A Systematic Review of the Correlation Between Web-based Query and Outbreak of Emerging Infectious Diseases and Meta-analysis of Influenza-like Illnesses

Minh-Duc Nguyen-Tran
Faculty of medicine, University of Medicine and Pharmacy, Ho Chi Minh City, Vietnam

Nu-Vuong Nguyen-Tran
Faculty of Medicine, Vietnam National University Ho Chi Minh City

Nguyen Nhu Do Phuc
Department of Epidemiology, Institute of Public Health Ho Chi Minh City, Vietnam

Huu Loc Thai Tran
School of medicine, Viet Nam National University, Ho Chi Minh City, Vietnam

Ahmad Helmy Zayan
Department of Otolaryngology, Menoufia University, Menoufia, Egypt

Ahmed Saber Abdelrahman
Faculty of Medicine, Misr University for Science and Technology, Giza, Egypt

Thao Thanh Vu
School of Health and Biomedical Science, RMIT University, Australia

Khuong An Luu Nguyen
Ho Chi Minh City University of Medicine and Pharmacy Faculty of Medicine

Trang Thi Nguyen
Quy Nhon University

Y Nhi Le Vo
Ho Chi Minh City University of Medicine and Pharmacy

Nhat Minh Nguyen Chu
School of Medicine, Vietnam National University, Ho Chi Minh City, Vietnam

Quynh Thuy Huong Tran
School of Medicine, Vietnam National University, Ho Chi Minh City, Vietnam

My Hoan Pham
Ho Chi Minh City University of Medicine and Pharmacy

Hai Ngoc Vu
Ho Chi Minh City University of Medicine and Pharmacy

Nguyen Tien Huy (tienhuy@nagasaki-u.ac.jp)
Institute of Tropical Medicine (NEKKEN), Nagasaki University https://orcid.org/0000-0002-9543-9440

Research article

Keywords: Infectious diseases, ILI, Influenza-like illnesses, Outbreaks, Systematic review, Web-based queries

Posted Date: August 31st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-66084/v1

A systematic review of the correlation between web-based query and outbreak of emerging infectious diseases and meta-analysis of influenza-like illnesses

Nguyen Tran Minh Duc[1, 2], Nguyen Tran Nu Vuong[2, 3], Do Phuc Nhu Nguyen[2, 7], Tran Thai Huu Loc[2, 3], Ahmad Helmy Zayan[2,4], Ahmed Saber Abdelrahman[2, 5], Thao Thanh Vu[2,6], Luu Nguyen An Khuong[1,2], Nguyen Thi Trang[2,8], Vo Le Y Nhi[1, 2], Chu Nguyen Nhat Minh[2, 3], Tran Thuy Huong Quynh[2, 3], Pham Hoan My[1,2], Vu Ngoc Hai[1, 2], Nguyen Tien Huy[9, 10,11*]

Affiliations:

[1]Faculty of Medicine, University of Medicine and Pharmacy, Ho Chi Minh City, Vietnam

[2]Online research Club (http://www.onlineresearchclub.org/).

[3]School of Medicine, Viet Nam National University Ho Chi Minh City, Vietnam

[4]Department of Otolaryngology, Menoufia University, Menoufia, Egypt

[5]Faculty of Medicine, Misr University for Science and Technology, Giza, Egypt.

[6]School of Health and Biomedical Science, RMIT University, Australia.

[7]Department of Epidemiology, Institute of Public Health Ho Chi Minh City, Vietnam

[8]Quy Nhon City Hospital, Binh Dinh Province, Ho Chi Minh City, Viet Nam

[9]Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam.

[11]Department of Clinical Product Development, Institute of Tropical Medicine (NEKKEN), Leading Graduate School Program, and Graduate School of Biomedical Sciences, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan

Authors equally contributed the work.

*Corresponding author: Nguyen Tien Huy, Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam (E-Mail: nguyentienhuy4@duytan.edu.vn).

Emails:

NTMD: minhduc1298@gmail.com *(ORCID: 0000-0002-9333-7539)*

NTNV: ntnvuong.y2016@medvnu.edu.vn *(ORCID: 0000-0003-1167-7650)*

DPNN: dophucnhunguyen@iph.org.vn *(ORCID: 0000-0002-7644-4658)*

TTHL: tthloc.stu13@medvnu.edu.vn *(ORCID: 0000-0001-5518-0887)*

AHZ: ahmad_zayan@yahoo.com *(ORCID: 0000-0003-2581-0459)*

ASA: dr.ahmedsaber@icloud.com *(ORCID: 0000-0003-2049-6085)*

TTV: thaovuthanhthao@gmail.com *(ORCID: 0000-0003-0980-0293)*

LNAK: louis.lnak1410@gmail.com *(ORCID: 0000-0002-8558-5946)*

NTT: trangnguyenqn.dr@gmail.com *(ORCID: 0000-0003-4676-0905)*

VLYN: vly.nhi1995@gmail.com *(ORCID: 0000-0002-3738-623X)*

CNNM: cnnminh.y2016@medvnu.edu.vn *(ORCID: 0000-0002-3229-636)*

TTHQ: tranquynh.su@gmail.com *(ORCID: 0000-0002-6077-4783)*

PHM: hoanmy.yds@gmail.com *(ORCID: 0000-0003-1686-1465)*

VNH: dr.vungochai@gmail.com *(ORCID: 0000-0002-3202-8871)*

NTH: nguyentienhuy4@duytan.edu.vn *(ORCID: 0000-0002-9543-9440)*

Abstract

Background: Emerging infectious diseases (EIDs) are among the widespread ever-changing threats to public health. Web-based queries using information gathered from social media can enhance global syndromic surveillance to trace EIDs activity. This systematic review aimed to investigate the correlation of web-based queries to outbreak of EIDs.

Methods: Nine electronic databases were systematically searched and updated in August 2018 including; PubMed, Virtual Health Library (VHL), WHO Global Health Library (GHL), Scopus, ISI, Google Scholar, POPLINE, and Systems for Information of Grey Literature in Europe (SIGLE), New York Academy of Medicine (NYAM Grey Literature Report). A prior protocol was registered at Prospero (CRD42016038104). In a total five included articles, 47 datasets were included for reviewing. The correlation was assessed through Spearman and Pearson tests using either google trends or number of tweets.

Results: Meta-analysis of influenza-like illness data revealed that correlation was significant (0.784 (0.743-0.820, 0.964 (0.918-0.985) for both Spearman and Pearson tests respectively.

Conclusions: Web-based surveillance systems could serve as a good method in predicting events of EIDs.

Keywords

Infectious diseases; ILI; Influenza-like illnesses; Outbreaks; Systematic review; Web-based queries

Background:

Emerging infectious diseases (EIDs) have been escalating in the past 20 years and threatening to over 17 million deaths worldwide in public health every year [1]. During ongoing emerging infectious diseases, prediction using search query data provides an optimal robust and sensitive solution for rapidly detecting the distribution of diseases and other health conditions over time, forecasting disease outbreaks in different geographical areas and controlling an outbreak [2] . This query system proved its power in most recent epidemics, such as influenza epidemics [3]. Traditional surveillance systems reply on both virological and clinical data, then national and regional data is published on a weekly basis, frequently with a 1-2 week reporting lag [3]. In developing countries, surveillance for such detection is costly, and lack the public health framework to determine outbreaks at their earliest stages. Furthermore, the internet has freely available web-based sources of information and subsequently faster detection at low cost. Eighty percent of American internet users, or about 113 million adults, are believed to search online for health information about specific diseases or medical conditions [4], millions of people worldwide use online to search for health-related information each day, making web-based queries a valuable source of information on recent health trends [5]. This calls into question about the precision of these queries on the detection and estimation of the global EIDs burden. Therefore, this study aims to investigate the correlation of web-based queries to the outbreak of EIDs.

Methods

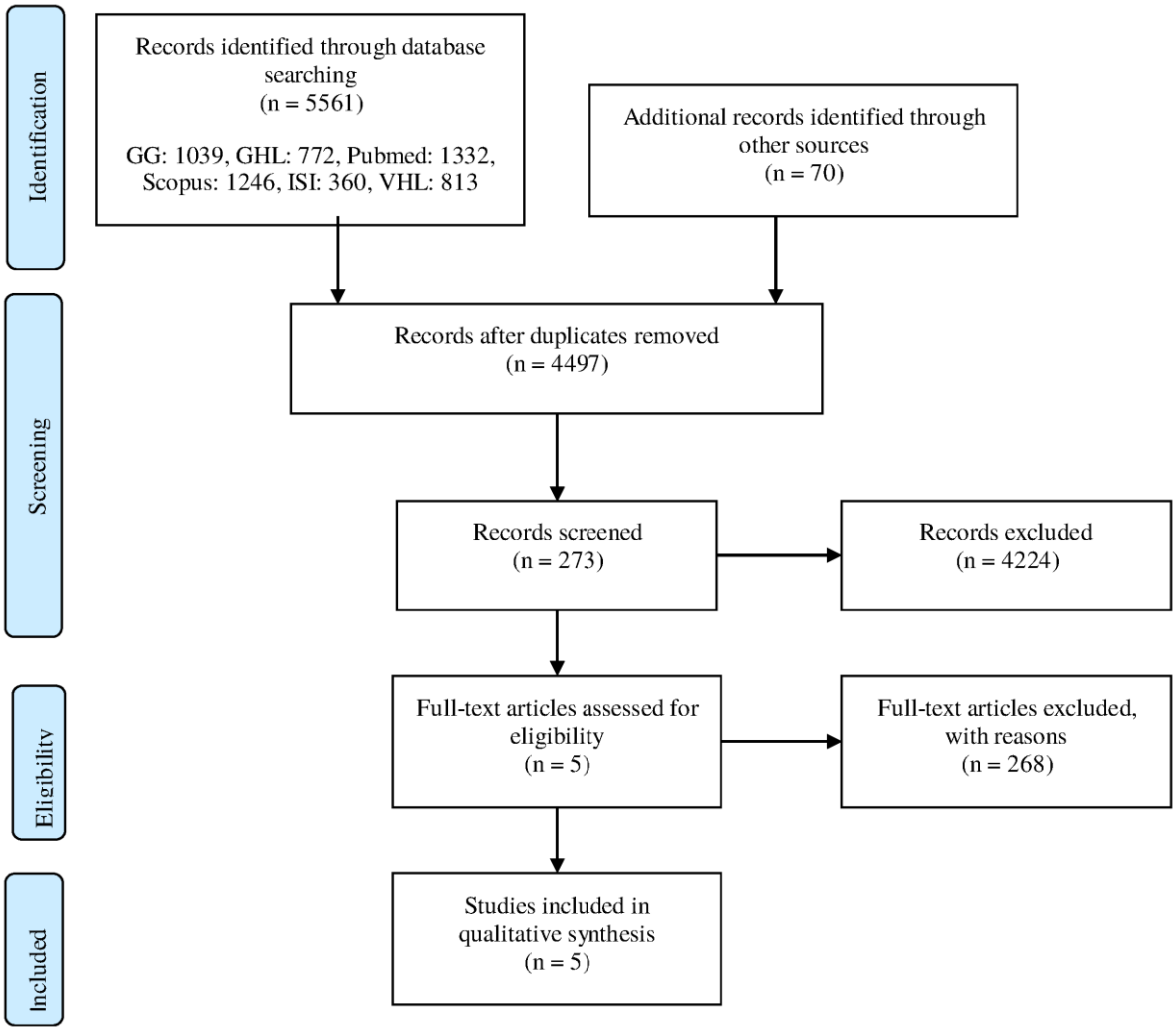

Figure 1. Flow diagram of study design

No.	ID	Country	Year of collecting data	Web query	Disease	Sample size	Correlation coefficient

1. Spearman's correlation

1	Milinovich/2014 [6]	Australia	2009-2013	Google	Pneumococcal disease (invasive)	8096	.882
					Gonococcal infection	55068	.847
					Ross River virus infection	23000	.811
					Varicella zoster (Unspecified)	37204	.783
					Influenza (laboratory confirmed)	165797	.775
					Dengue virus infection	6392	.754
					Chlamydia infection	362141	.699
					Pertussis	136494	.814
					Varicella zoster (Shingles)	17504	.783
					Varicella zoster (Chickenpox)	8843	.689
					Hepatitis C (unspecified)	48731	.617
					Barmah Forest virus infection	10327	.683
					Meningococcal disease (invasive)	1067	.586
					Hepatitis B (unspecified)	31965	.542
					Measles	635	.534
					Murray Valley encephalitis virus	23	.544
					infection	14297	.449
					Cryptosporidiosis	239	.436
					Chikungunya virus infection		
2	Aslam/2014 [9]	USA	2013-2014	Twitter	Sentinel-provided ILI rates in	3813	0.100
					Boston	5116	0.640
					Sentinel-provided ILI rates in	1497	0.600
					Chicago	1034	-0.240
					Sentinel-provided ILI rates in	1942	0.690
					Cleveland	2195	0.760
					Sentinel-provided ILI rates in	1236	0.850
					Columbus	1630	0.830
					Sentinel-provided ILI rates in	12632	0.550
					Denver	1808	0.880
					Sentinel-provided ILI rates in	3813	0.610
					Detroit	5116	0.800
					Sentinel-provided ILI rates in Fort	1497	0.750
					Worth	1034	0.870
					Sentinel-provided ILI rates in	1808	0.880
					Nashville-Davidson	2941	0.820
					Sentinel-provided ILI rates in New		
					York		
					Sentinel-provided ILI rates in San		
					Diego		
					Emergency department ILI rates in		
					Boston		
					Emergency department ILI rates in		
					Chicago		
					Emergency department ILI rates in		
					Cleveland		
					Emergency department ILI rates in		
					Columbus		

					Emergency department ILI rates in San Diego		
					Emergency department ILI rates in Seattle		
3	Nagar/2014 [10]	USA	2012-2013	Twitter	Influenza-like illness emergency department visit	2972	0.763
				Google	Influenza-like illness emergency department visit	2972	0.683
2. Pearson's correlation							
4	Broniatowski/2013 [8]	USA	2012-2013	Twitter	Weekly CDC ILI outpatient counts in US	104200	0.930
					Patient illness data in US	104200	0.750
5	Gesualdo/2013 [7]	Italy	2012-2013	Twitter	Traditional surveillance trends for US	5508	0.981
					ILINet data and tweets containing the words "flu" or "influenza"	5508	0.899
					ILINet with the control series of tweets containing ILI non-related Keywords	5580	0.292
					ILINet data and tweets responded to the US-GEO criteria	262853	0.769
					ILINet data and tweets responded to the US-WDE criteria	262853	0.980
					ILINet data and US NARROW tweets	262853	0.997
					ILINet data and tweets responded to the US-NARROW criteria	262853	0.974
					ILINet data and all tweets independently from localization	262853	0.977
					Tweet trend consistent with the ECDC case definition	262853	0.944
				Google	ILINet data with Google Flu Trends	5508	0.966

Table 1. Basic characteristics of the included studies.

We included original articles reporting the correlation of the web search query and number or prediction of cases and the correlation coefficient values such as R^2, Pearson or Spearman's coefficient without restriction in human (Supplementary Table 1). In February 2016, nine electronic databases were searched carefully to identify relevant articles and an update was conducted in February 2019. We used the search term which is presented in Supplementary Table 2. A manual search - of relevant articles, related searches, and

citations from the included articles, PubMed, and Google Scholar, respectively- was performed by three independent reviewers. Search results were imported into Endnote X7 (Thompson Reuter, CA, USA) to remove duplicates. Three independent reviewers screened the title and abstract by using the pre-determined eligibility criteria. If the title and abstract did not provide enough information, full-text reading is required. In the case of disagreement, the consensus was reached by discussion among reviewers or guidance from the supervisors. The standardized template was developed through a pilot extraction with the two most relevant references. Extracted data included basic and special information which are presented in **(Table 1)**. For the meta-analysis, we used the correlation between Google Trends or number of tweets and the best fit model of representative influenza-like illness data (emergency department visits, influenza illness surveillance program data (ILINet), Centers for Disease Control and Prevention counts (CDC), and laboratory-confirmed cases). The summary estimates with 95% Confidence Interval (CI) of each correlation coefficient would be calculated using the R statistical software 3.5 Heterogeneity would be assessed through the Chi-squared-based Q test or the I2 method. Fixed-effects model would be applied if no evident significant heterogeneity ($P < 0.05$., $I2 < 50\%$). Otherwise, the random-effects model was used [6]. All P values were two-sided and were considered statistically significantly less than 0.05.

Results

Among the 5674 records were found, distributed among the nine databases and after removing the duplicated articles, 4478 remained for analysis of the titles and abstracts. During the phase of screening the titles and abstract, four thousand and five studies were

excluded, and thus 273 studies were eligible for analysis of the complete text. After excluding 268 articles, five articles were eligible for review (Figure 1 and Table 1).

The included articles were published between 2013–2014, the most included infectious diseases were influenza and influenza-like illnesses. Of the five included studies, three were conducted in the United States, one in Italy and one in Australia. The study period varied with a duration ranging from 5 months to 9 years. Google trends were investigated in two articles (Millinovich et. al[6], Gesualdo et. al[7]) and Twitter data were investigated in the rest of articles (Broniatowski et.al[8], Aslam et al[9], Nagar et al[10])

Spearman correlation was investigated in three articles (Millinovich et. al[6], Aslam et al[9], Nagar et al[10]) and ranged between -0.24 to 0.88. Millinovich et al [6], investigated correlation google trends and 17 different infectious diseases using time series analysis on monthly notifications. They also concluded the potential applicability of web-based queries for vector-borne disease. In the second article (Aslam et al [9]), the weekly number of tweets of influenza/influenza-like illness without URL demonstrated higher accuracy values in predicting laboratory-confirmed cases for most categories compared to URL based ones (r=0.93). A spatiotemporal time analysis was used in Nagar et al[10] and revealed better correlation coefficients for daily flu tweets than daily google search volume (r=0.763 vs 0.683)

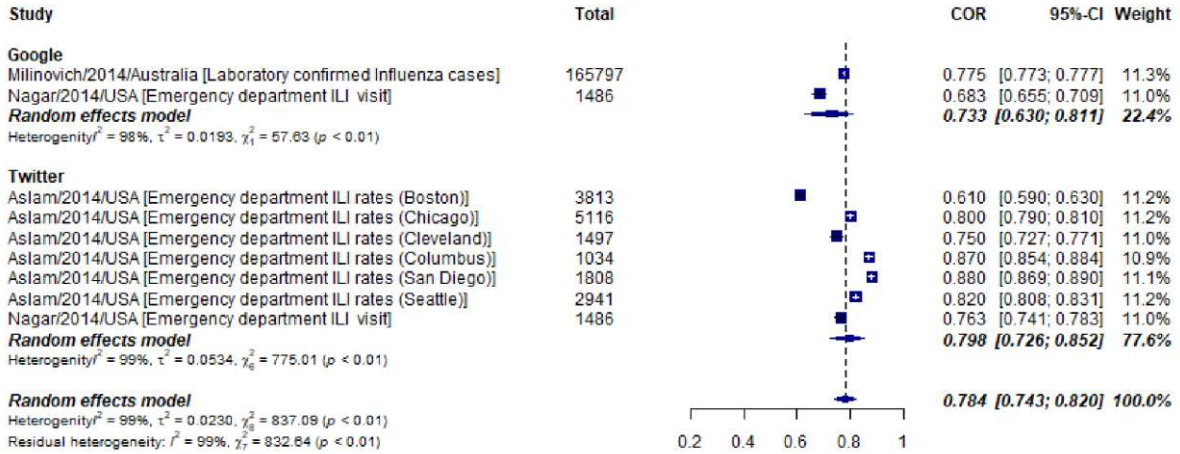

Figure 2. Meta-analysis of spearman correlation of influenza like illness across web-based queries, ILI=influenza like illness

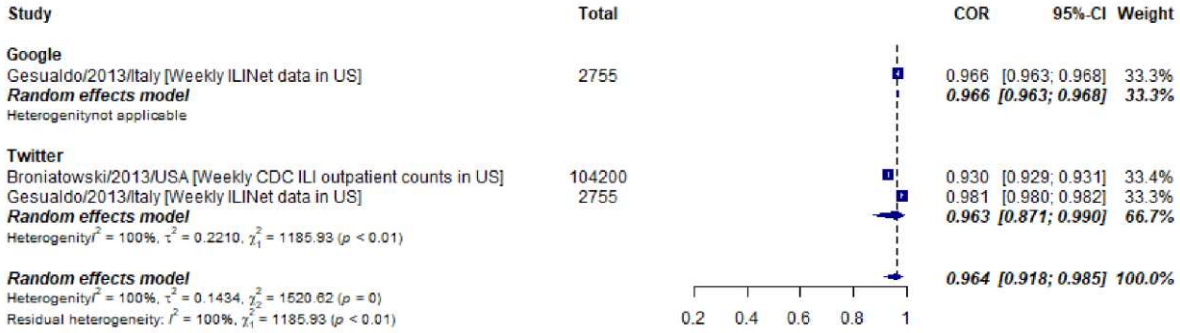

Figure 3. Meta-analysis of Pearson correlation of influenza like illness across web-based queries

Meanwhile, Pearson's correlation was studied in the other two articles (Broniatowski et al [8], Gesualdo et al [7]) and ranged between 0.29 to 0.98. Broniatowski et al [8] demonstrated the same significant efficacy of twitter data compared to CDC surveillance data for influenza-like illness patients for a lag of one week(r=0.93). In the same context, Gesualdo et al [7], demonstrated a significant correlation between influenza-positive tweets or google trends with ILINet data (r=0.98, 0.96 respectively)

Using the random-effects model, a meta-analysis of influenza-like illness data showed that correlation was significant (0.784 (0.743-0.820), 0.964 (0.918-0.985) for both Spearman and

Pearson tests respectively (Figure 2, 3). Subgroup analysis for results of Spearman correlation demonstrated better efficacy for twitter data (7 datasets, three articles) (0.798 (0.726-0.818) vs google trends (two datasets, two articles) (0.733 (0.630-0.811). Further subgroup analysis based on a period of data collection revealed that weekly data outperforms both daily and monthly data (0.803 (0.773-0.863), 0.725 (0.637-0.795), 0.775 (0.773-0.777), respectively).

Discussion

We investigated how well web queries submitted to the social media mimic the results from other systems for emerging infectious diseases (EIDs) surveillance. For the most common advantages, the web-based query could help track and predict ongoing pandemics for the most popular infectious diseases worldwide, and therefore planning for better prophylaxis and prevention [11]. As well, when these data were combined with other applications such as air traffic data, the query could enhance tremendously to the prediction of the spread of certain infectious diseases [12].

Through a meta-analysis of influenza/influenza-like illness data, we found a significant correlation at both Spearman and Pearson tests. Moreover, our results demonstrated better prediction values of twitter data versus google search, these were supported by results of Aslam et al. [9], Broniatowski et al. [8], and Nagar et al [7]. Regarding a period of data collection for trend analysis, we also demonstrated the better performance of weekly based model data. This can be explained by the temporal resolution of data if it was used on a monthly basis. [6]

From a different perspective, other internet-based surveillance, as in the study of Milinovich et al., was not only used in tracking and predicting influenza prevalence but also in the

management of other infectious diseases such as dengue fever. Through using a wide range of specific search terms, 17 infectious diseases (26.6%) were found to be significantly correlated [6]. They also recommended that search terms that present highly significant correlation should be kept for re-using as they can help in providing a quicker response on future emerging disease management

However, social media such as Twitter or Google can have a few limitations. A significant one is that we could not collect demographic data like age, sex and racial characteristics of patients via tweets, which could cause difficulty for the public health sector to make a response [9]. Another concern is that Twitter is used mainly in one group of the population, for example, people living in metropolitan areas, which may cause unavoidable bias in data retrieving as the data cannot represent the characteristic of the whole population [7]. Twitter also required a large human resource to classify the tweets, and therefore implied the potential of human error [10]. In addition, the search engine does not show the IP-address which could show us the specific location of the users; therefore, it is only possible to track the epidemic on the national scale [13]. But using the data with precise information connected to individuals could violate their privacy. Fifthly, because of the huge synonyms of query terms, the web-based tool may underestimate the real disease activity since a lot of terms may not be gathered [14]. Sixthly, forecasting health and disease-related phenomena have a very high chance of false-positive because people perceive their health status is very subjective. Thus, data sources must be carefully collected before all analyses [15]. Finally, due to the limited number of included articles, our study could hardly reflect accurately the clear correlation between web-based query prediction results and government announcement. Hence, we are looking forward to more research on how certain factors could alter

predictive results and in this way, developing tools to filter those factors in the attempt to complete the capacity of prediction thanks to web-based queries.

Conclusions

In conclusion, web-based surveillance systems could serve as a good method in predicting events of emerging infectious diseases.

List of abbreviations

EIDs: Emerging infectious diseases

ILI: Influenza-like illnesses

Declarations

Ethics approval and consent to participate:

Not applicable

Consent for publication:

Not applicable

Availability of data and material:

All data generated or analysed during this study are included in this published article.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This work was supported in part by a "Grant-in-Aid for Scientific Research (B)" (16H05844, 2016–2019 for Nguyen Tien Huy) from Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Authors' contributions:

NTMD, VNH, NTH: conceptualization, supervision; NTMD, NTNV, DPNN, TTV, LNAK, NTT, VLYN, CNNM, TTHQ, PHM: screening and data collecting, formal analysis and writing-original draft preparation; TTHL, AHZ, ASA: methodology, formal analysis, writing-reviewing, editing; and software. The authors read and approved the final manuscript.

Acknowledgements:

Not applicable.

Supplementary materials

Supplementary Table 1. PRISMA 2009 Checklist.

Section/topic	#	Checklist item	Reported on page #
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 4

Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 5
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Page 6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Page 7
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 5
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I2) for each meta-analysis.	Page 7

Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Page 7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	Page 17
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Page 7
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Page 7
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Page 5
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Page 16
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Page 17
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Page 7
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Page 17
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Page 12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Page 12
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Page 12
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	Page 13

Supplementary Table 2. Search terms of each included database.

Database	Result	Search term
PubMed	1332	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or “syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))
ISI:	360	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or “syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))
Scopus	1246	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or

		“syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))
SIGLE	0	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or “syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))
NYAM	0	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or “syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))
POPLINE	0	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or “syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or

		correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))
VHL	813	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or “syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))
GHL	772	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or “syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))
GOOGLE SCHOLAR	1039	(outbreaks or outbreak or epidemic or pandemic or dengue or infectious or influenza or flu or zika or foodborne or “food-borne” or waterborne or “water-borne” or “mosquito-borne” or vectorborne or “vector-borne”) and (forecasting or forecast or prediction or predictions or predict or predictive or predicted or “early warning” or monitor or monitoring) and (online or internet or web or google or googling or baidu or query or queries or infodemiology or “digital disease detection” or infoveillance or “real-time disease surveillance” or “syndromic surveillance” or “social media” or “social network” or Twitter or Facebook or Instagram) and (relationship or relation or correlation or correlations or correlated or correlate or Pearson or Spearman or (sensitivity and specificity))

References:

1. Key KK, DeNoon DJ. Infectious diseases kill over 17 million people a year. AIDS Wkly 1995. 1996;22–27.

2. 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;11(1):e11.

3. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457(7232):1012.

4. Fox S. Online health research is widespread, but few check the source and date. Medscape Gen Med. 2007;9(1):30.

5. Johnson HA, Wagner MM, Hogan WR, Chapman WW, Olszewski RT, Dowling JN, et al. Analysis of Web access logs for surveillance of influenza. In: Medinfo. 2004. p. 1202–1206.

6. Milinovich GJ, Avril SM, Clements AC, Brownstein JS, Tong S, Hu W. Using internet search queries for infectious disease surveillance: screening diseases for suitability. BMC Infect Dis. 2014;14(1):690.

7. Gesualdo F, Stilo G, Agricola E, Gonfiantini MV, Pandolfi E, Velardi P, et al. Influenza-Like Illness Surveillance on Twitter through Automated Learning of Naïve Language. PLOS ONE. 2013 Dec 4;8(12):e82489.

8. Broniatowski DA, Paul MJ, Dredze M. National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic. PLOS ONE. 2013 Dec 9;8(12):e83672.

9. JMIR - The Reliability of Tweets as a Supplementary Method of Seasonal Influenza Surveillance | Aslam | Journal of Medical Internet Research [Internet]. [cited 2019 Apr 10]. Available from: https://www.jmir.org/2014/11/e250

10. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, et al. A Case Study of the New York City 2012-2013 Influenza Season With Daily Geocoded Twitter Data From Temporal and Spatiotemporal Perspectives. J Med Internet Res. 2014;16(10):e236.

11. Wu S, Weng Y, Ye W, Wang L, Yan Y, Hong R, et al. [Role of syndromic surveillance program in the practice of early detection on disease outbreak]. Zhonghua Liu Xing Bing Xue Za Zhi Zhonghua Liuxingbingxue Zazhi. 2016 Apr;37(4):531–4.

12. Huang Z, Das A, Qiu Y, Tatem AJ. Web-based GIS: the vector-borne disease airline importation risk (VBD-AIR) tool. Int J Health Geogr. 2012 Aug 14;11(1):33.

13. Domnich A, Panatto D, Signori A, Lai PL, Gasparini R, Amicizia D. Age-related differences in the accuracy of web query-based predictions of influenza-like illness. PLoS One. 2015;10(5):e0127754.

14. Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. PloS One. 2009;4(2):e4378.

15. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic. PLOS ONE. 2011 Aug 19;6(8):e23610.

Figures

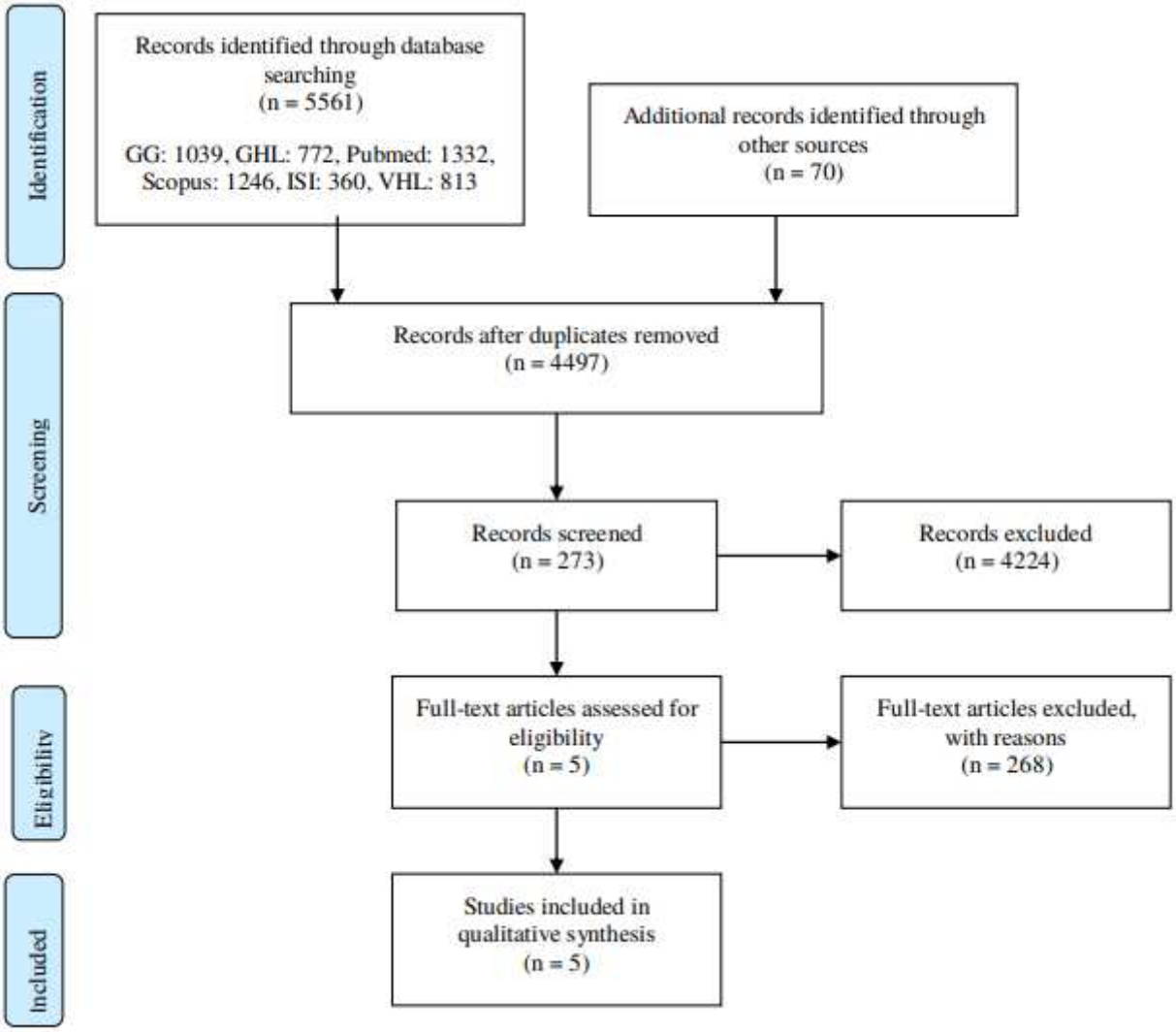

Figure 1

Flow diagram of study design.

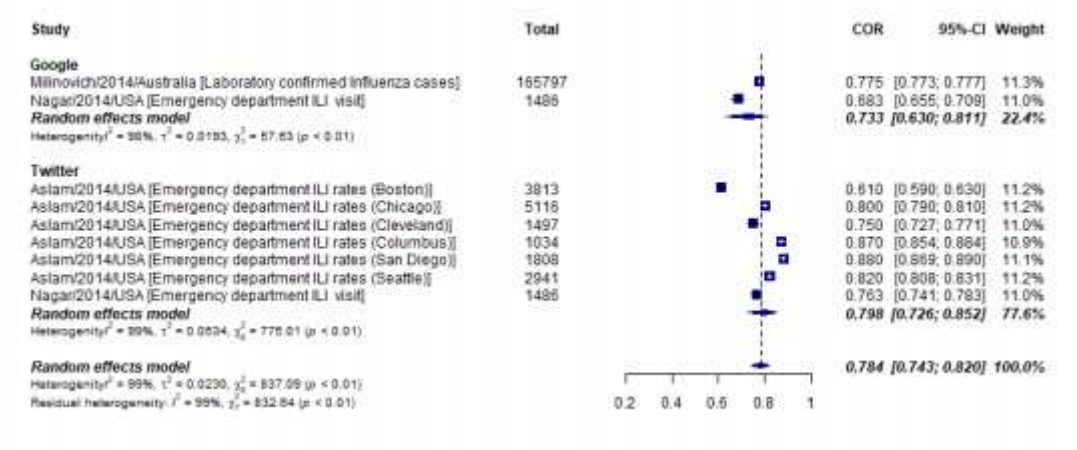

Figure 2

Meta-analysis of spearman correlation of influenza like illness across webbased queries, ILI=influenza like illness

Figure 3

Meta-analysis of Pearson correlation of influenza like illness across webbased queries, ILINet= influenza illness surveillance program

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table.pdf