Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Hanako Osuka Date

Temporal Relationship between Healthcare-Associated and Non-Healthcare-Associated - Norovirus Outbreaks, and Google Trends Data in the United States

By

Hanako Osuka
Degree to be awarded: Master of Public Health

Epidemiology

____
Benjamin Lopman, PhD, MSc
Committee Chair

____
Aron Hall, DVM, MSPH
Committee Member

Temporal Relationship between Healthcare-Associated and Non-Healthcare-Associated - Norovirus Outbreaks, and Google Trends Data in the United States

By

Hanako Osuka

M.D., University of Tsukuba, 2006
M.P.H., Emory University, 2017

Thesis Committee Chair: Benjamin Lopman, PhD, MSc

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health in Epidemiology
2017

Abstract

Temporal Relationship between Healthcare-Associated and Non-Healthcare-Associated - Norovirus Outbreaks, and Google Trends Data in the United States

By Hanako Osuka

Background

Norovirus is the leading cause of acute gastroenteritis in the United States. Because it is highly contagious, controlling norovirus outbreaks in healthcare settings is challenging. Since seasonal norovirus activity varies from year to year and rapid implementation of contact precautions is essential, early detection and prediction of the norovirus season may be useful for infection control in healthcare settings. Digital data, such as Google Trends and Twitter, have been increasingly used as a data source to examine infectious disease dynamics. We examined temporal relationships of norovirus outbreaks among healthcare settings, non-healthcare settings, and Google Trends search activity.

Methods

We analyzed norovirus outbreaks from 2009 - 2015 obtained from the National Outbreak Reporting System (NORS) database and Google Trends data in the same period in the United States. We categorized outbreaks into healthcare-associated and non-healthcare-associated, then examined temporal relationships between healthcare-associated outbreaks with (a) non-healthcare-associated norovirus outbreaks and (b) Google Trends data. We identified and compared the onset, peak, and end of the season and conducted linear regression analysis with a series of lags.

Results

11,212 confirmed and suspected norovirus outbreaks involving a total of 397,148 primary cases were reported to NORS during 2009-2015. Healthcare-associated outbreaks had more pronounced seasonality than non-healthcare outbreaks, as they had a higher peak-mean ratio (5.5 v.s. 3.3) and were more concentrated in winter; 63.6% v.s. 44.6% of total outbreaks occurred during November to February. There was weak correlations between weekly counts of healthcare-associated outbreaks with (a) non-healthcare-associated outbreaks ($R^2 = 0.39$) and (b) moderate correlation with Google Trends activity ($R^2 = 0.68$) overall. During the increasing phase of the season, healthcare-associated and non-healthcare-associated outbreaks with a seven-week lead showed the highest correlation ($R^2 = 0.43$). The strongest correlation was observed with no time lag between Google Trends activity and healthcare-associated outbreaks during the increasing phase of the season ($R^2 = 0.68$).

Conclusions

Non-healthcare-associated norovirus outbreaks are less seasonal but increased earlier than healthcare-associated outbreaks. Google Trends data showed moderate correlation with healthcare-associated outbreaks, but without preceding lag. Monitoring community norovirus activity and Google Trends data may have a potential to supplement existing norovirus surveillance and provide early warning of the season.

Temporal Relationship between Healthcare-Associated and Non-Healthcare-Associated - Norovirus Outbreaks, and Google Trends Data in the United States

By

Hanako Osuka

M.D., University of Tsukuba, 2006
M.P.H., Emory University, 2017

Thesis Committee Chair: Benjamin Lopman, PhD, MSc

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health in Epidemiology
2017

Acknowledgments

I would first like to express my sincere gratitude to my advisor, Ben Lopman, for his continuous guidance, input, and encouragement. He consistently allowed this paper to be my own work, but steered me in the right the direction whenever he thought I needed it. I would also like to thank Aron Hall for his support and feedback.

Finally, a very special thank you to my husband, Satoru, for his support and encouragement over the past few years to complete the program. Without his support, I would not have had the opportunity to obtain this degree.

Hanako Osuka

Table of Contents

Background

Norovirus is a single-stranded RNA virus within the *Caliciviridae* family. It is one of the most common causes of acute gastroenteritis and an estimated 19 - 21 million cases of norovirus infections occur in the United States annually (1). The virus mainly transmits person-to-person through the fecal-oral route, or can be transmitted through contaminated food, water, and the environment (2).

Norovirus often causes outbreaks in various settings, the most in the United States being long-term care facilities (LTCF) (3). Norovirus outbreaks in health care facilities cause substantial expenses of both material and human resources, leading to financial burden and disruption of healthcare services as well as morbidity and potential mortality (4). One report estimated the costs associated with a large outbreak in a tertiary care hospital in the United States to be $657,644 (4); nosocomial acute gastroenteritis outbreaks have been estimated to cost about $184 million a year in the United Kingdom (5). Although infected individuals with no underlying conditions usually recover in a few days without serious complications, norovirus outbreaks are associated with increased mortality and all-cause hospitalization in nursing homes (6). Because the virus is highly contagious, control in healthcare settings is challenging. Currently there is no vaccine or specific treatment and management strategies for norovirus focus on general infection control practices and supportive care. Hand hygiene, environmental disinfection, and initiation of control policy are the key strategies of norovirus infection control (7). Because immediate implementation of contact precautions is especially important for controlling norovirus (8), early detection and prediction of the norovirus season may be useful.

Most norovirus outbreaks occur during winter months in the northern hemisphere, but seasonal activity varies from year to year (9). This seasonal variation may be associated with the emergence of specific viral strains, the environmental factors and host-behavioral changes (10), and may be affected by climate change (11). Since norovirus is highly contagious and has a short incubation period(2), the disease activity can change rapidly. Therefore, timely surveillance is needed. In the United States, surveillance networks have been developed by the Centers for Disease Control and Prevention (CDC) for monitoring the activity and public health impacts of norovirus. Integrated surveillance provides nationwide epidemiologic data of outbreaks and genetic sequence of norovirus (12). However, these surveillance systems often have reporting lags which can limit the ability to rapidly and fully assess norovirus seasonal activity. Recently, the CDC launched Norovirus Sentinel Testing and Tracking (NoroSTAT), near-real time reporting system, to reduce reporting lags (12). Surrogate measures, such as syndromic surveillance, have also shown to supplement existing surveillance systems and to predict norovirus activity (13, 14).

Digital data platforms such as Google Trends and Twitter, have been increasingly used as a data source to examine infectious disease dynamics. Salathe proposed four uses of digital data (15): (1) early detection of disease outbreaks, (2) continuous monitoring of disease levels, (3) to assess disease-relevant health-related behaviors, and (4) to provide an additional method for examining the period before outbreaks. For example, Google Flu Trends has been used to estimate influenza activity for 29 countries and has provided an early warning of increase in influenza-like illness (16). The user distribution varies in platforms and users are not the same as the target population, but digital data can provide local and timely information and can be highly accessible. Previously, internet search terms using Google Internet query (17) and Websök, a Swedish search engine, have been demonstrated as possible early warnings of norovirus season (18). These studies, however, did not specifically examine norovirus outbreaks in healthcare settings.

Most healthcare-associated norovirus outbreaks occur in LTCF, where relatively small populations congregate and may have higher levels of contact and compromised hygiene. Healthcare workers have high proportion of asymptomatic infections but their role in transmission is not clear (19). They may introduce the virus into LTCF and cause worker-to-patient transmission, or visitor and newly admitted patients may carry the virus. Because the individuals in LTCF have less chance to expose to the virus than in communities, we suspect that norovirus infections increase in communities first then circulating viruses are subsequently transmitted by workers or visitors in LTCF. Therefore, we hypothesize that the non-healthcare norovirus outbreaks precede the healthcare-associated outbreaks.

The aim of this study is to describe characteristics of outbreaks in the healthcare and the non-healthcare settings and to quantify temporal relationships of norovirus outbreaks among the healthcare settings, the non-healthcare settings, and Google Trends search activity. We are especially interested in determining whether there is a time lag between outbreaks in healthcare settings and non-healthcare settings because we hypothesize that the non-healthcare norovirus outbreaks precede the healthcare-associated outbreaks and this would be a good early warning signal. We also examine whether Google Trends activity can be a predictive measure of healthcare-associated norovirus outbreaks, by analyzing for time lags between the number of healthcare-associated outbreaks and Google Trends activity.

Methods

Data Sources

Norovirus outbreak data were obtained from the National Outbreak Reporting System (NORS) database. NORS is a national outbreak surveillance system of acute gastroenteritis outbreaks, launched in February 2009. State, local, and territorial public health agencies can report acute gastroenteritis outbreaks to the CDC using a standard online data entry system. The primary mode of transmission (i.e., person-to-person contact, foodborne transmission, environmental contamination, water-borne, or unknown) as determined by each site is required in order to create a report. Outbreak characteristics, case characteristics, and laboratory information are reported through the system, but are not mandatory. An outbreak is defined as two or more cases of a similar illness epidemiologically linked to a common exposure. If two or more laboratory-confirmed cases are reported, the outbreak is considered "confirmed". If a reported etiology is associated with less than two laboratory-confirmed cases, or based solely on clinical or epidemiologic criteria, the outbreak is classified as "suspected" (3). We included both suspected and confirmed norovirus outbreaks in this analysis. Available outbreak data with a first illness onset date of January 2009 - December 2015 were obtained. Water-borne outbreak data were only available from 2009 to 2012 (total 13 norovirus outbreaks), therefore we excluded them from the analysis.

Google Trends (GT) is a free online tool that provides data on search patterns for user-specified terms. GT provides geographically-stratified temporal patterns among all search inquiries using Google, the most popular search engine worldwide. It then provides a relative score for the user-specified term, with a given location and time period. Scores represent a relative point to the highest point on a scale from 0 to 100; A score of 100 is the peak popularity of the term over a given time period, a score of 50 means that the term is half as popular in a given time period compared to the peak. We considered the same ten norovirus-related terms examined in a previous study (17): “norovirus”, “vomiting”, “diarrhea”, “nausea”, “abdominal pain”, “stomach virus”, “food poisoning”, “gastroenteritis”, “Norwalk virus”, “rotavirus”. We abstracted GT data from the United States for the same period as the NORS data.

Categorization

For analyzing the NORS data, we categorized the outbreak settings into healthcare-associated and non-healthcare-associated. Regardless of mode of transmission, we categorized them as healthcare-associated if the reported outbreak setting is “Hospital”, “Long-term care/nursing home/assisted living facility”, “Other healthcare facility.” For foodborne outbreaks, the place where food was prepared and the place where food was eaten are reported separately in NORS. We categorized foodborne outbreaks as healthcare-associated when the place where food was eaten was “Hospital”, “Long-term care/nursing home/assisted living facility”, “Other healthcare facility.” “Unknown” and missing settings are categorized as setting unknown and other settings were categorized as non-healthcare-associated outbreaks.

For weekly analysis, we grouped the outbreaks using Morbidity and Mortality Weekly Report (MMWR) week (20) for both NORS and GT data. Weeks last from

Sunday through Saturday. Thus, we exclude the incomplete weeks (2009 week 1 and 2015 week 52) from the weekly analysis.

Analysis

A. Descriptive analysis

First, we conducted the pairwise correlation analysis with the monthly number of outbreaks of each setting to see the occurrence pattern of each setting and confirm the categorization scheme. Then we conducted a descriptive analysis of healthcare-associated and non-healthcare-associated outbreaks. The number of outbreak-associated cases in each sex and age categories was calculated by multiplying the percentages by the total number of primary cases, for those outbreaks that only reported the percentages. Differences in primary mode of transmission, sex, deaths and hospitalization between healthcare-associated outbreaks and non-healthcare-associated were tested by chi-square test. Only outbreaks that reported information on each characteristic (mode of transmission, sex, age, deaths, hospitalization) were included in this analysis. All statistical analysis was performed with SAS version 9.4 (SAS Institute Inc., Cary, North Carolina) and significance was assessed with p-value <0.05 for the analysis.

B. Temporal analysis

We conducted two types of temporal analysis between healthcare-associated and (a) non-healthcare -associated outbreaks and (b) GT activity. For GT activity, we first conducted correlation analysis between all ten terms with the monthly number of healthcare-associated outbreaks, then used the term with the highest correlation coefficient for weekly analysis. The seasonal year was defined as from week 27 in one year to week 26 of the next.

1) Identifying the peak, onset, and the end of the norovirus season: we used the same definition used by Rha et al for the peak, the onset, and the end of the season (13) and compared these indices among healthcare-associated, non-healthcare-associated, and GT. The peak was defined as the week that has the highest number of outbreaks or the highest GT score in the seasonal year. The onset of the season was defined as the week that exceeds 10% of total outbreaks, or score for GT, in the seasonal year had occurred. The end of the season was defined as the week that exceeds 90% of total outbreaks in the seasonal year, or score for GT, had occurred.

2) Linear regression analysis with lag-time: to identify whether a time lag exists between healthcare-associated outbreaks, non-healthcare-associated outbreaks, and GT activity, we used a simple linear regression model (specified below) and estimated the regression coefficient with 95% confidence interval and R^2. The model is described as;

$$Y_w = \alpha + \beta X_{w+t} + \varepsilon$$

where

Y_w is the number of healthcare-associated or non-healthcare -associated outbreaks in a week w

X_{w+t} is the number of non-healthcare -associated outbreaks or GT score in a week $w+t$

t is the week of lag (ranging from: -12, -11, -10…+11, +12)

When t = -8, for example, we regressed the number of healthcare-associated outbreaks on the number of non-healthcare-associated outbreaks 8 weeks prior. Then we identified the lag week which has the strongest correlation.

We first conducted this analysis using all weeks in the study period, then analyzed the period before/after the peak separately (the week from the preceding trough to the peak of the healthcare-associated outbreaks, the week from the peak to the subsequent trough of the healthcare-associated outbreaks) to examine increasing phase and decreasing phase of norovirus activity.

Results

Descriptive analysis

11,212 confirmed and suspected norovirus outbreaks involving a total of 397,148 primary cases were reported during 2009 - 2015. LTCF accounted for nearly half of the total outbreaks, followed by restaurants (Table 1). The settings we categorized as "Healthcare-associated" were mutually moderately correlated (mean $R^2 = 0.53$), compared with other settings (Table 2). Other settings, including prison/jail, had relatively lower correlation coefficient with healthcare settings (mean $R^2 = 0.15$). Therefore, we categorized the rest of the settings as "non-healthcare-associated."

After categorization, there were 5,542 healthcare-associated outbreaks, 3,247 non-healthcare-associated outbreaks and 2,423 outbreaks with unknown settings. Primary mode of transmission differed between healthcare-associated and non-healthcare-associated outbreaks ($p<0.001$); most (94.7%) healthcare-associated outbreaks were spread through person-to-person transmission whereas 38.2% of non-healthcare-associated outbreaks were person-to-person transmission. Information on the number of deaths was available in 9,962 outbreaks and over 86% of outbreaks with deaths were from healthcare-associated outbreaks. Similarly, over 72.9% of outbreaks with hospitalizations occurred in healthcare-associated settings (Table 3). The case-fatality rate was 2.4 per 1,000 in healthcare-associated outbreaks and 0.1 per 1,000 in non-healthcare -associated outbreaks. The case-hospitalization rate was 2.1 per 100 in healthcare-associated and 0.6 per 100 in non-healthcare -associated outbreaks. Age and sex distributions were also significantly different ($p<0.0001$); in healthcare-associated outbreaks, 62.3% of cases were aged 75 years and older and 75% were female. In non-healthcare-associated outbreaks, 3.3% were aged 75 years and older and 52.1% were female.

Temporal analysis:

Among the ten GT search terms, “stomach virus” showed the highest correlation with the monthly number of the healthcare-associated outbreaks ($R^2 = 0.79$, table 4). Therefore, we used this term for further analysis.

Healthcare-associated outbreaks, non-healthcare-associated outbreaks, and GT score were all more frequent during winter, with healthcare-related outbreaks showing the most pronounced seasonal pattern (Figure 1). 63.6% of healthcare-associated outbreaks occurred during November to February, compared with 44.6% of non-healthcare-associated outbreak and 40.9% of the GT score. The number of weekly healthcare-associated outbreaks ranged 0 - 87 (mean 15.9), non-healthcare-associated outbreaks ranged 0 - 29 (mean 8.9), and GT score ranged 9 - 100 (mean 39.0). The peak-mean ratio was also higher in healthcare-associated outbreaks than non-healthcare-associated or GT score (5.5 vs. 3.3, 2.6).

1) Identifying the norovirus season

Healthcare-associated outbreaks had the latest onset in all six seasons (week 45 - 52), while GT had the earliest onset in five of the six seasons (week 32 - 37) (Figure 2). The difference of the onset between non-healthcare-associated and GT was 3.5 weeks on average (range 1-7 weeks), which is smaller than the difference between healthcare-associated and GT (13.2 on average, range 9-18 weeks). Among all six seasonal years, healthcare-associated outbreaks had the earliest end of the season (week 12 - 18). Again, this difference of the end was smaller between non-healthcare-associated and GT than between healthcare-associated and GT (average 1.3, range 1 - 2 vs. average 4.5, range 1 - 8). Healthcare-associated outbreaks and GT had more similar peaks than healthcare-associated and non-healthcare-associated (average difference 2.7, range 0 - 9 vs. average difference 5.2, range 1 - 10 weeks). The duration of the seasons was the shortest for healthcare-associated outbreaks (average 18.3, range 16 - 20 weeks), compared with non-healthcare-associated and GT (average 33.0, range 30 - 36 and average 36.1, range 35 - 39 weeks).

2) Linear regression analysis with lag-time

There were modest correlations between weekly healthcare-associated outbreak counts and non-healthcare-associated outbreaks (R^2 = 0.39) and moderate correlation with GT score (R^2 = 0.68) overall (Table 5). The R^2 was the highest when there was no time lag between non-healthcare and healthcare-associated outbreaks. The R^2 was also the highest when there was no time lag between GT score and healthcare-associated outbreaks. GT score had weaker correlation with non-healthcare-associated outbreaks than with healthcare-associated outbreaks (R^2 = 0.46, at the highest). R^2 was the highest with two weeks later data of GT score than the number of non-healthcare-associated outbreaks, but the correlation coefficient was smaller compared with healthcare-associated outbreak and GT score (0.23 vs, 0.79).

In the increasing phase, the correlation between healthcare-associated and non-healthcare outbreaks was the strongest for a time lag of minus seven weeks ($R^2 = 0.43$, Table 6a) (i.e. the non-healthcare-associated outbreaks were seven weeks' prior to the healthcare outbreaks). In the decreasing phase, the correlation was strongest with no time lag ($R^2 = 0.54$). When comparing the healthcare-associated outbreaks and GT score (Table 6b), the correlation was the strongest with no time lag in the increasing phase and plus one week lag in the decreasing phase ($R^2 = 0.68, 0.72$, respectively).

Discussion

In this study, we examined temporal relationships between healthcare-associated and non-healthcare-associated norovirus outbreaks and found that non-healthcare-associated norovirus outbreak activity increased earlier than healthcare-associated outbreaks. We also found that the seasonality of healthcare-associated norovirus outbreaks is more distinctive than non-healthcare-associated outbreaks. Internet searches using Google Trends showed stronger correlations with healthcare-associated outbreaks than with non-healthcare-associated outbreaks. No time lag was observed between Google Trends activity and healthcare-associated outbreaks.

We conducted two types of analysis for examining temporal relationships and both of them are suggestive that non-healthcare norovirus outbreaks increase earlier than healthcare-associated outbreaks. Due to the definition of onset and end of the season in this study, duration of the season might have been longer than the reality; since it included 80% of a total number of outbreaks in a seasonal year, the season of non-healthcare norovirus outbreaks resulted in more than seven months on average with these definitions. However, healthcare-associated outbreaks also have a higher peak-mean ratio and more concentrated in winter, compared with non-healthcare-associated outbreaks. Therefore, we conclude that the seasonality of healthcare-associated norovirus outbreaks is more distinctive than non-healthcare outbreaks. The correlation analysis was conducted with three data sets: whole seasonal years, before the peak, and after the peak. When analyzing healthcare-associated and non-healthcare-associated norovirus outbreaks with whole seasonal years, there was no time lag. But when analyzing increasing phase only (from trough to the peak), healthcare-associated outbreaks were the most well correlated with the number of non-healthcare outbreaks seven weeks prior. Because healthcare-associated outbreaks tend to increase and decrease earlier than non-healthcare-associated outbreaks, a regression analysis using whole data might have canceled out these time lags in two phases. These results may support our hypothesis that the circulating viruses in the non-healthcare settings are subsequently transmitted by visitors or healthcare workers therefore non-healthcare-associated norovirus activity increases preceding to healthcare-associated outbreaks. Additionally, since non-healthcare outbreaks occur throughout the year, non-healthcare settings may be acting as a reservoir for virus introduction into healthcare settings. However, because we did not assess the actual transmission of the virus between healthcare settings and communities, there may be unique drivers of the seasonality in healthcare settings. During the winter season, there are increased incidence of diseases, such as cardiovascular diseases and respiratory infections, and increased hospitalizations in healthcare facilities (21, 22). Crowded wards in facilities due to other diseases in the winter may affect the seasonality of norovirus outbreaks. Additionally, different distribution of strains may contributes this seasonality, as one systematic review confirms that healthcare-related and winter outbreaks were associated with GII strains while waterborne outbreaks were associated with GI strains (23).

The previous norovirus studies using digital data showed preceding time lag between internet search activity and norovirus disease activity. In Desai's study in Boston, the correlation was strongest ($R^2 = 0.74$) between emergency department visits and GT data 2 weeks prior (17). Edelstein's study compared Swedish search engine activity and laboratory surveillance and showed that it detected the onset two to three weeks earlier (18). Since we used data from the entire United States, the different local disease activity and internet search activity in local areas may have affected our results and diluted any relationship. Similar to the previous studies, our descriptive analysis showed that healthcare-associated norovirus outbreaks involve more female and elderly people (24, 25). Healthcare-associated outbreaks have more hospitalizations and deaths than non-healthcare outbreaks; the case-fatality rate (0.2 per 100 cases) and case-hospitalization rate (2.1 per 100 cases), in our analysis were similar to other estimates in the literature (26). Since our analysis revealed that healthcare-associated outbreaks have shorter seasonal durations, these increased hospitalizations and deaths also occur over shorter periods. These findings highlight the value of early warnings for seasonal increases in norovirus activity and corresponding reinforcement of infection control practices in healthcare settings.

Our study has important strengths. The first is the large numbers of observations; we analyzed 11,212 outbreaks reported to NORS over six seasonal years. This is the one of the largest national datasets of norovirus outbreaks (3, 27). Another strength is that we conducted different types of analysis to assess temporal relationships. In addition to identifying and comparing the characteristics of season by point (onset, peak, end), we assessed the temporal relationship by regression analysis with whole outbreaks data, increasing phase, and decreasing phase. This multi-faceted approach helped to identify more subtle temporal associations, particularly those within specific portions of a given season.

NORS and Google Trends data each has their own limitations. First, there is variable levels of detail available on outbreaks reported through NORS. Although we categorized and focused on outbreak settings, 22% of outbreak settings were unknown. However, we still have a large sample size and analyzed over 8,800 outbreaks with settings. Second, since NORS only captures outbreaks, it may not represent the overall activity of norovirus. Third, we categorized the outbreak setting as either healthcare-associated or non-healthcare but non-healthcare involves a heterogeneous mix of various settings. Some of the settings, such as jails and prisons, may have some similar characteristics with healthcare-associated. However, our primary focus in this study was on healthcare settings and the number of other institutional settings included in the non-healthcare category was relatively small. Fourth, there may be reporting biases among outbreaks in different setting, and relatedly, there is a possibility that healthcare facilities are more likely to test norovirus during the winter, once the outbreak season is recognized.

Because we do not know user's search motives, Google Trends can be affected by various factors, such as media coverage and another disease's activity. In this study, for example, we observed a sudden increase in GT activity in 2013 (Figure 1) but we cannot tell the reason of this increase. A single news story by a major media outlet, for example, could lead to increased GT activity. Also, increasing activity of other etiologies that cause similar symptoms, such as rotavirus infections, may drive the GT activities because we used the term "stomach virus" for weekly analysis, which is not specific to norovirus. Furthermore, even though GT allows us to obtain the search activity data only in the United States, we do not know users' demographics. Google users might not have represented properly the population in the United States. Additionally, GT only counts search activity, not the number of users. In other words, we cannot distinguish one search activity of ten individuals from ten search activities of one individual. Therefore, GT should be interpreted with caution; we cannot assess magnitude or severity of the disease with GT because of these limitations, although it is useful to see temporal trends.

Our study suggests two things: that there is a relationship of norovirus outbreak between non-healthcare and healthcare settings, and the possible use of digital data for norovirus surveillance. This study implicates that the activity of norovirus increase in the non-healthcare first, then later in healthcare settings. This study did not assess the specific genotype so it did not track the exact dynamics of the virus. Further studies using more detailed molecular epidemiology, is needed to assess these temporal relationships. If we track specific genotype with molecular methods, the virus activity between non-healthcare and healthcare settings may become clear. Applying our study into a mathematical model is another possibility for future study. One study has examined the effectiveness of ward closure of nosocomial norovirus outbreak by mathematical model (28), but it did not take into account of seasonality or the transmission links between healthcare and community. Incorporating our results in network model may improve the explanatory power in the results in similar studies.

Since this study showed non-healthcare norovirus outbreaks precede healthcare-associated outbreaks by several weeks, it might be useful to predict the onset of norovirus season for healthcare settings by monitoring community outbreaks. Although there was no preceding time lag in increasing phase, Google Trend showed a strong correlation with healthcare-associated norovirus outbreaks. Because it is real-time data and highly accessible to the public, this may help augment current national surveillance systems. It may provide timely data to state and local health departments to enable early detection, response, and reporting of outbreaks. By monitoring Google Trend activity, healthcare workers may have a timely signal of the norovirus season onset and improve the speed of infection control response.

In conclusion, non-healthcare norovirus outbreaks and GT data may provide a good early warning signal of healthcare outbreak because of their time lead and data timeliness, respectively. To assess the actual transmission dynamics of norovirus between community and healthcare settings, further study is needed.

Figure 1. Number of Weekly Reported Healthcare-Associated and Non-Healthcare-Associated Norovirus Outbreaks and Google Trends Score with "Stomach Virus" between 2009 and 2015, United States

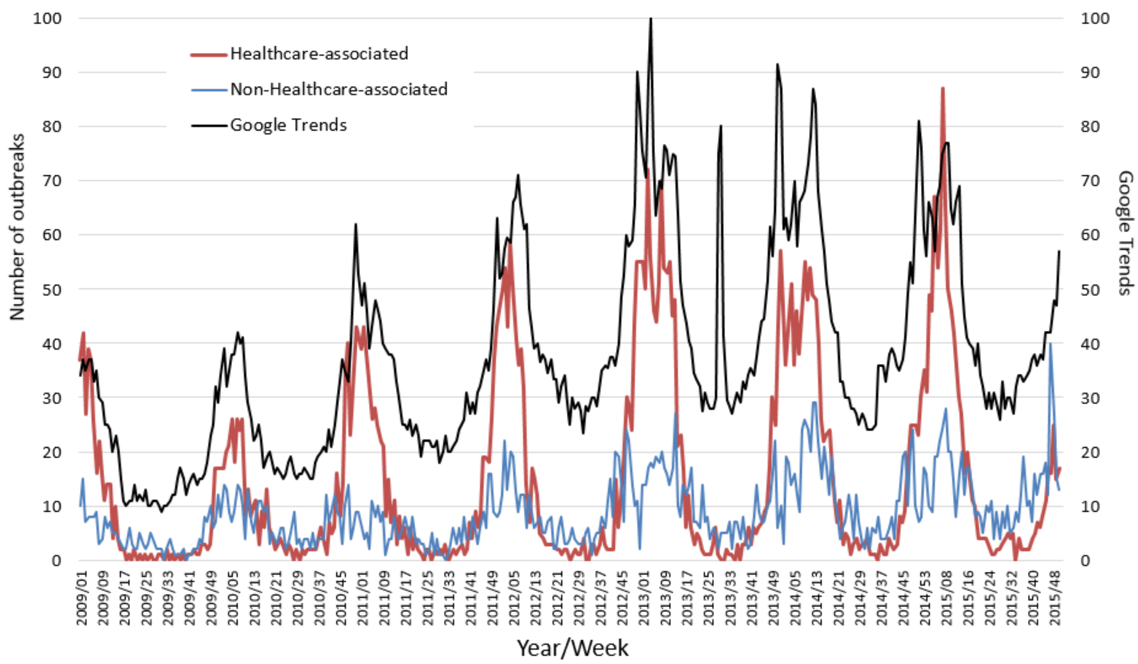
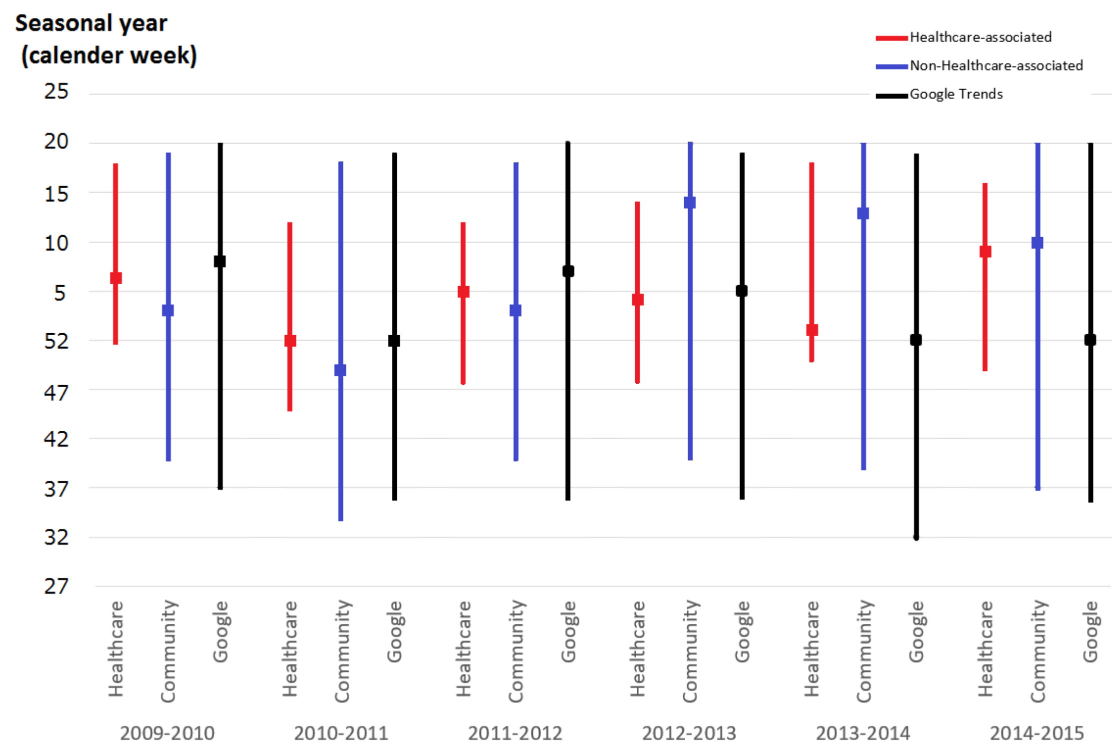


Figure 2. The Week of Onset, Peak, and End of Season for Healthcare-Associated and Non-Healthcare-Associated Norovirus Outbreaks and Google Trends Score with "Stomach Virus" between 2009 and 2015, United States



*The lines show the onset to the end of each season, the boxes represent the peak
**The seasonal year was defined as from week 27 in one year to week 26 of the next

Table 1. Number and Percentage of Reported Norovirus Outbreaks by Setting, National Outbreak Reporting System, 2009-2015, United States

Setting	Frequency	(%)
Healthcare-associated		
Long-term care/nursing home/ assisted living facility	5,165	(46.1)
Hospital	279	(2.5)
Other healthcare facility	98	(0.9)
Non-healthcare-associated		
Restaurant	1,367	(12.2)
School/college/university	617	(5.5)
Event space	292	(2.6)
Private home/residence	223	(2.0)
Child day care	208	(1.9)
Office/indoor workplace	112	(1.0)
Camp	77	(0.7)
Hotel/motel	52	(0.5)
Prison/jail	42	(0.4)
Religious facility	33	(0.3)
Festival/fair	7	(0.1)
Ship/boat	5	(0.1)
Other	212	(1.9)
Unknown	13	(0.1)
Missing	2,410	(21.5)
Total	11,212	(100.0)

Table 2. Pearson Correlation Coefficients Matrix among Outbreaks Settings, National Outbreak Reporting System, 2009-2015, United States

Setting (n)	LTCF**	Hospital	Other healthcare facility	Restaurant	School/ college/ university	Event space	Private home/ Residence	Child day care	Office/ indoor workplace	Camp	Hotel/motel	Prison/jail	Religious facility
LTCF (5,165)**	1.00												
Hospital (279)	0.87 (<0.01)	1.00											
Other healthcare facility (98)	0.71 (<.001)	0.60 (<.001)	1.00										
Restaurant (1,367)	0.77 (<.001)	0.64 (<.001)	0.49 (<.001)	1.00									
School/college/university (617)	0.55 (<.001)	0.48 (<.001)	0.31 (0.003)	0.63 (<0.001)	1.00								
Event space (292)	0.46 (<.001)	0.44 (<.001)	0.23 (0.034)	0.57 (<0.001)	0.45 (<0.001)	1.00							
Private home/residence (223)	0.36 (<.001)	0.34 (0.002)	0.17 (0.117*)	0.46 (<0.001)	0.29 (0.008)	0.27 (0.001)	1.00						
Child day care (208)	0.43 (<.001)	0.35 (0.001)	0.27 (0.015)	0.43 (<0.001)	0.71 (<0.001)	0.36 (0.001)	0.23 (0.032)	1.00					
Office/indoor workplace (112)	0.37 (<.001)	0.42 (<.001)	0.23 (0.039)	0.51 (<0.001)	0.30 (0.005)	0.39 (0.001)	0.41 (<0.001)	0.29 (0.007)	1.00				
Camp (77)	-0.24 (0.026)	-0.17 (0.117*)	-0.18 (0.102*)	-0.12 (0.272*)	-0.26 (0.02)	0.01 (0.971*)	-0.03 (0.804*)	-0.20 (0.065*)	-0.13 (0.234*)	1.00			
Hotel/motel (52)	0.44 (<.001)	0.37 (0.001)	0.35 (0.001)	0.56 (<0.001)	0.50 (<0.001)	0.44 (<0.001)	0.22 (0.047)	0.26 (0.019)	0.21 (0.055*)	0.10 (0.367)	1.00		
Prison/jail (42)	0.48 (<.001)	0.43 (<.001)	0.20 (0.071)	0.52 (<0.001)	0.45 (<0.001)	0.32 (0.003)	0.22 (0.040)	0.27 (0.012)	0.30 (0.006)	0.00 (0.967*)	0.49 (<0.001)	1.00	
Religious facility (33)	0.21 (0.055*)	0.17 (0.116*)	0.15 (0.182*)	0.31 (<0.001)	0.21 (0.050)	0.25 (0.022)	0.32 (0.003)	0.23 (0.040)	0.31 (0.005)	0.05 (0.658*)	0.25 (0.020)	0.28 (0.009)	1.00

*not significant
**LTCF; Long term care/nursing home/ assisted living

Table 3. Characteristics of Outbreak by Each Setting, from the National Outbreak Reporting System (NORS) Database, 2009-2015, the United States

Outbreaks	Healthcare-associated (5,542)	Non-healthcare-associated (3,260)	Setting-unknown (2,410)	Total (11,212)	p-value*5
Primary mode of transmission (n, column %)					<0.0001
Person-to-person	5,248 (94.7)	1,240 (38.2)	1,983 (81.8)	8,471 (75.6)	
Food	37 (0.7)	1,723 (53.1)	122 (5.0)	1,882 (16.8)	
Environment	4 (0.1)	17 (0.5)	18 (0.7)	39 (0.4)	
Other/Unknown	253 (4.6)	267 (8.2)	300 (12.4)	820 (7.31)	
Number of primary cases					
Range	2 - 670	2 - 2,500	2 - 699	2 - 2,500	
Median	30	15	25	25	
Total Primary Cases	207,314	99,091	90,743	397,148	
Age*1 (cases, column %)					<0.0001
under 1	54 (0.1)	382 (0.7)	15 (0.1)	451 (0.3)	
1 - 4	68 (0.1)	2,881 (5.0)	214 (1.7)	3,163 (1.9)	
5 - 9	76 (0.1)	10,618 (18.3)	942 (7.4)	11,636 (7.1)	
10 - 19	960 (1.0)	11,861 (20.5)	1,602 (12.6)	14,423 (8.8)	
20 - 49	15,466 (16.5)	19,178 (33.1)	2,971 (23.3)	37,615 (22.9)	
50 - 74	18,668 (19.9)	11,165 (19.3)	2,458 (19.3)	32,291 (19.6)	
75 +	58,360 (62.3)	1,905 (3.3)	4,541 (35.6)	64,806 (39.4)	
unknown	94,526	29,844	9,807	134,177	
Sex*2 (cases, column %)					<0.0001
Female	104,046 (75.1%)	34,452 (52.1%)	11,135 (63.1%)	149,633 (67.3%)	
Male	34,454 (24.9%)	31,732 (47.9%)	6,520 (36.9%)	72,706 (32.7%)	
Death* 3 (cases, row %)					
Number of outbreaks with deaths	347 (86.8%)	12 (3.0%)	42 (10.5%)	401	<0.0001
Total number of deaths	479 (86.6%)	12 (2.2%)	62 (11.2%)	553	
Case-fatality rate	2.4 / 1,000	0.1 / 1,000	1.4 / 1,000	1.6 / 1,000	
Hospitalization*4 (cases, row %)					
Number of outbreaks with hospitalizations	1,738 (72.9%)	387 (16.2%)	258 (10.8%)	2383	<0.0001
Total number of hospitalizations	4,119 (75.7%)	602 (11.1%)	720 (13.2%)	5441	
Case-hospitalization rate	2.1 / 100	0.6 / 100	1.7 / 100	1.6 / 100	

*1 age information was available in 8,334 outbreaks

*2 sex information was available in 8,472 outbreaks

*3 deaths information available in 9,692 outbreaks

*4 hospitalization information available in 9,753 outbreaks

*5 p-value was obtained by Chi-square test between healthcare-associated and non-healthcare-associated

Table 4. R^2 between Monthly Number of Hospital-Associated Norovirus Outbreak and Monthly Google Trends Activity

Search Term	**Highest R^2**
Norovirus	0.569
Vomiting	0.476
Diarrhea	0.152
Nausea	0.142
Abdominal pain	0.095
Stomach virus	**0.790**
Food poisoning	0.454
Gastroenteritis	0.467
Norwalk virus	0.334
Rotavirus	0.640

Table 5. The Coefficient with 95% Confidence Interval, R-square with the Regression Model with Lag-time

	Healthcare with Non-healthcare			Healthcare with Google Trends			Non-healthcare and Google Trends		
Lag (week)*	**β**	**(95% CI)**	**R^2**	**β**	**(95% CI)**	**R^2**	**β**	**(95% CI)**	**R^2**
-12	0.53	(0.22, 0.84)	0.03	0.20	(0.10, 0.30)	0.04	0.11	(0.07, 0.14)	0.10
-11	0.66	(0.36, 0.97)	0.05	0.26	(0.17, 0.36)	0.08	0.12	(0.08, 0.15)	0.12
-10	0.77	(0.46, 1.07)	0.07	0.34	(0.24, 0.43)	0.13	0.14	(0.10, 0.17)	0.17
-9	0.92	(0.63, 1.22)	0.10	0.41	(0.32, 0.50)	0.19	0.16	(0.13, 0.19)	0.23
-8	1.12	(0.83, 1.41)	0.14	0.46	(0.38, 0.55)	0.24	0.17	(0.14, 0.20)	0.25
-7	1.26	(0.98, 1.54)	0.18	0.51	(0.43, 0.60)	0.30	0.18	(0.15, 0.21)	0.28
-6	1.39	(1.17, 1.66)	0.22	0.57	(0.49, 0.65)	0.37	0.19	(0.16, 0.22)	0.32
-5	1.47	(1.21, 1.74)	0.25	0.63	(0.55, 0.70)	0.44	0.19	(0.17, 0.22)	0.33
-4	1.64	(1.38, 1.90)	0.31	0.68	(0.61, 0.75)	0.51	0.20	(0.17, 0.23)	0.36
-3	1.64	(1.40, 1.89)	0.33	0.70	(0.65, 0.77)	0.55	0.21	(0.18, 0.24)	0.40
-2	1.66	(1.42, 1.90)	0.34	0.74	(0.68, 0.80)	0.60	0.22	(0.19, 0.24)	0.41
-1	1.68	(1.44, 1.92)	0.35	0.77	(0.72, 0.83)	0.66	0.22	(0.19, 0.24)	0.41
0	**1.78**	**(1.55, 2.01)**	**0.39**	**0.79**	**(0.73, 0.84)**	**0.68**	0.22	(0.19, 0.24)	0.42
+1	1.73	(1.50, 1.97)	0.37	0.78	(0.73, 0.84)	0.68	0.22	(0.20, 0.25)	0.44
+2	1.67	(1.43, 1.91)	0.34	0.76	(0.70, 0.82)	0.63	**0.23**	**(0.20, 0.25)**	**0.46**
+3	1.59	(1.35, 1.84)	0.32	0.73	(0.66, 0.79)	0.58	0.22	(0.19, 0.25)	0.44
+4	1.54	(1.29, 1.79)	0.29	0.68	(0.61, 0.75)	0.51	0.20	(0.18, 0.23)	0.40
+5	1.45	(1.20, 1.71)	0.26	0.63	(0.55, 0.70)	0.43	0.19	(0.16, 0.22)	0.36
+6	1.34	(1.08, 1.60)	0.22	0.56	(0.48, 0.64)	0.35	0.18	(0.15, 0.21)	0.31
+7	1.20	(0.93, 1.47)	0.18	0.49	(0.41, 0.58)	0.27	0.17	(0.14, 0.19)	0.27
+8	1.07	(0.80, 1.35)	0.14	0.43	(0.34, 0.55)	0.20	0.15	(0.12, 0.18)	0.22
+9	0.98	(0.70, 1.26)	0.12	0.36	(0.27, 0.45)	0.14	0.13	(0.10, 0.16)	0.16
+10	0.87	(0.59, 1.16)	0.09	0.29	(0.20, 0.39)	0.09	0.11	(0.07, 0.14)	0.11
+11	0.64	(0.35, 0.93)	0.05	0.22	(0.12, 0.32)	0.05	0.09	(0.06, 0.13)	0.09
+12	0.47	(0.17, 0.76)	0.03	0.15	(0.05, 0.25)	0.03	0.08	(0.05, 0.11)	0.07

*Negative lag means that the second data is prior to the first data and positive lag means that the second data is after the first data.

Table 6. The Coefficient with 95% Confidence Interval, R-square with the Regression Model with Lag-time: Before and After the Peak

a. **Healthcare and Non-healthcare-associated norovirus outbreak**

	Before the peak			After the peak		
Lag (week)*	**β**	**(95% CI)**	**R^2**	**β**	**(95% CI)**	**R^2**
-12	2.07	(1.42, 2.71)	0.20	0.18**	(-0.27, 0.63)	0.00
-11	2.31	(1.75, 2.86)	0.29	0.27**	(-0.18, 0.72)	0.01
-10	2.33	(1.78, 2.88)	0.30	0.30**	(-0.15, 0.75)	0.01
-9	2.29	(1.75, 2.82)	0.30	0.50	(0.05, 0.75)	0.02
-8	2.42	(1.96, 2.88)	0.39	0.81	(0.38, 1.24)	0.07
-7	**2.38**	**(1.95, 2.80)**	**0.43**	0.98	(0.98, 1.40)	0.10
-6	2.21	(1.78, 2.64)	0.39	1.14	(1.14, 1.55)	0.13
-5	2.01	(1.57, 2.45)	0.33	1.27	(1.27, 1.66)	0.17
-4	2.06	(1.62, 2.49)	0.35	1.52	(1.52, 1.89)	0.25
-3	1.61	(1.21, 2.01)	0.28	1.77	(1.77, 2.11)	0.34
-2	1.45	(1.07, 1.83)	0.26	1.98	(1.98, 2.30)	0.43
-1	1.34	(0.95, 1.72)	0.22	2.08	(2.08, 2.40)	0.45
0	1.42	(1.05, 1.78)	0.26	**2.28**	**(2.28, 2.57)**	**0.54**
+1	1.38	(1.02, 1.74)	0.26	2.26	(2.26, 2.57)	0.51
+2	1.30	(0.93, 1.66)	0.24	2.33	(2.33, 2.65)	0.51
+3	1.26	(0.89, 1.63)	0.22	2.31	(2.31, 2.64)	0.48
+4	1.22	(0.84, 1.60)	0.20	2.26	(2.26, 2.61)	0.45
+5	1.04	(0.64, 1.44)	0.14	2.28	(2.28, 2.63)	0.46
+6	0.95	(0.54, 1.36)	0.12	2.28	(2.28, 2.64)	0.44
+7	0.85	(0.42, 1.27)	0.09	2.22	(2.22, 2.59)	0.41
+8	0.73	(0.29, 1.17)	0.07	2.09	(2.09, 2.48)	0.36
+9	0.73	(0.30, 1.17)	0.07	2.01	(1.60, 2.42)	0.32
+10	0.80	(0.37, 1.23)	0.08	1.84	(1.40, 2.27)	0.25
+11	0.55	(0.11, 0.99)	0.04	1.39	(0.90, 1.89)	0.13
+12	0.39	(-0.06, 0.83)	0.02	1.08	(0.58, 1.58)	0.08

*Negative lag means that the non-healthcare outbreaks' data is prior to the healthcare-associated outbreaks' data.

**not significant

b. Healthcare-associated norovirus outbreak and Google Trend

	Before the peak			After the peak		
Lag (week)*	**β**	**(95% CI)**	**R^2**	**β**	**(95% CI)**	**R^2**
-12	0.70	(0.45, 0.94)	0.16	0.13	(-0.01, 0.28)	0.02
-11	0.76	(0.52, 0.99)	0.2	0.23	(0.09, 0.38)	0.05
-10	0.82	(0.61, 1.04)	0.26	0.35	(0.21, 0.49)	0.12
-9	0.84	(0.65, 1.03)	0.32	0.45	(0.32, 0.58)	0.2
-8	0.84	(0.67, 1.01)	0.36	0.50	(0.38, 0.63)	0.25
-7	0.85	(0.69, 1.01)	0.40	0.55	(0.43, 0.67)	0.31
-6	0.87	(0.71, 1.02)	0.44	0.61	(0.50, 0.71)	0.38
-5	0.94	(0.80, 1.07)	0.54	0.65	(0.55, 0.75)	0.46
-4	0.97	(0.86, 1.08)	0.63	0.67	(0.58, 0.77)	0.50
-3	0.92	(0.81, 1.03)	0.63	0.69	(0.60, 0.78)	0.53
-2	0.89	(0.79, 1.00)	0.65	0.73	(0.65, 0.82)	0.59
-1	0.86	(0.77, 0.95)	0.68	0.77	(0.69, 0.85)	0.65
0	**0.81**	**(0.72, 0.89)**	**0.68**	0.81	(0.74, 0.89)	0.70
+1	0.76	(0.67, 0.84)	0.65	**0.86**	**(0.79, 0.93)**	**0.72**
+2	0.72	(0.63, 0.81)	0.62	0.88	(0.80, 0.96)	0.70
+3	0.70	(0.61, 0.79)	0.58	0.87	(0.79, 0.96)	0.65
+4	0.67	(0.57, 0.77)	0.53	0.87	(0.77, 0.96)	0.62
+5	0.63	(0.52, 0.74)	0.46	0.85	(0.75, 0.95)	0.56
+6	0.58	(0.47, 0.70)	0.39	0.83	(0.71, 0.94)	0.50
+7	0.52	(0.40, 0.64)	0.31	0.79	(0.66, 0.92)	0.42
+8	0.47	(0.34, 0.60)	0.24	0.75	(0.61, 0.89)	0.35
+9	0.40	(0.26, 0.54)	0.17	0.71	(0.56, 0.86)	0.29
+10	0.32	(0.18, 0.47)	0.11	0.66	(0.49, 0.82)	0.23
+11	0.24	(0.09, 0.39)	0.06	0.60	(0.42, 0.78)	0.18
+12	0.16	(0.01, 0.31)	0.03	0.51	(0.31, 0.70)	0.12

*Negative lag means that the non-healthcare outbreaks' data is prior to the healthcare-associated outbreaks' data.

Reference

1. Hall AJ, Lopman BA, Payne DC, et al. Norovirus disease in the United States. *Emerg Infect Dis*. 2013;19(8):1198–1205.

2. Robilotti E, Deresinski S, Pinsky BA. Norovirus. *Clin Microbiol Rev*. 2015;28(1):134–164.

3. Wikswo ME, Kambhampati A, Shioda K, et al. Outbreaks of Acute Gastroenteritis Transmitted by Person-to-Person Contact, Environmental Contamination, and Unknown Modes of Transmission--United States, 2009-2013. *MMWR Surveill Summ*. 2015;64(12):1–16.

4. Johnston CP, Qiu H, Ticehurst JR, et al. Outbreak management and implications of a nosocomial norovirus outbreak. *Clin Infect Dis*. 2007;45(5):534–540.

5. Lopman BA, Reacher MH, Vipond IB, et al. Epidemiology and cost of nosocomial gastroenteritis, Avon, England, 2002-2003. *Emerg Infect Dis*. 2004;10(10):1827–1834.

6. Trivedi TK, DeSalvo T, Lee L, et al. Hospitalizations and mortality associated with norovirus outbreaks in nursing homes, 2009-2010. *JAMA*. 2012;308(16):1668–1675.

7. MacCannell T, Umscheid CA, Agarwal RK, et al. Guideline for the prevention and control of norovirus gastroenteritis outbreaks in healthcare settings. *Infect Control Hosp Epidemiol*. 2011;32(10):939–969.

8. Friesema IHM, Vennema H, Heijne JCM, et al. Norovirus outbreaks in nursing homes: the evaluation of infection control measures. *Epidemiol Infect*. 2009;137(12):1722–1733.

9. Ahmed SM, Hall AJ, Robinson AE, et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *Lancet Infect Dis*. 2014;14(8):725–730.

10. Dowell SF. Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg Infect Dis*. 2001;7(3):369–374.

11. Rohayem J. Norovirus seasonality and the potential impact of climate change. *Clin Microbiol Infect*. 2009;15(6):524–527.

12. Shah MP, Wikswo ME, Barclay L, et al. Near Real-Time Surveillance of U.S. Norovirus Outbreaks by the Norovirus Sentinel Testing and Tracking Network - United States, August 2009-July 2015. *MMWR Morb Mortal Wkly Rep*. 2017;66(7):185–189.

13. Rha B, Burrer S, Park S, et al. Emergency department visit data for rapid detection and monitoring of norovirus activity, United States. *Emerging Infect Dis*. 2013;19(8):1214–1221.

14. Loveridge P, Cooper D, Elliot AJ, et al. Vomiting calls to NHS Direct provide an early warning of norovirus outbreaks in hospitals. *Journal of Hospital Infection*. 2010;74(4):385–393.

15. Salathé M, Freifeld CC, Mekaru SR, et al. Influenza A (H7N9) and the importance of digital epidemiology. *N Engl J Med*. 2013;369(5):401–404.

16. Martin LJ, Xu B, Yasui Y. Improving Google Flu Trends estimates for the United States through transformation. *PLoS ONE*. 2014;9(12):e109209.

17. Desai R, Hall AJ, Lopman BA, et al. Norovirus disease surveillance using Google Internet query share data. *Clin Infect Dis*. 2012;55(8):e75–8.

18. Edelstein M, Wallensten A, Zetterqvist I, et al. Detecting the Norovirus Season in Sweden Using Search Engine Data – Meeting the Needs of Hospital Infection Control Teams. *PLoS ONE*. 2014;9(6):e100309.

19. Sabrià A, Pintó RM, Bosch A, et al. Norovirus shedding among food and healthcare workers exposed to the virus in outbreak settings. *J Clin Virol*. 2016;82:119–125.

20. CDC. MMWR_Week_Fact_Sheet.

21. Foster ED, Cavanaugh JE, Haynes WG, et al. Acute myocardial infarctions, strokes and influenza: seasonal and pandemic effects. *Epidemiol Infect*. 2013;141(4):735–744.

22. Kumar N, Venkatraman A, Garg N. Seasonality in acute ischemic stroke related hospitalizations and case fatality rate in the United States. *Int J Cardiol*. 2015;195:134–135.

23. Matthews JE, Dickey BW, Miller RD, et al. The epidemiology of published norovirus outbreaks: a review of risk factors associated with attack rate and genogroup. *Epidemiol Infect*. 2012;140(7):1161–1172.

24. Lopman BA, Reacher MH, Vipond IB, et al. Clinical manifestation of norovirus gastroenteritis in health care settings. *Clin Infect Dis*. 2004;39(3):318–324.

25. Costantini VP, Cooper EM, Hardaker HL, et al. Epidemiologic, Virologic, and Host Genetic Factors of Norovirus Outbreaks in Long-term Care Facilities. *Clin Infect Dis*. 2015;62(1):1–10.

26. Rosenthal NA, Lee LE, Vermeulen BAJ, et al. Epidemiological and genetic characteristics of norovirus outbreaks in long-term care facilities, 2003-2006. *Epidemiol Infect*. 2010;139(2):286–294.

27. Vega E, Barclay L, Gregoricus N, et al. Genotypic and Epidemiologic Trends of Norovirus Outbreaks in the United States, 2009 to 2013. *J Clin Microbiol*. 2013;52(1):147–155.

28. Vanderpas J, Louis J, Reynders M, et al. Mathematical model for the control of nosocomial norovirus. 2009;71(3):214–222.