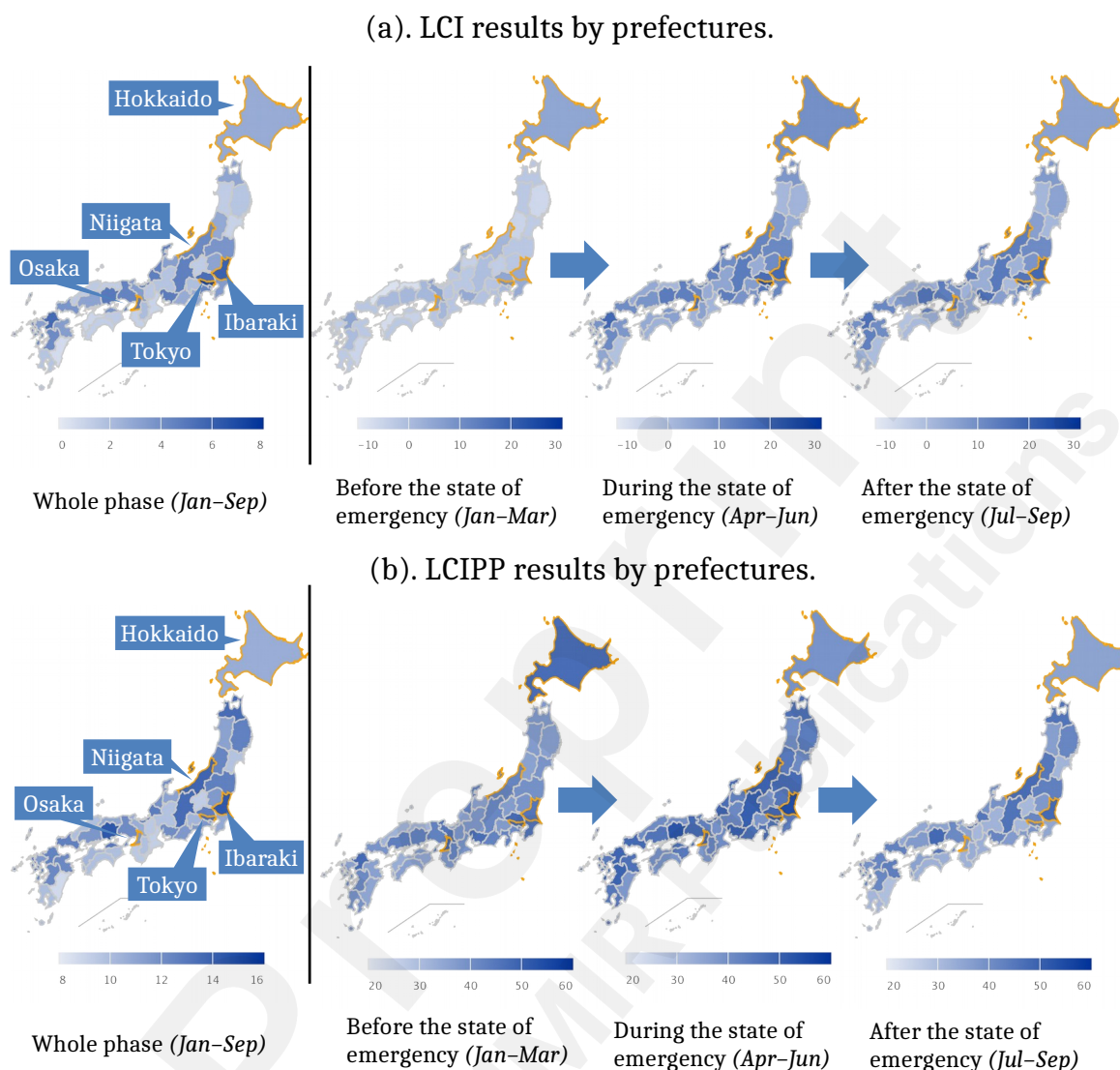
Figure 2(a) and Figure 2(b) show the geographical results of the LCI and LCIPP from January to September 2020, respectively. Table 3 shows the top prefectures with the highest and the lowest LCI and LCIPP. These are also shown as the darkest and the lightest regions, respectively, in Figure 2. In addition, we examined the LCI and LCIPP across three phases following the timing of the state of emergency: (1) before the state of emergency (January–March), (2) during the state of emergency (April–June), and (3) after the state of emergency (July–September). For a breakdown of the LCI and LCIPP by prefecture, please refer to our online supplementary material [22]. The LCI results show that before the state of emergency declaration, there was a low level of concern about COVID-19 nationwide. However, with the state of emergency declaration, some prefectures such as Tokyo and Ibaraki began to show a high level of concern. At the end of the state of emergency declaration, the overall level of concern in the country did not change much, though prefectures such as Hokkaido showed a slight decrease. The LCIPP results show a high level of concern for COVID-19 in Hokkaido even before the national state of emergency declaration. This is because Hokkaido activated COVID-19 response measures at an earlier stage (March 2020) [23], which resulted in a high level of concern for COVID-19 among local residents. With the nationwide state of emergency declaration in April 2020, most of the prefectures also showed a high level of concern, and with the end of the state of emergency declaration, the level of concern declined accordingly.

Table 3. Top prefectures with the highest and the lowest localized concern index and localized concern index by patient percentage.

Prefecture	LCI	Prefecture	LCI	Prefecture	LCIPP	Prefecture	LCIPP
Tokyo	7.04	Miyazaki	0.63	Okayama	14.36	Miyazaki	8.68
Ibaraki	5.90	Ehime	0.87	Ibaraki	14.31	Ehime	9.18
Fukuoka	5.54	Mie	0.96	Niigata	13.78	Gunma	9.23
Saitama	5.31	Miyagi	1.02	Nagano	13.48	Okinawa	9.30
Okayama	4.92	Wakayama	1.07	Aomori	13.46	Shiga	9.43

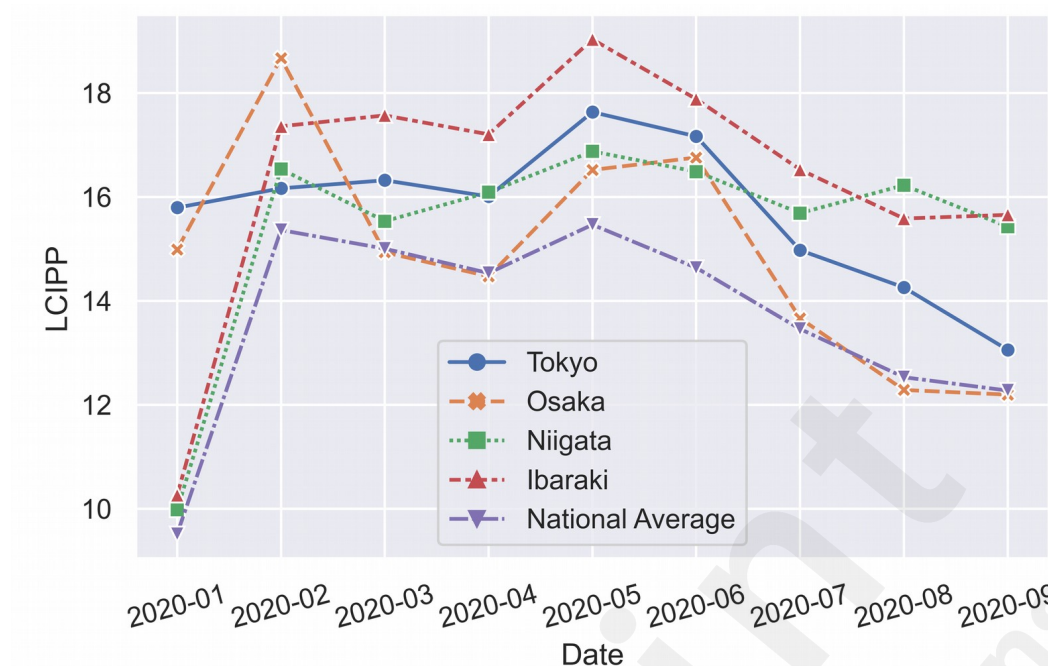
LCI = localized concern index, LCIPP = localized concern index by patient percentage.

Figure 2. Geographical results. The prefectures with orange boundary lines are the ones mentioned in this study: Hokkaido, Niigata, Ibaraki, Tokyo, and Osaka. LCI = localized concern index and LCIPP = localized concern index by patient percentage.



Of all the prefectures, we selected four to be specifically displayed as shown in Figure 3: Tokyo, Osaka, Niigata, and Ibaraki. We used the number of farming households as a criterion for distinguishing between urban and rural areas. According to a survey by the Statistics Bureau of Japan [24], Tokyo and Osaka have the lowest number of farming households in Japan, while Ibaraki and Niigata have the highest number. This suggests that Niigata and Ibaraki have relatively lower urbanization rates than Tokyo and Osaka. Nevertheless, Figure 3 indicates that some urban areas (i.e., Tokyo and Osaka) have a relatively lower LCIPP than some rural areas (i.e., Ibaraki) in terms of the general trends.

Figure 3. Localized concern index by patient percentage results of Tokyo, Osaka, Niigata, and Ibaraki and the national average from January to September. LCIPP = localized concern index by patient percentage.



Correlations with the LCI and LCIPP

By calculated the correlation coefficients of LCI and LCIPP on the aforementioned indicators, our results showed that the LCI was significantly correlated with the COVID-19 infection risk (cumulative cases), and urbanized prefectures. The latter was identified through the population density and ease of accessibility. The LCI was higher in prefectures that have a high proportion of reachable areas within one hour and in prefectures that had growing populations. In contrast, the LCIPP was higher in prefectures that were more rural. This was marked by prefectures that had a higher number of farming households, and rice production. Table 4 presents the correlation results.

Age-Based LCIPP Results

We also investigated the LCIPP results for three different age groups: (1) age 25–44 years, (2) age 45–64 years, and (3) age over 65. At this point, the *population* in the LCIPP equation was replaced by the number of people in each age group. Figure 4 shows that elderly people over the age of 65 had fewer concerns than the 25–44 and 45–64 age groups, while the 25–44 and 45–64 age groups had almost the same level of concerns about COVID-19.

Table 4. Correlation coefficients of the whole phase localized concern index and localized concern index by patient percentage and some indicators.

Indicator		LCI	LCIPP
No. of farming households	Pearson's <i>r</i>	0.147	0.403
	<i>P</i> -value	.323	.005
Percentage of farmland	Pearson's <i>r</i>	0.235	0.232
	<i>P</i> -value	.111	.116
Rice production	Pearson's <i>r</i>	0.020	0.309

	<i>P</i> -value	.895	.034
Rate of population change	Pearson's <i>r</i>	0.522	0.022
	<i>P</i> -value	<.001	.886
Reachable areas within one hour	Pearson's <i>r</i>	0.585	0.258
	<i>P</i> -value	<.001	.080
Travel time from Tokyo to major stations in each prefecture	Pearson's <i>r</i>	-0.264	-0.164
	<i>P</i> -value	.073	.270
Reported symptoms	Pearson's <i>r</i>	-0.053	-0.160
	<i>P</i> -value	.725	.289
Daily outpatients	Pearson's <i>r</i>	-0.146	-0.019
	<i>P</i> -value	.326	.897
COVID-19 cases per million residents	Pearson's <i>r</i>	0.465	-0.058
	<i>P</i> -value	<.001	.696

LCI = localized concern index, LCIPP = localized concern index by patient percentage.

Figure 4. Localized concern index by patient percentage results for ages 25–44, 45–64, and over 65.
LCIPP = localized concern index by patient percentage.



Discussion

Principal Findings

Nature of the LCI and LCIPP

As expected, the LCI appeared to represent the overall concern about COVID-19. To a large extent, this was influenced by the actual prevalence of COVID-19 within the prefecture. Similarly, concern about COVID-19 was heightened in prefectures that are dense, highly accessible, and growing in population. Most likely, these are reflections of urbanized prefectures that have a highly developed infrastructure and that attract migration from younger workers due to a higher number of job opportunities. As such, these fast-paced, young, and dense prefectures naturally present a greater risk of COVID-19 infection, and the increased concern shown in the search queries that was observed in the LCI is no surprise.

However, once we removed the variance explained by the inclusion of daily increases in COVID-19 cases from the equation, we found the opposite effect. The LCIPP ceased to reflect the COVID-19 risk in urban areas, and where was no significant relationship with the cumulative COVID-19 cases. Furthermore, the pattern of significant correlations revealed an association with the rural prefectures. The larger the proportion of farming households and the greater the rice production, the higher the public concern about COVID-19, which is beyond that explained by the risk of infection. Why does the LCIPP reflect increased ruralization? We posit an explanation in the form of collectivism afforded by farming, specifically in rice farming societies [25, 26]. These societies tend to be more collectivistic and residents have a greater psychological desire to protect the community from internal and external threats. COVID-19 is one such threat; thus, rural communities may have greater vigilance and concern about preventing COVID-19 from becoming prevalent in their community. This is also consistent with prior studies that have established links between COVID-19 concern and collectivism [27] and collectivism with COVID-19 prevention behavior [28].

This explanation has potential implications for policy. Specifically, the discrepancy between the LCI and LCIPP shows that more care must be taken when comparing public attitudes toward COVID-19 between rural and urban communities. If the LCIPP is indeed a result of collectivism and greater vigilance against COVID-19, this may also indicate a broader adoption of preventive measures, such as handwashing and mask-wearing. In contrast, the LCI does not seem to be similarly indicative of such concerns. For greater effectiveness, public campaigns that promote such behaviors should therefore use different strategies when targeting rural and urban societies.

Interestingly, we note that the public health measures in Table 3 did not correlate significantly with the LCIPP. This suggests that the preexisting or general health of a prefecture's population does not appear to affect public concern about COVID-19.

Preliminary Analysis of the Age-Based Results

Japan, which is one of the fastest aging countries, has the highest proportion of elderly people worldwide [29]. Emerging studies suggested that elderly people are more susceptible to COVID-19 and likely to have poor outcomes [19, 30, 31]. However, we found that individuals over 65 had reduced LCIPP scores, which suggests that their concern might be lower than those of the 25–44 and 45–64 age groups. Alternatively, this could be a consequence of Internet literacy, as users above 65 years of age may have less proficiency in using the Internet, or are simply less accustomed to using search terms and queries for topics of concern. However, more research is needed to contextualize this result.

Effectiveness of Search Queries as Public Concern Indicators

Finally, we evaluated the usefulness of our method by extracting search queries and combining them with actual COVID-19 infection rates to quantify public concern. Our results showed the correlations for the LCI and LCIPP demonstrate external validity, as they were both associated with constructs that could be explained by previous research. This study joins a growing body of literature that uses web-based search queries to track public health (e.g., Murayama et al. [32]), and the LCIPP adds the dimension of using prefecture-level infection rates to control for expected outcomes. Therefore, this study was able to effectively quantify public concern at a deeper level, which we propose is explained by the collectivistic psychological tendencies of a society.

Limitations

We note the limitations of our approach. Our method of extracting prefecture information from search queries relied on searches for the prefecture name + “coronavirus cases”, and not location-based information, such as IP addresses. This may not necessarily represent queries that are only

from residents of these prefectures; it may also reflect queries from non-residents who are interested in the COVID-19 situation in these prefectures (e.g., a user who may have family in these prefectures). In addition, although Yahoo JAPAN! has a large user base in Japan, but in the search engine market, such as Google still has a significant share. Search queries and results might vary between search engines due to user preferences and service provider settings. Therefore, from the perspective of data diversity, it would be a better choice to consider a combination of search queries from multiple search engines.

Conclusions

In summary, the present research utilized search queries from Yahoo! JAPAN users to quantify the degree of concern about COVID-19 in rural and urban areas. We first established that Yahoo search queries could be used to quantify COVID-19-related concerns. Next, we defined the LCI and LCIPP as quantitative indicators of prefecture-level COVID-19 concern. The LCI was indicative of COVID-19 concern in urban areas, whereas the LCIPP appeared to be indicative of COVID-19 concern in rural prefectures. By investigating the relationships between these concern indices and prefecture-level information, we showed that the LCI and LCIPP have external validity. These results suggest that one potential application could be conducting differentiated public campaigns about COVID-19 prevention and misinformation. Because of the different sources of concern in rural and urban areas, such campaigns could adopt different risk communication strategies in these areas.

Acknowledgements

This study was supported in part by JSPS KAKENHI Grant Number JP19K20279 and Health and Labor Sciences Research Grant Number H30-shinkougousei-shitei-004.

Abbreviations

LCI: Localized concern index

LCIPP: Localized concern index by patient percentage

References

1. Fauci AS, Lane HC, Redfield RR. Covid-19 — Navigating the Uncharted. *New England Journal of Medicine*. 2020;382(13):1268-9. PMID: 32109011. doi: 10.1056/NEJMe2002387.
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*. 2020;20(5):533-4. doi: 10.1016/S1473-3099(20)30120-1.
3. Wang C, Tee M, Roy AE, Fardin MA, Srichokchatchawan W, Habib HA, et al. The impact of COVID-19 pandemic on physical and mental health of Asians: A study of seven middle-income countries in Asia. *PLOS ONE*. 2021;16(2):e0246824. doi: 10.1371/journal.pone.0246824.
4. Zacher H, Rudolph CW. Individual differences and changes in subjective wellbeing during the early stages of the COVID-19 pandemic. *Am Psychol*. 2021 Jan;76(1):50-62. PMID: 32700938. doi: 10.1037/amp0000702.
5. Cullen W, Gulati G, Kelly BD. Mental health in the COVID-19 pandemic. *Qjm*. 2020 May 1;113(5):311-2. PMID: 32227218. doi: 10.1093/qjmed/hcaa110.
6. Ahorsu DK, Lin C-Y, Imani V, Saffari M, Griffiths MD, Pakpour AH. The Fear of COVID-19 Scale: Development and Initial Validation. *Int J Ment Health Addict*. 2020;1-9. PMID: 32226353. doi: 10.1007/s11469-020-00270-8.
7. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. *PloS one*. 2020;15(4):e0231924-e. PMID: 32298385. doi: 10.1371/journal.pone.0231924.
8. Su Y, Venkat A, Yadav Y, Puglisi LB, Fodeh SJ. Twitter-based analysis reveals differential COVID-19 concerns across areas with socioeconomic disparities. *Computers in Biology and*

Medicine. 2021 2021/03/13/:104336. doi: 10.1016/j.compbmed.2021.104336.

9. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res*. 2009 Mar 27;11(1):e11. PMID: 19329408. doi: 10.2196/jmir.1157.

10. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med*. 2011 May;40(5 Suppl 2):S154-8. PMID: 21521589. doi: 10.1016/j.amepre.2011.02.006.

11. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. *J Med Internet Res*. 2013 Jul 18;15(7):e147. PMID: 23896182. doi: 10.2196/jmir.2740.

12. Daughton AR, Shelley CD, Barnard M, Gerts D, Watson Ross C, Crooker I, et al. Mining and Validating Social Media Data for COVID-19–Related Human Behaviors Between January and July 2020: Infodemiology Study. *J Med Internet Res*. 2021 2021/5/25;23(5):e27059. doi: 10.2196/27059.

13. Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. *JMIR Public Health Surveill*. 2019;5(2):e13439-e. PMID: 31144671. doi: 10.2196/13439.

14. Koeze E, Popper N. The Virus Changed the Way We Internet. 2020; Available from: <https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html>. Accessed: January 25, 2021.

15. Leatherby L. The Worst Virus Outbreaks in the U.S. Are Now in Rural Areas. 2020; Available from: <https://www.nytimes.com/interactive/2020/10/22/us/covid-rural-us.html>. Accessed: January 25, 2021.

16. Singh KD, Gettleman J. ‘Rural Surge’ Propels India Toward More Covid-19 Infections Than U.S. 2020; Available from: <https://www.japantimes.co.jp/news/2020/10/09/asia-pacific/rural-surge-india-coronavirus-infections/>. Accessed: January 25, 2021.

17. Yahoo Japan Corporation. Yahoo! JAPAN Overview. 2021; Available from: https://s.yimg.jp/images/listing/pdfs/yj_mediaguide_eng.pdf. Accessed: May 26, 2021.

18. Prime Minister of Japan and His Cabinet. [COVID-19] Guidelines for Lifting the State of Emergency. 2020; Available from: https://japan.kantei.go.jp/ongoingtopics/_00025.html. Accessed: January 18, 2021.

19. Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? *Aging (Albany NY)*. 2020;12(10):9959-81. PMID: 32470948. doi: 10.18632/aging.103344.

20. Gamonal Limcaoco RS, Mateos EM, Fernández JM, Roncero C. Anxiety, worry and perceived stress in the world due to the COVID-19 pandemic, March 2020. Preliminary results. *medRxiv*. 2020:2020.04.03.20043992. doi: 10.1101/2020.04.03.20043992.

21. Seidansha. (Japanese) Returning home after the Corona disaster. 2020; Available from: <https://bunshun.jp/articles/-/39985>. Accessed: January 25, 2021.

22. Available from: <https://aoi.naist.jp/concern>. Accessed: January 30, 2021.

23. Kyodo. Hokkaido set to lift coronavirus state of emergency. 2020; Available from: <https://www.japantimes.co.jp/news/2020/03/19/national/hokkaido-set-to-lift-coronavirus-emergency/>. Accessed: January 23, 2021.

24. Statistics Bureau of Japan. Japan Statistical Yearbook 2020. 2020; Available from: <https://www.stat.go.jp/english/data/nenkan/69nenkan/1431-08.html>. Accessed: January 20, 2021.

25. Liu SS, Morris MW, Talhelm T, Yang Q. Ingroup vigilance in collectivistic cultures. *Proceedings of the National Academy of Sciences*. 2019;116(29):14538-46. doi: 10.1073/pnas.1817588116.

26. Uchida Y, Takemura K, Fukushima S, Saizen I, Kawamura Y, Hitokoto H, et al. Farming cultivates a community-level shared culture through collective activities: Examining contextual effects with multilevel analyses. *J Pers Soc Psychol*. 2019 Jan;116(1):1-14. PMID: 30382740. doi:

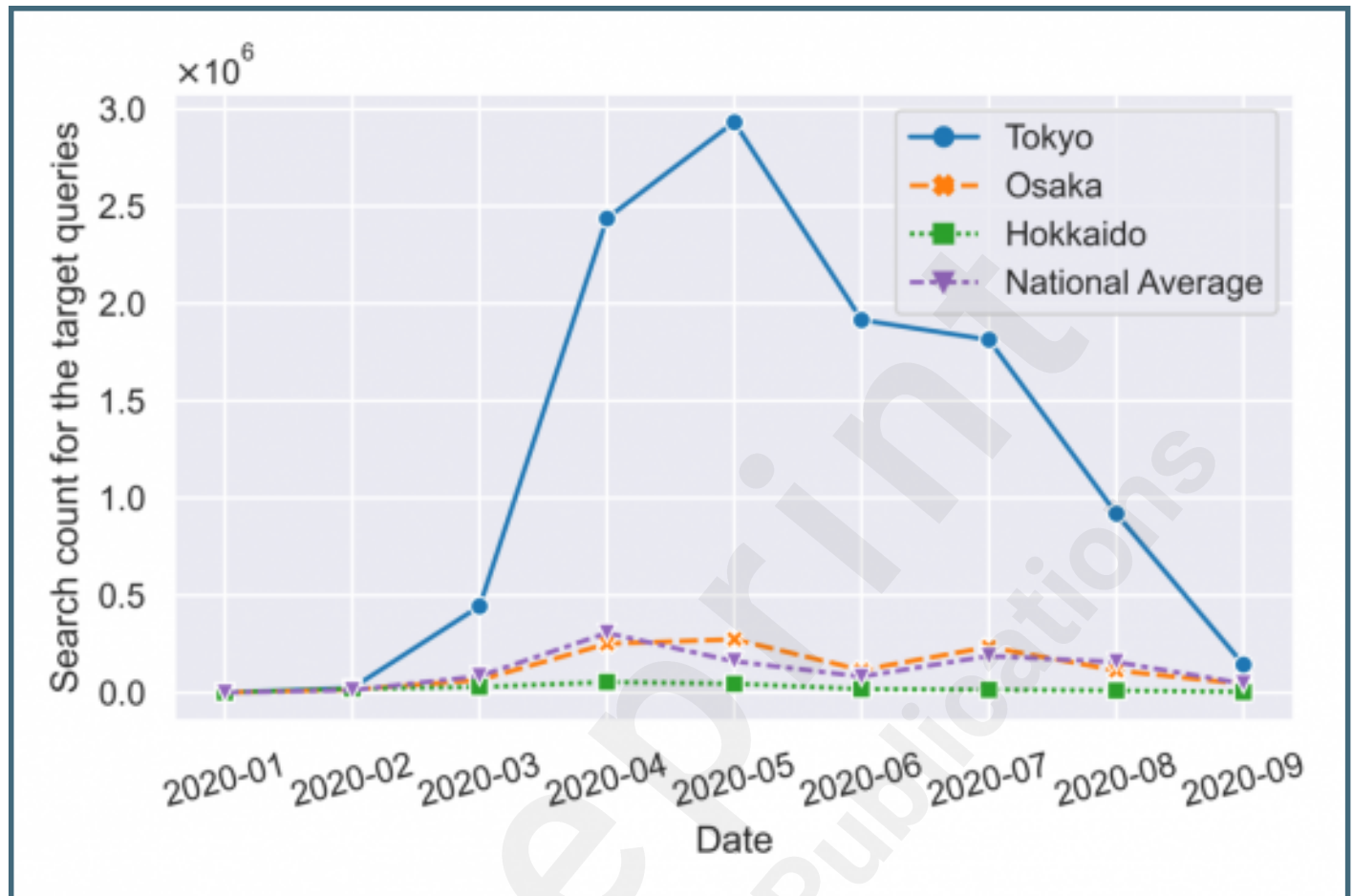
10.1037/pspa0000138.

27. Germani A, Buratta L, Delvecchio E, Mazzeschi C. Emerging Adults and COVID-19: The Role of Individualism-Collectivism on Perceived Risks and Psychological Maladjustment. *Int J Environ Res Public Health*. 2020;17(10):3497. PMID: 32429536. doi: 10.3390/ijerph17103497.
28. Huang F, Ding H, Liu Z, Wu P, Zhu M, Li A, et al. How fear and collectivism influence public's preventive intention towards COVID-19 infection: a study based on big data from the social media. *BMC Public Health*. 2020 2020/11/16;20(1):1707. doi: 10.1186/s12889-020-09674-6.
29. Muramatsu N, Akiyama H. Japan: super-aging society preparing for the future. *Gerontologist*. 2011 Aug;51(4):425-32. PMID: 21804114. doi: 10.1093/geront/gnr067.
30. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. *BMJ*. 2020;368:m1198. doi: 10.1136/bmj.m1198.
31. Aw D, Woodrow L, Ogliari G, Harwood R. Association of frailty with mortality in older inpatients with Covid-19: a cohort study. *Age and Ageing*. 2020;49(6):915-22. doi: 10.1093/ageing/afaa184.
32. Murayama T, Shimizu N, Fujita S, Wakamiya S, Aramaki E. Robust two-stage influenza prediction model considering regular and irregular trends. *PLOS ONE*. 2020;15(5):e0233126. doi: 10.1371/journal.pone.0233126.

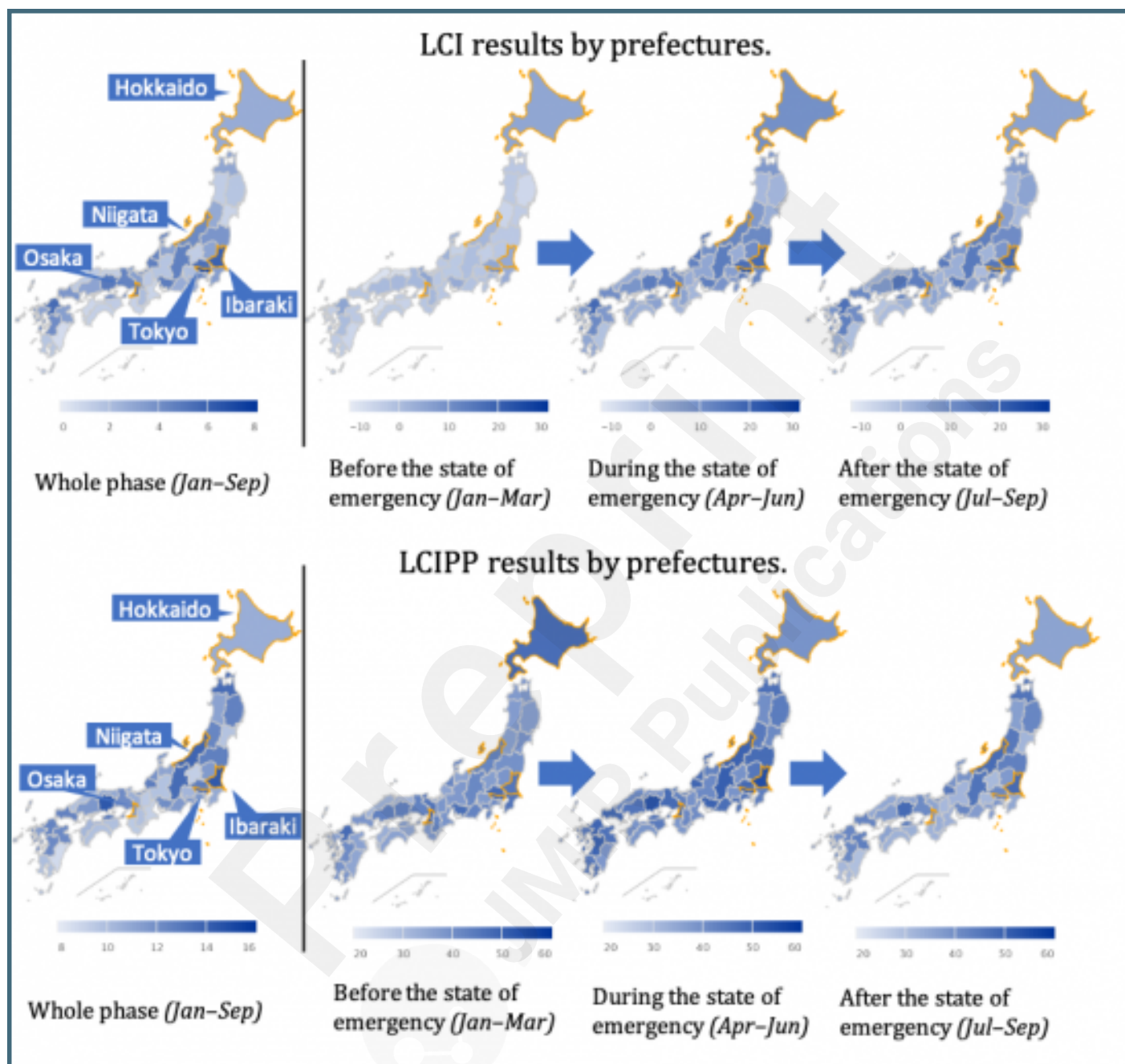
Supplementary Files

Figures

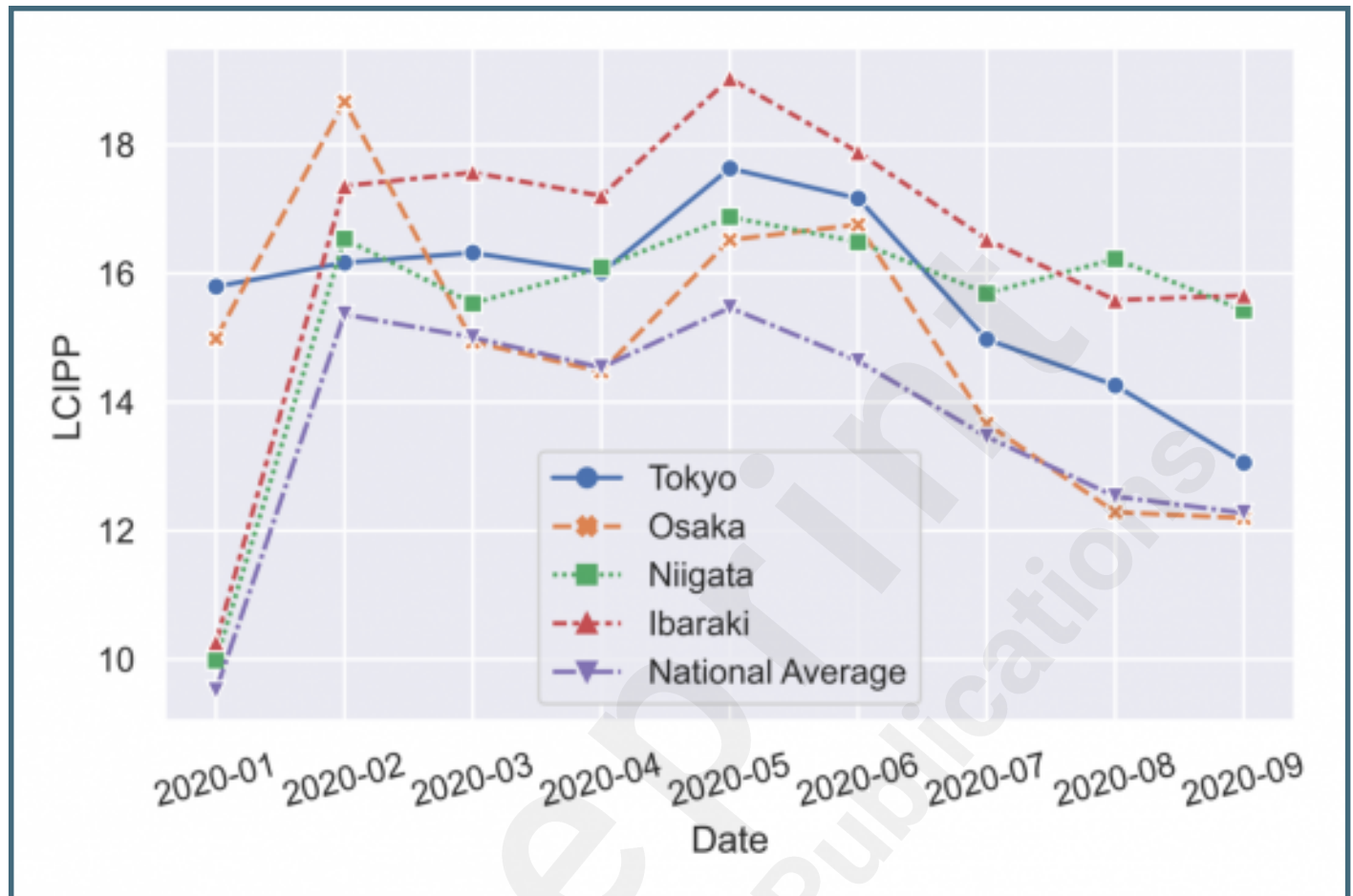
Search counts of targeted queries in Tokyo, Osaka, Hokkaido and the national average from January to September.



Geographical results. The prefectures with orange boundary lines are the ones mentioned in this study: Hokkaido, Niigata, Ibaraki, Tokyo, and Osaka. LCI = localized concern index and LCIPP = localized concern index by patient percentage.



Localized concern index by patient percentage results of Tokyo, Osaka, Niigata, and Ibaraki and the national average from January to September. LCIPP = localized concern index by patient percentage.



Localized concern index by patient percentage results for ages 25–44, 45–64, and over 65. LCIPP = localized concern index by patient percentage.

