Original Paper

Identifying Potential Norovirus Epidemics in China via Internet Surveillance

Kui Liu[1,2*], MSc; Sichao Huang[3*], MSc; Zi-Ping Miao[1,2*], BSc; Bin Chen[1,2], MSc; Tao Jiang[1], MPH; Gaofeng Cai[1], MPH; Zhenggang Jiang[1], MSc; Yongdi Chen[1], MPH; Zhengting Wang[1], MPH; Hua Gu[1], MD, PhD; Chengliang Chai[1,2], MPH; Jianmin Jiang[1,2], MD, PhD

[1]Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China

[2]Key Laboratory of Vaccine, Prevention and Control of Infectious Disease of Zhejiang Province, Hangzhou, China

[3]School of Medicine, Ningbo University, Ningbo, China

*these authors contributed equally

Corresponding Author:
Jianmin Jiang, MD, PhD
Zhejiang Provincial Center for Disease Control and Prevention
3399 Binsheng Rd
Binjiang District
Hangzhou, 310051
China
Phone: 86 57187115009
Email: jmjiang@cdc.zj.cn

Abstract

Background: Norovirus is a common virus that causes acute gastroenteritis worldwide, but a monitoring system for norovirus is unavailable in China.

Objective: We aimed to identify norovirus epidemics through Internet surveillance and construct an appropriate model to predict potential norovirus infections.

Methods: The norovirus-related data of a selected outbreak in Jiaxing Municipality, Zhejiang Province of China, in 2014 were collected from immediate epidemiological investigation, and the Internet search volume, as indicated by the Baidu Index, was acquired from the Baidu search engine. All correlated search keywords in relation to norovirus were captured, screened, and composited to establish the composite Baidu Index at different time lags by Spearman rank correlation. The optimal model was chosen and possibly predicted maps in Zhejiang Province were presented by ArcGIS software.

Results: The combination of two vital keywords at a time lag of 1 day was ultimately identified as optimal (ρ=.924, P<.001). The exponential curve model was constructed to fit the trend of this epidemic, suggesting that a one-unit increase in the mean composite Baidu Index contributed to an increase of norovirus infections by 2.15 times during the outbreak. In addition to Jiaxing Municipality, Hangzhou Municipality might have had some potential epidemics in the study time from the predicted model.

Conclusions: Although there are limitations with early warning and unavoidable biases, Internet surveillance may be still useful for the monitoring of norovirus epidemics when a monitoring system is unavailable.

Introduction

Acute gastroenteritis, inflammation of the gastrointestinal tract, is defined as the sudden onset of diarrhea, with or without signs of nausea, vomiting, fever, or abdominal pain [1,2]. The known pathogenic causes of acute gastroenteritis include various infective pathogens and other noninfectious causes. Infectious acute gastroenteritis is generally caused by enteric viruses, bacteria, and protozoal pathogens. Norovirus, a single-stranded RNA virus of the *Caliciviridae* family, is a leading cause of infectious acute gastroenteritis worldwide across all age groups, particularly in health care and community settings [3,4]. In the

XSL•FO
RenderX

United States, norovirus causes nearly 21 million cases of acute gastroenteritis annually, and nearly 50% of acute gastroenteritis occurrences across Europe were attributed to the norovirus infection [5,6]. Although norovirus infection is viewed as a self-limited illness, it might be still responsible for severe dehydration, and even potential death, in children and the elderly population [7-9]. Infection might be attributable to closed bedside care and exposure to vomit of contaminated food, water and aerosol of norovirus, and other factors such as prolonged time for viral shedding and the enhanced viability in the environment might play vital roles in improving transmissibility of norovirus [4,10-12]. Additionally, no obvious evidence supports the idea of there being a specific reservoir, and only the scattered speculation from available literature viewed immunocompromised individuals, elderly, and malnourished hosts as potential norovirus reservoirs [13]. Available studies from surveillance systems demonstrate that 0.7% of reported outbreaks were foodborne, 28.5% were person-to-person, and the remaining 70.8% were unclear or not described [10]. Transmission of norovirus commonly occurs before the typical symptoms appear, which further increases the difficulty for interventions [11]. Thus, how to effectively supervise and control norovirus infection has aroused substantial concern around the world.

Norovirus was ranked the second most common etiological agent only after rotavirus in children younger than 5 years in China [14-16]. Field surveys imply that unboiled water and contaminated food are the common causes of norovirus infection [12,17]. The high contagiosity, frequent virus mutation, and limited immune protection has resulted in more frequent epidemics of norovirus outbreak on the Chinese mainland since the winter of 2014, especially in schools [18-20]. Norovirus infection is classified as "other infectious diarrhea" (excluding the illness of cholera, bacillary dysentery, amebic dysentery, and typhoid/paratyphoid) in the Chinese National Notifiable Infectious Disease Reporting System. Accordingly, norovirus cases are not reported in an independent module, and some norovirus cases manifesting the main symptom of vomiting are inevitably omitted as well. The Public Health Emergency Management Information System can only focus on some clustered epidemics; therefore, some sporadic cases or subclinical infections of norovirus may be omitted. Thus, effective interventions for norovirus at the early stage are a pressing issue for public health.

With the rapid development of Internet technology, an increasing number of researchers have tried to take advantage of the retrieval function of search engines to forecast and warn against infectious diseases [21,22]. In China, both wired and wireless networks have been booming to meet the increasing demands of the cyber citizens. The China Internet Development Statistical Report released on July 23, 2015, revealed a total of 668 million Internet users (48.8% of the population), 18.94 million more than 6 months previous [23]. As the most frequently used search engine, Baidu has a priority selection incidence of 89.1% among Chinese cyber users [24]. The Baidu Index, based on the search frequency of some keywords within the Baidu search engine, can be viewed as awareness and requirement of cyber users [25]. Evidence from previous studies of communicable diseases suggested a potential relationship between search volume and the number of infected cases [26,27]. One study also used the comprehensive Baidu Index to construct a linear regression model to predict the potential cases of epidemic erythromelalgia, suggesting that the Baidu Index may serve as a good early indicator for epidemics [22]. Given that there is no specific monitor system for norovirus in China, the purpose of this study was to determine whether Internet surveillance was a helpful supplement to traditional surveillance of norovirus epidemics in China.

Methods

Ethics

This study was approved by the Ethics Committee of Zhejiang Provincial Center for Disease Control and Prevention. Given no privacy information involving human participants, it was granted an exemption from informed consent.

Information of Epidemic

Clustered cases in schools of diarrhea and vomiting were notified in Haining and Haiyan, two counties in the Jiaxing Municipality, on February 17, 2014. After receiving notice, a field investigation conducted by the Zhejiang Provincial Center for Disease Control and Prevention was performed to find the potential causes. Using a standard questionnaire, potential cases were searched within seven schools in Haiyan and six schools in Haining. Through the field survey, the first case was identified on February 12, 2014, and the epidemic lasted for 10 days. The inclusion criteria of possible cases were vomiting more than once or having diarrhea more than three times in one day, including probable cases and confirmed cases. There were a total of 924 cases in this outbreak with a ratio of 1 male to 1.2 females (Figure 1).

Figure 1. Daily cases of norovirus outbreak in Jiaxing Municipality from February 12 to 21, 2014, within the counties of Haiyan and Haining.
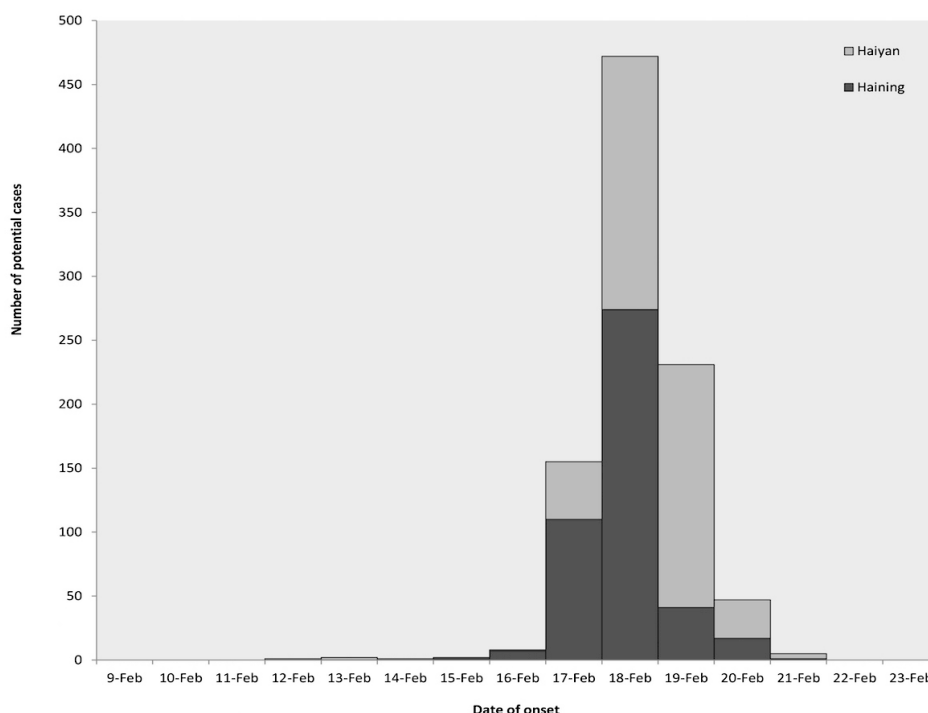
Baidu Index

The Baidu Index comprehensively reflects media exposure and users' concerns based on certain keywords used by cyber users in the past day. Although the specific algorithm of Baidu Index is not available to the public because it is proprietary information, it was proven to be similar to Google Flu Trend in identifying public behavior on the Internet for different areas on diverse days [25]. Given the potential time lag between the onset of the symptoms in cases and related Internet searching, we collected the Baidu Index of Jiaxing Municipality from February 10 to 28, 2014. The data in the same period during 2013 and 2015 were also extracted.

Keyword Screening and Data Collection

In China, the same idea can be expressed with diverse characters among different populations. That is to say, the retrieval of disease-related information may be distinctive through the search engine. Thus, how to recognize keywords specific to norovirus was vital for Internet surveillance. Furthermore, no standardized guidance was available for this issue; the disease names and main presentations had been commonly chosen as the primary keywords [21,22,28]. In this study, the primary keywords chosen (in Chinese) were "norovirus," "nausea," "emesis," "abdominal pain," and "diarrhea." More norovirus-related keywords searched by cyber users were acquired at Keywords Mining, a website that uses semantic correlation analysis [29]. All

keywords obtained from the website were stemmed from search engines and also websites, blogs, and other online sources.

We retrieved from this website the top 100 keywords for each of the five initial keywords. Two individuals evaluated these 500 keywords to exclude unrelated ones. In case of discrepancy in opinion, a third person made the final decision. Then, correlations between potential case number and Baidu Index of the screened keywords with possible time lags were calculated. If *P*<.05 combined with a Spearman rank correlation coefficient (ρ) >0.4, the keyword was brought into the group with the special time lag. Similar to a previous study, we chose the onset of illness as the beginning of the study time and had five groups with time lags of 0 to 4 days [22].

Composite Baidu Index

After excluding keywords with no significance, the meaningful keywords were grouped by different time lags. In each group, the weight of each keyword (weight$_{ti}$) and composite Baidu Index (composite BDI$_{ti}$) were calculated as equation a and b in Figure 2, where *t* represents the potential time lag, *i* indicates the order number of present keyword, *n* is the number of keywords included in a specific time lag, ρ$_{ti}$ is the Spearman rank correlation of included keyword (i) with specific time lag (t), and *BDI*$_{ti}$ denotes the daily Baidu Index of included keywords (i) with specific time lag (t) [22].

Figure 2. Equations used in the study.

$$\text{weight}_{ti} = \frac{\rho_{ti}}{\sum_{i=1}^{n} \rho_{ti}} \qquad \text{a}$$

$$\text{composite } BDI_{ti} = \sum_{i=1}^{n} \text{weight}_{ti} \, BDI_{ti} \qquad \text{b}$$

$$y_{\text{LRM}} = b_0 + (b_1 * x) \qquad \text{c}$$

$$y_{\text{QCRM}} = b_0 + (b_1 * x) + (b_2 * x^2) \qquad \text{d}$$

$$y_{\text{CCRM}} = b_0 + (b_1 * x) + (b_2 * x^2) + (b_3 * x^3) \qquad \text{e}$$

$$y_{\text{ECM}} = b_0 * e^{b_1 * x} \qquad \text{f}$$

$$y_{\text{GCM}} = e^{(b_0 + (b_1 * x))} \qquad \text{g}$$

$$\text{OCR} = (1 - \frac{|T_p - T_r|}{T_r}) * 100\% \qquad \text{h}$$

Optimal Time Lag and Model Construction

According to the Spearman rank correlation of the composite Baidu Index and potential case number in each time lag, the time lags with the superior coefficient were determined optimal. Additionally, the mean composite Baidu Index normalized by local netizens (1/100 million) was calculated to avoid potential bias stemming from the distinction of Internet users in diverse area (Multimedia Appendix 1). Prediction models were constructed to explore the relationship of the mean composite Baidu Index and potential case number under diverse optimal lag periods, respectively.

Given the distribution characteristics of potential norovirus cases, we constructed a linear regression model and latent curve models. In this study, five possible models shown in Figure 2 (c-g) were evaluated, including linear regression model (LRM), quadratic curve regression model (QCRM), cubic curve regression model (CCRM), exponential curve model (ECM), and growth curve model (GCM).

In these models, b_1, b_2, and b_3 indicate the coefficients; x and y represent the mean composite Baidu Index and case number, respectively. The optimal model was examined by the P value of variance (ANOVA) and t test for coefficient. The overall coincidence rate (OCR) shown in Figure 2 (h) would be employed to determine the optimal model, in which T_p represents the total predicted case number from the specific prediction model during the study period, and T_r denotes the case number during the study period. The more the value of OCR trended to 100%, the better the selected model would be.

Spatial Presentation for Predicted Norovirus

The composite Baidu Index of 11 municipalities in Zhejiang Province was calculated from February 10 to 28, 2014. Based on the optimal prediction model described previously, the predicted norovirus cases in 11 municipalities were acquired and shown by ArcGIS software. The predicted norovirus cases in the same period of 2013 and 2015 were also shown on the map.

Statistical Analysis

All analyses were performed using SPSS Statistics 20.0 (SPSS Inc, Chicago, IL, USA). Spatial display was done with the Geographic Information System version 10.1 (SERI Inc, Redlands, CA, USA). Results were considered statistically significant if $P<.05$ with two sides.

Results

Epidemiological Characteristics of the Norovirus Outbreak

This epidemic was first reported in Haiyan, followed by Haining, within the Jiaxing Municipality in Zhejiang Province of China. A total of 924 cases (420 male and 504 female) involving 13 schools were detected from February 12 to 21, 2014; five were teachers and the rest were students. The clinical symptoms were mild and the main symptom was vomiting accompanied by nausea, diarrhea, fever, and abdominal pain, but no death occurred. The local departments of disease control in both counties responded rapidly to the epidemic, with such interventions as class suspension, disinfection within dormitories and classrooms, and sealing barreled water. Vomitus and anal swabs were retained from some cases. Given the potential exposure to barreled water among most cases, local centers for disease control sampled the different brands of barreled water in the schools. After 15:00 on February 21, no new cases were reported. Laboratory tests detected norovirus genogroup II in eight samples of anal swabs, five samples of opened barreled

water, and one sample of unopened barreled water, suggesting that water contaminated by norovirus caused the epidemic. Through in-depth investigation of the drinking water, we eventually deemed the occurrences at two sites as one outbreak because the contaminated water was supplied by the same supplier. This detailed information had been described in another study [30].

Optimal Time Lags of the Composite Baidu Index

Five possible time lags (0, 1, 2, 3, and 4 days) were considered to screen the considerable time lags, and details of all inclusion keywords for each time lag are listed in Table 1. The correlation coefficient peaked at the time lag of 2 days with five keywords (ρ=.945, *P*<.001). Considering the potential epidemiology significance and delicate difference at the time lag of 1 day with two keywords (ρ=.924, *P*<.001), both time lags were included to construct appropriate models.

Prediction Model for Norovirus

The composite Baidu Index was calculated with the weight and Baidu Index of the different keywords at lag times of 1 and 2 days, respectively. After the standardization of local netizens, the regression models were then constructed to predict the potential norovirus cases by the mean composite Baidu Index independent variable. Of the five candidate models considered in our study, ECM was determined the optimal model for the time lag of 1 day (Figure 3), whereas the top model for the time lag of 2 days was GCM (Table 2). Then, OCR values of both models in Jiaxing Municipality were calculated, demonstrating that OCR in ECM was 90.69% and in GCM was 66.00% (Table 3). Consequently, the optimal model was decided as ECM with 1 day lag. In this model, $y=1.809*e^{0.764*x}$, which was interpreted as a one-unit increase in the mean composite Baidu Index contributed to an increase of norovirus infections by 2.15 times during the outbreak.

Table 1. Inclusive keywords at time lags of 0 to 4 days after screening.

Time lag and keyword	Indicators for keyword			Indicators for composition	
	ρ	*P*	Weight	ρ	*P*
Day 0				.740	.01
Norovirus	.740	.01	0.348		
Noro	.735	.02	0.346		
Vomiting and bleeding	.650	.04	0.306		
Day 1				.924	<.001
Noro	.950	<.001	0.507		
Norovirus	.924	<.001	0.493		
Day 2				.945	<.001
Norovirus	.945	<.001	0.237		
Noro	.932	<.001	0.234		
Vomiting and diarrhea	.715	.02	0.180		
Nausea and vomiting	.701	.02	0.176		
Viral diarrhea in infants	.688	.03	0.173		
Day 3				.707	.02
Norovirus	.707	.02	0.266		
Why feel headache and nausea	.665	.04	0.250		
Noro	.648	.04	0.243		
Why feel dizziness and nausea	.642	.045	0.241		
Day 4					
Why feel headache and nausea	.673	.03	1	.673	.03

Figure 3. Fluctuant trend of potential case number and mean composite Baidu Index at the time lag of 1 day during February 12 to 21, 2014.
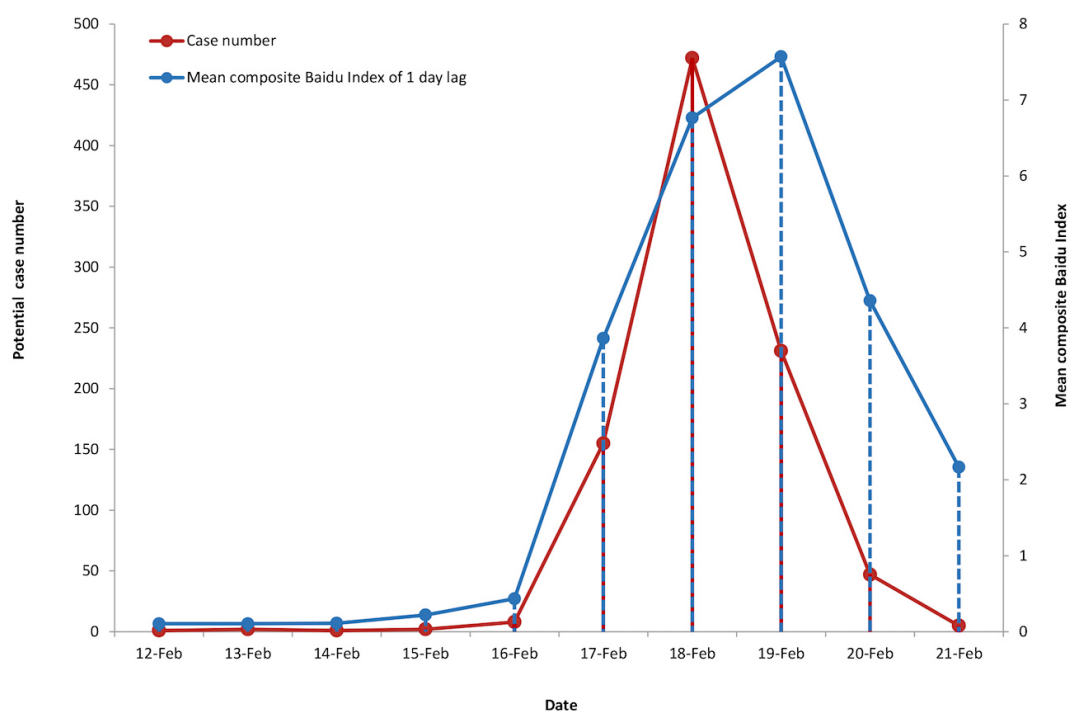

Table 2. Details of model screening for five potential candidate models.

Time lag and model	F (df1,df2)	P	R^2	Coefficient							
				b_0	P	b_1	P	b_2	P	b_3	P
Day 1											
LRM[a]	18.292 (1,8)	.003	.696	−22.213	.59	44.610	.003				
QCRM[b]	9.254 (2,7)	.01	.726	−4.24	.93	10.957	.79	4.827	.41		
CCRM[c]	7.901 (3,6)	.02	.798	30.292	.56	−151.010	.24	66.333	.17	−5.496	.19
ECM[d]	59.664 (1,8)	<.001	.882	1.809	.03	0.764	<.001				
GCM[e]	59.664 (1,8)	<.001	.882	0.593	.15	0.764	<.001				
Day 2											
LRM	19.215 (1,8)	.002	.706	−35.840	.40	95.996	.002				
QCRM	14.546 (2,7)	.003	.806	7.440	.86	−30.034	.68	36.277	.10		
CCRM	9.450 (3,6)	.01	.825	−11.621	.82	103.430	.58	−67.747	.62	19.205	.45
ECM	41.870 (1,8)	<.001	.840	1.535	.06	1.593	<.001				
GCM	41.870 (1,8)	<.001	.840	0.428	.38	1.593	<.001				

[a]LRM: linear regression model.

[b]QCRM: quadratic curve regression model.

[c]CCRM: cubic curve regression model.

[d]ECM: exponential curve model.

[e]GCM: growth curve model.

Table 3. The overall coincidence rate (OCR) value of the exponential curve model (ECM) and growth curve model (GCM) with different time lags.

Indicators	ECM with 1 day time lag	GCM with 2 days time lag
Total predicted case number	1010	610
Case number during the study period	924	924
OCR	90.69%	66.00%

Figure 4. Predicted norovirus infections in Zhejiang Province from February 12 to 21 in each year from 2013 to 2015.
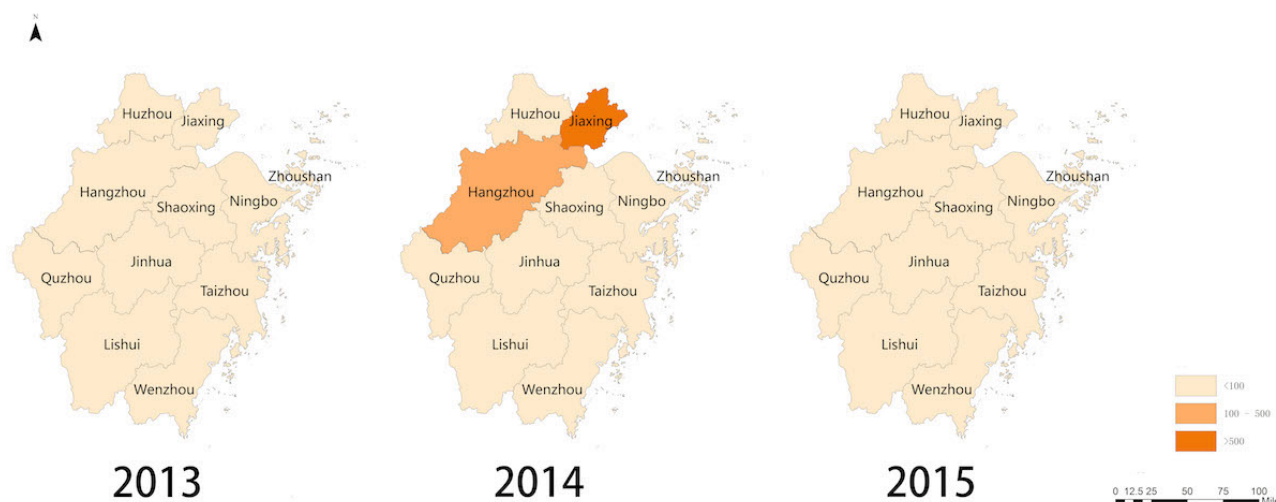
Spatial Presentation for Predicted Norovirus

Based on the preceding optimal model, potential norovirus cases of 11 municipalities in Zhejiang Province during the study time were evaluated and displayed on the map. The number of possible cases in the same period in 2013 and 2015 was predicted (Figure 4). From the displayed map, Jiaxing Municipality in 2014 showed the peak of the norovirus infection than other areas in the same period. Moreover, there might have been potential norovirus epidemics in other municipalities, such as Hangzhou.

Discussion

Principal Findings

Studies demonstrated that norovirus, a common pathogen of acute gastroenteritis, caused several serious outbreaks in China especially in Zhejiang Province in the last decade, implying imminent demand for effective control and prevention of norovirus epidemics [31-34]. With an independent reporting module yet to be constructed in the Chinese National Notifiable Infectious Disease Reporting System, only norovirus outbreaks were recorded by the Public Health Emergency Management Information System. To some extent, such a circumstance limited the detection of norovirus epidemics at the early stage. Fortunately, Internet-based surveillance offers a potential means for monitoring emergent infectious diseases, whose effectiveness and dependability have been explored and examined in some studies [22,27,35].

In this study, we used Internet-based surveillance to identify the association between possible norovirus case number and the fluctuant retrieval index of norovirus-related keywords from the Baidu search engine, and explored an optimal model with specific time lag for the prediction of norovirus epidemics. Possible norovirus-related keywords were first captured at the Keywords Mining website, which involved the technology of text mining and semantic analysis. Then, the Spearman rank correlations between possible norovirus case numbers and the search index of norovirus-related keywords were calculated to determine the inclusion keywords with different time lags. After that, all inclusion keywords for each time lag were combined to obtain the composite Baidu Index and calculate its related Spearman rank correlation coefficients. In this research, the results suggested that the composite Baidu Index of five included keywords was significantly related to this outbreak at the time lag of 2 days with the largest Spearman rank correlation coefficient of $\rho=.945$ followed by two included keywords at the time lag of 1 day with Spearman rank correlation coefficient of $\rho=.924$. Combined with the OCR values of different models, ECM was shown to be optimal including two included keywords with a time lag of 1 day. These findings are similar to previous study implying Internet-based surveillance based on some specific keywords might be effective in identifying epidemics [22]. In contrast to a previous study, the optimal mean composite Baidu Index related to norovirus in this study was essentially on a parallel track with the new cases reported in Figure 3 [22]. Given the short disease course, the serious symptom of acute gastroenteritis, and the characteristics of self-healing, the clustered epidemics of norovirus infection in young groups were more likely to be monitored, particularly in schools. Thus, the early warning of norovirus epidemics by Internet surveillance might be limited for epidemics with short incubation periods and rapid disease progression. Moreover, according to ECM with a time lag of 1 day, the optimal model indicated that an increase of one unit in the composite Baidu Index from 1/100 million netizens contributed to the rise of norovirus infections by 2.15 times during the outbreak, which further supports the

quantitative relationship between Internet surveillance and potential norovirus cases.

Previous studies have identified different optimal time lags in the analysis of diverse diseases, suggesting that this might be attributed to diverse study purposes, various incubation periods, and population susceptibility of different diseases [21,22,36]. Also, the ultimate time lag (1 day) selected in this paper was not the absolutely optimal time lag (2 days) in our study, implying more in-depth studies should be performed at more exquisite scope to explore its potential mechanism, which could mine significantly targeted interventions in public health.

Geographic information system technology has been adopted to present predicted norovirus cases in scale of Zhejiang Province (Figure 4). Compared with the same study period in each year from 2013 to 2015, our predicted results also demonstrated that the selected norovirus outbreak that occurred in Jiaxing Municipality was the largest one, which may prove the reliability of our prediction by Internet surveillance to some extent. Interestingly, some potential cases were identified in Hangzhou Municipality. Although no direct evidence was provided, Hangzhou was shown to be a high-risk region of norovirus infection in the available literature, which also certified the efficiency of Internet surveillance [37-39].

Limitations

Some limitations should be mentioned in our study:

1. The representation of the study was limited because the norovirus outbreak in question involved only schools. Therefore, the conclusion extrapolated to the whole population was insufficient.

2. Although the clinical and epidemiological evidence could be obtained in this study, laboratory tests were not performed for all cases. Thus, the accuracy of the prediction might be affected.

3. The R^2 of ECM was .882, whereas the rest (<12%) were not explored. Other external factors, such as environmental factors and economic factors, might influence the eventual results, which were not considered in this study.

4. Despite some technical means employed to search as many related keywords as possible, omission of subordinate keywords might be inevitable.

5. Other models that were not explored in this study might have a better goodness of fit, which could influence the accuracy of this study.

Conclusions

Over the past decades, the development of the Internet and search engines in China have experienced rapid leaps. A majority of the public sought medical information and expressed personal concerns on the Internet, which provided the underlying possibility for disease surveillance through the Internet, particularly in the field of emergent infectious diseases. The role of forecasting and warning against infectious diseases through the Internet has been identified in some available studies, whereas there is still no record reporting acute infectious diseases such as norovirus that have a short incubation period. In this study, we try to explore the significant keywords involving norovirus, construct an effective model, and eventually identify the potential epidemics of norovirus in Zhejiang Province using Internet surveillance. Despite existing limitations in early warning and unavoidable biases, Internet surveillance may be still useful for the monitoring of norovirus epidemics when a monitoring system is unavailable.

Authors' Contributions

KL designed the study and drafted the manuscript. SCH screened, analyzed the data, and drafted the manuscript. ZPM revised the manuscript and drafted the manuscript. KL, SCH, AND ZPM contributed equally to the manuscript. CC and JJ also made equal contributions to the manuscript. BC, TJ, GFC, and ZGJ interpreted and revised the manuscript. YDC and ZTW collected and screened the data. HG screened the data and revised the manuscript. CLC and JMJ designed and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The Internet users of different municipalities in Zhejiang Province by the end of 2013.

[PDF File (Adobe PDF File), 96KB-Multimedia Appendix 1]

References

XSL•FO

RenderX

1. Florez ID, Al-Khalifah R, Sierra JM, Granados CM, Yepes-Nuñez JJ, Cuello-Garcia C, et al. The effectiveness and safety of treatments used for acute diarrhea and acute gastroenteritis in children: protocol for a systematic review and network meta-analysis. Syst Rev 2016 Jan 20;5:14 [FREE Full text] [doi: 10.1186/s13643-016-0186-8] [Medline: 26818403]

2. Pieścik-Lech M, Shamir R, Guarino A, Szajewska H. Review article: the management of acute gastroenteritis in children. Aliment Pharmacol Ther 2013 Feb;37(3):289-303 [FREE Full text] [doi: 10.1111/apt.12163] [Medline: 23190209]

3. Belliot G, Lopman BA, Ambert-Balay K, Pothier P. The burden of norovirus gastroenteritis: an important foodborne and healthcare-related infection. Clin Microbiol Infect 2014 Aug;20(8):724-730 [FREE Full text] [doi: 10.1111/1469-0691.12722] [Medline: 24943671]

4. Robilotti E, Deresinski S, Pinsky BA. Norovirus. Clin Microbiol Rev 2015 Jan;28(1):134-164 [FREE Full text] [doi: 10.1128/CMR.00075-14] [Medline: 25567225]

5. Hall AJ, Wikswo ME, Manikonda K, Roberts VA, Yoder JS, Gould LH. Acute gastroenteritis surveillance through the National Outbreak Reporting System, United States. Emerg Infect Dis 2013 Aug;19(8):1305-1309 [FREE Full text] [doi: 10.3201/eid1908.130482] [Medline: 23876187]

6. Division of Viral Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention. Updated norovirus outbreak management and disease prevention guidelines. MMWR Recomm Rep 2011 Mar 04;60(RR-3):1-18 [FREE Full text] [Medline: 21368741]

7. Mattner F, Sohr D, Heim A, Gastmeier P, Vennema H, Koopmans M. Risk groups for clinical complications of norovirus infections: an outbreak investigation. Clin Microbiol Infect 2006 Jan;12(1):69-74 [FREE Full text] [doi: 10.1111/j.1469-0691.2005.01299.x] [Medline: 16460549]

8. Lopman BA, Steele D, Kirkwood CD, Parashar UD. The vast and varied global burden of norovirus: prospects for prevention and control. PLoS Med 2016 Apr;13(4):e1001999 [FREE Full text] [doi: 10.1371/journal.pmed.1001999] [Medline: 27115709]

9. Mulatu G, Zeynudin A, Zemene E, Debalke S, Beyene G. Intestinal parasitic infections among children under five years of age presenting with diarrhoeal diseases to two public health facilities in Hawassa, South Ethiopia. Infect Dis Poverty 2015 Nov 04;4:49 [FREE Full text] [doi: 10.1186/s40249-015-0081-x] [Medline: 26530964]

10. Petrignani M, van Beek J, Borsboom G, Richardus J, Koopmans M. Norovirus introduction routes into nursing homes and risk factors for spread: a systematic review and meta-analysis of observational studies. J Hosp Infect 2015 Mar;89(3):163-178. [doi: 10.1016/j.jhin.2014.11.015] [Medline: 25601744]

11. Glass RI, Parashar UD, Estes MK. Norovirus gastroenteritis. N Engl J Med 2009 Oct 29;361(18):1776-1785. [doi: 10.1056/NEJMra0804575] [Medline: 19864676]

12. Zhou N, Zhang H, Lin X, Hou P, Wang S, Tao Z, et al. A waterborne norovirus gastroenteritis outbreak in a school, eastern China. Epidemiol Infect 2016 Apr;144(6):1212-1219. [doi: 10.1017/S0950268815002526] [Medline: 26482884]

13. Karst SM, Baric RS. What is the reservoir of emergent human norovirus strains? J Virol 2015 Jun;89(11):5756-5759 [FREE Full text] [doi: 10.1128/JVI.03063-14] [Medline: 25787285]

14. Sai L, Sun J, Shao L, Chen S, Liu H, Ma L. Epidemiology and clinical features of rotavirus and norovirus infection among children in Ji'nan, China. Virol J 2013 Oct 08;10:302 [FREE Full text] [doi: 10.1186/1743-422X-10-302] [Medline: 24099150]

15. Xue Y, Pan H, Hu J, Wu H, Li J, Xiao W, et al. Epidemiology of norovirus infections among diarrhea outpatients in a diarrhea surveillance system in Shanghai, China: a cross-sectional study. BMC Infect Dis 2015 Apr 15;15:183 [FREE Full text] [doi: 10.1186/s12879-015-0922-z] [Medline: 25884557]

16. Zhang S, Zhou Y, Xu W, Tian L, Chen J, Chen S, et al. Impact of co-infections with enteric pathogens on children suffering from acute diarrhea in southwest China. Infect Dis Poverty 2016 Jun 27;5(1):64 [FREE Full text] [doi: 10.1186/s40249-016-0157-2] [Medline: 27349521]

17. Wang X, Yong W, Shi L, Qiao M, He M, Zhang H, et al. An outbreak of multiple norovirus strains on a cruise ship in China, 2014. J Appl Microbiol 2016 Jan;120(1):226-233. [doi: 10.1111/jam.12978] [Medline: 26481457]

18. Zhang X, Huang Q, Long Y, Jiang X, Zhang T, Tan M, et al. An outbreak caused by GII.17 norovirus with a wide spectrum of HBGA-associated susceptibility. Sci Rep 2015 Dec 07;5:17687 [FREE Full text] [doi: 10.1038/srep17687] [Medline: 26639056]

19. Chan MC, Lee N, Hung T, Kwok K, Cheung K, Tin EK, et al. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. Nat Commun 2015 Dec 02;6:10061 [FREE Full text] [doi: 10.1038/ncomms10061] [Medline: 26625712]

20. Tao Z, Xu M, Lin X, Wang H, Song L, Wang S, et al. Environmental surveillance of genogroup I and II noroviruses in Shandong Province, China in 2013. Sci Rep 2015 Nov 30;5:17444 [FREE Full text] [doi: 10.1038/srep17444] [Medline: 26616566]

21. Kang M, Zhong H, He J, Rutherford S, Yang F. Using Google Trends for influenza surveillance in South China. PLoS One 2013;8(1):e55205 [FREE Full text] [doi: 10.1371/journal.pone.0055205] [Medline: 23372837]

22. Gu Y, Chen F, Liu T, Lv X, Shao Z, Lin H, et al. Early detection of an epidemic erythromelalgia outbreak using Baidu search data. Sci Rep 2015 Jul 28;5:12649 [FREE Full text] [doi: 10.1038/srep12649] [Medline: 26218589]

23. China Internet Network Information Center. 2015 Jul. The 36th statistical report on Internet development in China URL: http://www.cnnic.cn/gywm/xwzx/rdxw/2015/201507/t20150723_52626.htm [accessed 2017-07-23] [WebCite Cache ID 6sAfX3iw0]

24. China Internet Network Information Center. 2014. 2014 Chinese Internet users search behavior study URL: http://www.cnnic.net.cn/hlwfzyj/hlwmrtj/201410/t20141017_49359.htm [accessed 2017-07-23] [WebCite Cache ID 6sAfikIYf]

25. Chen T, Lin J. [Comparative analysis of temporal-spatial evolution of online public opinion based on search engine attention-cases of Google Trends and Baidu Index]. J Intell 2013;3(32):7-16.

26. Liu B, Wang Z, Qi X, Zhang X, Chen H. Assessing cyber-user awareness of an emerging infectious disease: evidence from human infections with avian influenza A H7N9 in Zhejiang, China. Int J Infect Dis 2015 Nov;40:34-36 [FREE Full text] [doi: 10.1016/j.ijid.2015.09.017] [Medline: 26432409]

27. Xie T, Yang Z, Yang S, Wu N, Li L. Correlation between reported human infection with avian influenza A H7N9 virus and cyber user awareness: what can we learn from digital epidemiology? Int J Infect Dis 2014 May;22:1-3 [FREE Full text] [doi: 10.1016/j.ijid.2013.11.013] [Medline: 24576842]

28. Althouse BM, Ng YY, Cummings DAT. Prediction of dengue incidence using search query surveillance. PLoS Negl Trop Dis 2011 Aug;5(8):e1258 [FREE Full text] [doi: 10.1371/journal.pntd.0001258] [Medline: 21829744]

29. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. PLoS One 2013;8(5):e64323 [FREE Full text] [doi: 10.1371/journal.pone.0064323] [Medline: 23750192]

30. Shang X, Fu X, Zhang P, Sheng M, Song J, He F, et al. An outbreak of norovirus-associated acute gastroenteritis associated with contaminated barrelled water in many schools in Zhejiang, China. PLoS One 2017;12(2):e0171307 [FREE Full text] [doi: 10.1371/journal.pone.0171307] [Medline: 28170414]

31. Xue L, Wu Q, Cai W, Zhang J, Guo W. Molecular characterization of new emerging GII.17 norovirus strains from South China. Infect Genet Evol 2016 Jun;40:1-7. [doi: 10.1016/j.meegid.2016.02.026] [Medline: 26923075]

32. Xue L, Dong R, Wu Q, Li Y, Cai W, Kou X, et al. Molecular epidemiology of noroviruses associated with sporadic gastroenteritis in Guangzhou, China, 2013-2015. Arch Virol 2016 May;161(5):1377-1384. [doi: 10.1007/s00705-016-2784-0] [Medline: 26906693]

33. Chen H, Qian F, Xu J, Chan M, Shen Z, Zai S, et al. A novel norovirus GII.17 lineage contributed to adult gastroenteritis in Shanghai, China, during the winter of 2014–2015. Emerg Microbes Infect 2015 Nov;4(11):e67 [FREE Full text] [Medline: 26975060]

34. Han J, Ji L, Shen Y, Wu X, Xu D, Chen L. Emergence and predominance of norovirus GII.17 in Huzhou, China, 2014-2015. Virol J 2015 Sep 11;12:139 [FREE Full text] [doi: 10.1186/s12985-015-0370-9] [Medline: 26362650]

35. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of Internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]

36. Cho S, Sohn CH, Jo MW, Shin S, Lee JH, Ryoo SM, et al. Correlation between national influenza surveillance data and google trends in South Korea. PLoS One 2013;8(12):e81422 [FREE Full text] [doi: 10.1371/journal.pone.0081422] [Medline: 24339927]

37. Song CL, Liu Y. [Molecular epidemiology of norovirus in clustering of gastroenteritis in China]. Pract Prevent Med 2013 Dec;20:1294-1296.

38. XU XQ, Wang Z, Zhen YM, Liu BY. [Epidemiology of norovirus gastroenteritis outbreaks in Zhejiang Province]. Disease Surveillance 2011 Sep;26(9):710-713.

39. Lu QB, Lin JF, Li FD, Wu HC, Wu C, Wang Z, et al. [Epidemiology characteristics of school norovirus gastroenteritis outbreaks in Zhejiang Province]. Chin J Sch Health 2015;36(10):1537-1539.

Abbreviations

CCRM: cubic curve regression model
ECM: exponential curve model
GCM: growth curve model
LRM: linear regression model
QCRM: quadratic curve regression model
OCR: overall coincidence rate

XSL•FO
RenderX