View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by eprints Iran University of Medical Sciences

and Libraries Journal

Health Libraries Group

CILIP and information association

DOI: 10.1111/hir.12216

Review Article

Trends of infodemiology studies: a scoping review

Kimia Zeraatkar & Maryam Ahmadi

Department of Health Information Technology, Iran University of Medical Sciences, Tehran, Iran

Abstract

Introduction: The health care industry is rich in data and information. Web technologies, such as search engines and social media, have provided an opportunity for the management of user generated data in real time in the form of infodemiology studies. The aim of this study was to investigate infodemiology studies conducted during 2002–2016, and compare them based on developed, developing and in transition countries.

Methods: This scoping review was conducted in 2017 with the help of the PRISMA guidelines. PubMed, Scopus, Science Direct, Web of Knowledge, Google Scholar, Wiley and Springer databases were searched between the years 2002 and 2016. Finally, 56 articles were included in the review and analysed.

Results: The initial infodemiology studies pertain to the quality assessment of the hospital's websites. Most of the studies were on developed countries, based on flu, and published in the Journal of Medical Internet Research.

Conclusion: The infodemiology approach provides unmatched opportunities for the management of health data and information generated by the users. Using this potential will provide unique opportunities for the health information need assessment in real time by health librarians and thereby provide evidence based health information to the people.

Keywords: consumer health information; data mining; internet; review, scoping; social media; social networking; Web 2.0

Key Messages

- Data mining course construction for Health librarianship students and professionals, in formal education and continuing professional education (CPE), with the aim of conducting infodemiology studies by them.
- Conducting health information need assessment in real time, use of infodemiology approach by health librarians.
- Production of reliable, evidence based content by health librarians conducting infodemiology studies and, as a result, improve the health literacy of the people in real time.
- Conducting infodemiology studies in the field of health information librarianship using different Web tools, such as Telegram, Twitter, and Google Trends.

Introduction

Internet usage has been increasing in the present day with almost a third of the world's population relying on it, as reported in 2017. According to The International Telecommunication Union (2017) report, the number of young Internet users has increased in developed and developing countries to 94% and 67% in 2017, respectively. The Internet has changed the ways in which people search for health information. Some studies have shown that 177 million Americans use social

Correspondence: Maryam Ahmadi, Health Information Technology department, Iran University of Medical Sciences, Rashid Yasemi St, Tehran Province, Tehran 1996713883, Iran. E-mail: ahmadi.m@iums.ac.ir

© 2018 Health Libraries Group

media platforms and Internet queries to determine their medical condition or that of others (Casey, 2016; Jacobs, Amuta & Jeon, 2017). Recently, social media and other Web based data sources have been used to spread awareness on the outbreak of diseases (Brownstein, Freifeld, Reis & Mandl, 2008; Paul & Drezde, 2011; Signorini, Segre & Polgreen, 2011).

With the advent of the Web 2.0 paradigm, the Internet is being used as a means for the distribution of personal health information rather than simply as an information source. Also, with the advent of the Web 2.0 technologies, huge amounts of content are produced daily by users in the form of web pages, blogs and social networks (Chew & Eysenbach, 2010; Paul & Drezde, 2011; Santos & Matos, 2014; Scanfeld, Scanfeld & Larson, 2010).

This user generated content (UGC) or consumer created content (UCC) includes personal experiences, health information and knowledge (Bruns, 2016; Neiburger, 2010; Wyrwoll, 2014). Exploration, mining and analysis of user generated content (UGC) provide an image of the information seeking behaviour of people and tracking them over time can lead to the identification of the changes in their behaviour (Eysenbach, 2006).

Information and communication technologies have affected every aspect of the modern society enabling people to share and exchange knowledge. Interaction through social media and online communications helps people to make intelligent decisions (Wang, Zeng, Carley & Mao, 2007).

Internet researchers and developers have emphasised on the development of Health 2.0 or Medicine 2.0 (using Web 2.0 technologies for health or medicine). In this era of Health 2.0, patients share their health care experiences with other patients through Web technologies (Eysenbach, 2008).

Recent studies suggest that the information obtained from social media platforms, such as Twitter and Facebook, can be considered as supplements for epidemiological studies and traditional surveillance (Aslam et al., 2014; Yang, Horneffer & DiLisio, 2013). The information generated by social media regarding health care can be used for content analysis tracking in real time, knowledge translation as well as for the awareness of health policymakers (Eysenbach, 2006).

Google Trends is a particularly useful tool for the monitoring of Internet related activities concerning a particular topic, especially epidemics of infectious diseases over time. Google Trends offers several valuable insights into the people's health information seeking behaviour (Alicino et al., 2015; Ginsberg et al., 2009; Siri et al., 2016).

According to Statista (2017), Twitter is one of the most popular microblogging platforms with more than 328 million active users. Twitter is considered as a rich data source for performing public health surveillance. It offers an unprecedented opportunity for studying the people's information seeking behaviour (Kwak, Lee, Park & Moon, 2010; Paul & Drezde, 2011).

The Internet and social media platforms are considered as channels for the distribution and determination of health information in infodemiology. Although social media are potentially powerful tools for engaging and enabling users searching for relevant health information but the trustworthiness of the user generated content produced in them is questionable (Zhao & Zhang, 2017). As stated by Rutsaert, Pieniak, Regan, McConnon and Verbeke (2013), the main barrier preventing consumers from using social media as an information channel is trustworthiness.

The access to Internet data and its dissemination has created a new research field called infodemiology or the science of distribution and determination of health information in an electronic medium. The word 'infodemiology' was first used by Eysenbach in 2002. Although it was used for the first time for the identification of misinformation, it has been revealed that Internet queries could be used for predicting the influenza pandemic too. Eysenbach described the relationship between the flu related searches on Google and flu incidence data, and suggested that such an approach is better as it is faster than traditional epidemiologic surveillances. Since then, the term 'infodemiology' has been used to analyse the relationship between health information demands (through Web queries analysis) and health information supply (via social media data analysis). Using infodemiology with the aim of disease

© 2018 Health Libraries Group

surveillance is called infoveillance (Eysenbach, 2002, 2006, 2009).

Infodemiology is a new and emerging branch of science that deals with the occurrence, distribution and analysis of electronic health information to raise awareness in people on disease patterns. One of the main characteristics of infodemiology is the collection and analysis of data in real time (Culotta, 2010).

Infodemiology is useful in the field of public health and a wide range of other areas that include scientometrics 2.0 (The Pew Internet Project's Research, 2013). The research previously conducted showed that this methodology is valid for the identification of public health challenges (Koch-Weser, Bradshaw, Gualtieri & Gallagher, 2010). The real-time nature of this methodology means the results are fast while having a significant effect on the health policy. As people use the Internet and social media as information and news (McCully, Don & Updegraff, 2013; The Pew Internet Project's Research, 2014), these platforms can be seen as a new source of health data for public health surveillance (Eysenbach, 2002; Heaivilin, Gerbert, Page & Gibbs, 2011; Myslín, Zhu, Chapman & Conway, 2013; Somaiya, 2014), tracking health behaviours, attitudes (Cole-Lewis et al., 2015; Collier, Son & Nguyen, 2011; Kim et al., 2015; Sanders-Jackson, Brown & Prochaska, 2015; Zhang et al., 2013) and measuring the psychological traits of the community (Chan, Lopez & Sarkar, 2015; Eichstaedt et al., 2015).

The present study is the first attempt to systematically map the published literature on infodemiology studies. Therefore, a scoping review seemed the most appropriate research design (Levac, Colquhoun & O'Brien, 2010; Pham et al., 2014). The objective of this study was to conduct a scoping review of infodemiology studies conducted from 2002 onwards, as no such study has yet been carried out. Categorisation of infodemiology studies, based on developed, developing and in transition countries, is another concern of this review.

Methods

This scoping review was conducted in accordance with the PRISMA guidelines (Table 1).

The objective of this scoping review was to provide a descriptive overview of infodemiology studies without critically appraising individual studies, or synthesising evidence from different studies, then systematically map, and compare them based on developed, developing and in transition countries where they were performed. We identified the relevant studies by searching Scopus, Science Direct, Web of knowledge, Wiley, Springer, PubMed and Google Scholar, using a comprehensive search strategy. Our search was limited to 2002 onwards, as this is when the term infodemiology was first coined.

The search terms used were infodemiology, infoveillance, and e-epidemiology. No proximity operators or stop words were used. Boolean operators and free-text searching were used. The references of some articles were checked to identify relevant studies. The inclusion criteria for the research are as follows: (1) they should be in English language, (2) the publication year should be between 2002 and 2016, and (3) the infodemiology studies should be in the field of health care. The exclusion criteria were if they were in other languages or intended to present a model in infodemiology (Table 2).

A total of 1165 potential studies were identified for inclusion. At first, to avoid selection bias, the title and abstract of each identified study were assessed by the first reviewer (KZ) blinded to authors, affiliations and the publishing journal. Then, the titles and abstracts of the retrieved studies were independently reviewed by the second reviewer (MA). Finally, 95 studies met the inclusion criteria for the full-text review. The 39 studies that did not meet our inclusion criteria were excluded.

The 56 infodemiology studies that met our inclusion criteria and avoided the exclusion criteria were included in this review. To conduct these studies, the Web (1.0 & 2.0) tools were used. As our main aim was doing a scoping review of infodemiology studies during the 14 years, studies with no substantial use of the Web (1.0 & 2.0) tools were also included. Such studies usually relate to the initial days of inception of infodemiology when their authors called them infodemiology studies. They were based on the evaluation of hospital websites, identifying health

© 2018 Health Libraries Group

Table 1 PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
Title			
Title	1	Identify the report as a systematic review, meta-analysis, or both	1
Abstract			
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	1
Introduction			
Rationale	3	Describe the rationale for the review in the context of what is already known	3–4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)	3–4
Methods			
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	N/A
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	4
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	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated	4
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)	4
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	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made	5
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	10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means)	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis	N/A
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)	4, 10
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified	5
Results			
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	4
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	Table 4
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)	4,10
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (1) simple summary data for each intervention group (2) effect estimates and confidence intervals, ideally with a forest plot	N/A

(continued)

© 2018 Health Libraries Group

Table 1. (continued)

Section/topic	#	Checklist item	Reported on page #
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15)	4,10
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16])	5
Discussion			
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)	11
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)	9
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research	11, 12
Funding			
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	N/A

From: Moher, Liberati, Tetzlaff and Altman (2009). For more information, visit: www.prisma-statement.org.

information seeking behaviour and designing the information mining system.

At first, we examined all of the 56 articles included in this study and categorised them into two main groups based on the Web (1.0 & 2.0) tools that were used to conduct them. These two main groups were demand based and supply based infodemiology studies. The ‘demand based studies’ are those that were done using Web (1.0) tools, such as Google Trends and search engines queries. The ‘supply based studies’ were done using Web (2.0) tools such as Twitter, blogs, wikis, and online forums. The third group comprised of studies conducted using both Web (2.0) and Web (1.0) tools simultaneously; we called these studies ‘demand + supply studies’. The last group is ‘other studies’ in which no Web tools were used (Table 3).

Thereafter, a subanalysis of studies was conducted to extract further detail by topic domain, study aim, data sources, the journal in which each article was published, number of citations, search terms utilised, the country, analysis type and findings. These variables which were quoted from Nuti et al.’s systematic review (2014) were selected and defined in Table 4.

The data were extracted from the studies using the instrument presented in Table 3. Both authors performed data extractions, and disagreements between them were resolved by consensus. However, our primary aim for reporting these variables was not to state the reproducibility of the studies; it was to further analyse their content and provide the reader with an overview of how researchers are using Web tools to conduct the studies, if possible.

We tried to explain the variables for the demand based, supply based, demand + supply based studies. In the cases where the studies did not have the variables or they were not reported, such as in ‘other studies’, we used ‘N/A’, which means ‘not available’. We also tried to do a subgroup analysis in addition to performing data extraction for the main groups of studies. Efforts were made to capture the variables based on the aim of the study, and the analysis type was examined for every study included in this review.

Results

According to the inclusion and exclusion criteria of the 56 articles included in the review, the selection process of the studies and the reasons for withdrawal of articles are shown in the PRISMA flow chart (Fig. 1).

© 2018 Health Libraries Group

Table 2 Search string

Database	Google Scholar
Date of search	1 September 2016–30 December 2016
Search query	infodemiology or infoveillance, custom range (2002-2016)
Relative number of results	472
Database	PubMed
Date of search	1 September 2016–30 December 2016
Search query	((infodemiolog* OR infoveillance OR e-epidemiology)) AND ("2002"[Date - Publication]: "2016"[Date - Publication])
Relative number of results	98
Database	Scopus
Date of search	1 September 2016–30 December 2016
Search query	TITLE-ABS-KEY(infodemiolog* OR infoveillance OR e-epidemiology) AND PUBYEAR > 2002
Relative number of results	133
Database	Springer
Date of search	1 September 2016–30 December 2016
Search query	'(infodemiology OR infoveillance OR e-eoidemiology)' within 2002-2016
Relative number of results	54
Database	Wiley
Date of search	1 September 2016–30 December 2016
Search query	infodemiolog* OR infoveillance OR e-epidemiology in Keywords
Relative number of results	132
Database	ScienceDirect
Date of search	1 September 2016–30 December 2016
Search query	pub-date > 2001 and TITLE(infodemiolog* OR infoveillance OR e-epidemiology)
Relative number of results	151
Database	Web of Knowledge
Date of search	1 September 2016–30 December 2016
Search query	TS=(infodemiolog* OR infoveillance OR e-epidemiology)
Relative number of results	125

Data extracted from the articles are shown in Table 5.

All studies included in this review are original articles (100%). The results of this research indicate that the first studies, which were called infodemiology, were about the quality assessment of hospitals' websites (Kind, Wheeler, Robinson & Cabana, 2004; Liu, Bao, Liu & Wang, 2011). The other contexts that stand in the form of infodemiology studies include describing and assessing the Internet resources of patients in a specific field, designing the information mining system on specific topics (Chen, Chai, Zhang & Wang, 2014) and studying the information seeking behaviour.

Although some contemporary studies address both the demand based and supply based approaches (Bragazzi, Dini, Toletone, Brigo &

Table 3 Types of infodemiology studies

Type of study	Definition
Demand based studies	Studies carried out using Web (1.0) tools, such as Google Trends and search engines queries
Supply based studies	Studies carried out using Web (2.0) tools such as Twitter, blogs, wikis and online forums
Demand + Supply studies	Studies carried out using both Web (2.0) and Web (1.0) tools simultaneously
Other studies	Studies in which no Web tools were used

Durando, 2016a; Bragazzi, Watad, et al., 2016; Brigo & Erro, 2016; Lampos, Yom-Tov, Pebody & Cox, 2015; Nagar et al., 2014; Santos & Matos, 2014). The last group is categorised as 'other studies' in which no Web tools were used (Chen et al., 2014; Kim, Jung, Jung & Hur, 2014; Kind et al., 2004; Liu et al., 2011; Showghi & Williams, 2012).

The findings of the study indicates that out of the 56 papers included in the review, 27 articles (48.2%) were demand based studies, 18 (32.1%) were supply based studies, 6 (10.8%) were demand + supply studies, and the remaining 5 (8.9%) papers belonged to 'other studies' (Chart 1).

The publication rate of the demand based studies (27 papers) in comparison with the supply based studies (18 papers) was 3–2. It appears that the demand based studies had an increasing publication rate except in 2015, during which the publication rate remained stable. The publication rate of supply based studies remained stable in 2016. The results also showed that the articles were related to 13 countries in total.

Based on The United Nations' country classification (United Nations, 2014), the findings indicated that 46 (82%) articles were related to developed countries, as shown in Table 5, while seven (12.5%) were related to developing countries (Chan, Ho & Lam, 2013; Chen et al., 2014; Kim et al., 2014; Liu et al., 2011; Ocampo, Chunara & Brownstein, 2013; Wang et al., 2007; Wong et al., 2013). Three demand based articles (5%) belong to Russia as a country in transition (Domnich et al., 2014; Zheluk, Quinn, Hercz & Gillespie, 2013; Zheluk, Quinn & Meylakhs, 2014).

The list of developed countries that participated and the number of their papers are as follows: the United States (21), Italy (13), UK (5), Canada (4), Austria (1), Japan (1) and Portugal (1). The most supply based studies belong to USA with 12 papers (44%). The most demand based studies belong to Italy with nine papers (33%).

The list of developing countries that participated and the number of their papers is as follows: China (3) Hong Kong (1), South Korea (1), Thailand (1) and Taiwan (1) (Chart 2).

The Journal of Medical Internet Research was the most common journal publishing infodemiology studies, in total 25 (45%) of the 56 studies. Meanwhile, 37% of the demand based

Table 4 Variable definitions

1	Country	The country choose to study
2	Aim of study	
a	Descriptive	To describe temporal or geographic trends and general relationships
b	Surveillance	To evaluate the use of UGC* to predict or monitor real world phenomena
c	Casual inference	To evaluate a hypothesised causal relationship with UCG
3	Data sources	Web (1.0, 2.0) tools to conduct the studies such as Google Trends, Twitter, Wiki trends, online forums and search engine queries
4	Purpose	The primary aim of the study as stated in the introduction section of each article
5	Citations	The number of citations for each article as determined by Google Scholar on 20 May 2017
6	Search terms used	The search terms used to gain output. We report only five search terms for each paper, if provided
7	Analysis type	
a	Time trend analysis	Using UGC for comparisons across time period
b	Cross-sectional analysis	Using UGC for comparisons across different locations at a single time period or both
8	Primary findings	The main findings of the paper were abstracted

*UGC: User generated content, novel data streams generated by means of Web (1.0, 2.0) tools by users.

© 2018 Health Libraries Group

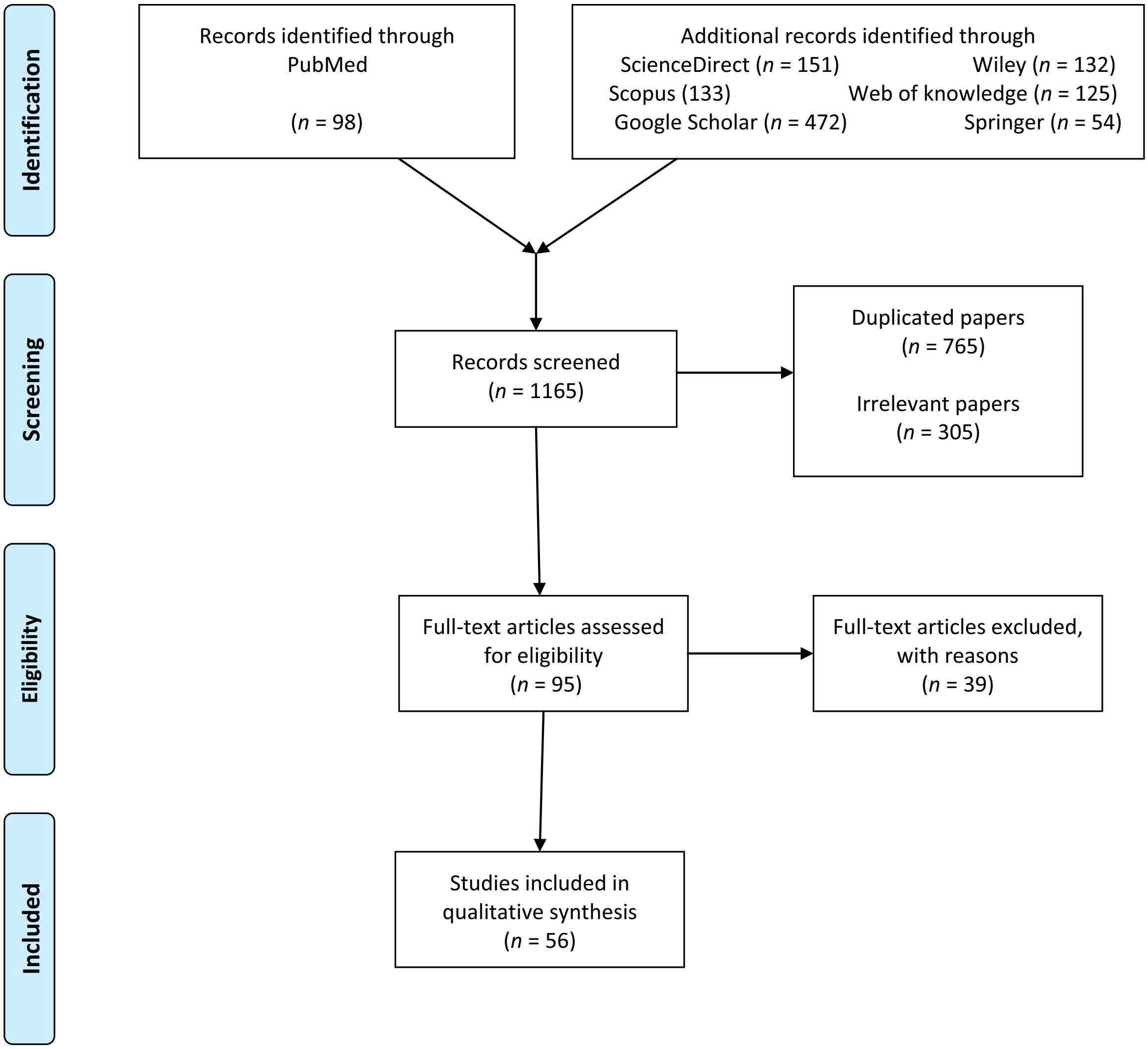

Figure 1 PRISMA flow diagram. [Colour figure can be viewed at wileyonlinelibrary.com].

studies (10 articles), 67% of the supply based studies (12 papers), 17% of demand + supply studies (1 article), and 40% of the 'other studies' (2 papers) were published in the Journal of Medical Internet Research. The median of the number of citations for each article was 33.6. Of course, the supply based studies with 61.7 (59%) citations per article had the maximum rate of citations of all the other groups. The demand based studies with 24 (34%) citations per article had the second closest rate of citations.

The topic that was covered most in the supply and demand based studies was influenza. Generally, 12 (21%) of the articles were related to 'flu' (Aslam et al., 2014; Chew & Eysenbach, 2010; Eysenbach, 2006; Fung et al., 2013; Hill et al., 2011; Lampos et al., 2015; Liang & Scammon, 2013; Nagar et al., 2014; Nagel et al., 2013; Santos & Matos, 2014; Tilston, Eames, Paolotti, Ealden & Edmunds, 2010; Yom-Tov, Johansson-Cox, Lampos & Hayward, 2015).

In the demand based studies, the highest topic frequency after flu was on multiple sclerosis (Bragazzi, 2013; Brigo, Lochner, Tezzon & Nardone, 2014) and suicide (Solano et al., 2016; Wong et al., 2013). In the supply based studies, the highest topic frequency after flu was on prescription drug abuse (Hanson, Cannon, Burton & Giraud-Carrier, 2013; Katsuki, Mackey & Cuomo, 2015) and e-cigarettes (Hua, Alfi & Talbot, 2013; Kim et al., 2015). The findings also showed that most articles have a descriptive aim

© 2018 Health Libraries Group

Table 5 Studies included in the review

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Alicino et al. (2015), Italy	Assessing Ebola-related web search behavior: Insights and implications from an analytical study of Google Trends-based query volumes	Descriptive	Google Trends	Demand-based	Ebola	Evaluate and interpret Google search queries for terms related to the Ebola outbreak both at the global level and in all countries where primary cases Ebola occurred	Infect Dis Poverty	9	Ebola"-related terms in local languages of the three most affected West Africa countries	Both	Google Trends showed a coarse-grained nature, strongly correlating with global epidemiological data, but was weaker at country level
Aslam et al. (2014), USA	The reliability of tweets as a supplementary method of seasonal influenza surveillance	Surveillance	Twitter	Supply-based	Flu	Improve the correlation of tweets to sentinel-provided influenza-like illness (ILI) rates by city through filtering and a machine-learning classifier	J Med Internet Res	23	Flu	Cross sectional	Increased accuracy in using Twitter as a supplementary surveillance tool for influenza as better filtering and classification methods yielded higher correlations for the 2013–2014 influenza season
Ayers et al. (2012), USA	A novel evaluation of World No Tobacco Day in Latin America	Surveillance	Health-related news stories and Internet search queries	Demand-based	World No Tobacco Day	Explore the potential of digital surveillance (infoveillance) to evaluate the impacts of WNTD on population awareness of and interest in cessation	J Med Internet Res	38	Cessation	Cross sectional	Cessation news coverage peaked around WNTD
Bragazzi (2013), Italy	Infodemiology and infoveillance of Multiple Sclerosis in Italy	Descriptive	Google Trends	Demand-based	Multiple sclerosis	Review the concept of "Multiple Sclerosis 2.0 Introduce a Google Trends-based approach for monitoring MS-related Google queries and searches	Mult Scler Int	11	Sclerosi multipla	Time series	Google Trends-based infodemiology and infoveillance system could be successfully applied also to monitor MS-related information
Bragazzi (2014), Italy	Google Trends-based approach for monitoring NSSI	Surveillance	Google Trends	Demand-based	NSSI	Introduce a Google Trends-based approach for monitoring NSSI	Psychol Res Behav Manag	19	'autolesionismo' (Italian for NSSI)	Both	There is the suspicion that NSSI is an increasing phenomenon
Bragazzi et al. (2016a), Italy	Infodemiology of status epilepticus: A systematic validation of the Google Trends-based search queries	Descriptive	Google Trends	Demand-based	Epilepsy	Providing a quantitative analysis of SE (status epilepticus) related web queries	Epilepsy Behav	5	Status epilepticus	Time series	Google Trends can foster new epidemiological studies in the field and can complement traditional epidemiological tools
	Infodemiological data of West-Nile virus disease in	Surveillance	Google Trends	Demand-based	West-Nile virus disease (WNVD)	Designing an approach for complementing traditional	Data brief	N/A	West-Nile virus disease (WNVD)	Both	Google Trends - generated data

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Bragazzi et al. (2016b), Italy	Italy in the study period 2004–2015					surveillance of the West-Nile virus disease					concerning WNVD well correlated with epidemiology and could be exploited for complementing traditional surveillance
Bragazzi et al. (2016c), Italy	How often people google for vaccination: Qualitative and quantitative insights from a systematic search of the web-based activities using Google Trends	Descriptive	Google Trends	Demand-based	Vaccination	How often people search the Internet for vaccination-related information if this search is spontaneous or induced by media which kind of information is in particular searched	Hum Vaccin Immunother	1	Vaccination	Time series	Google Trends can be used for monitoring the interest for vaccinations and the main Information searched
Bragazzi et al. (2016d), USA	Infodemiological data concerning silicosis in the USA in the period 2004–2010 correlating with real-world statistical data	Descriptive	Google Trends	Demand-based	Silicosis	Reporting data concerning silicosis-related web-activities using Google Trends(GT) Capturing the Internet behavior in the USA for the period 2004–2010	Data brief	N/A	Silicosis	Both	The temporal trend well correlated with the epidemiological data, as well as the geospatial distribution of the web-activities with the geographic epidemiology of silicosis
Bragazzi et al. (2016 e), Italy	Leveraging big data for exploring occupational diseases-related interest at the level of scientific community, media coverage and novel data streams: The example of silicosis as a pilot study	Descriptive	Google Trends Google news، Wiki trends، You Tube ،Twitter Pubmed/Medline Google scholar	Demand + Supply based	Silicosis	Investigated silicosis-related web-activities using Google Trends (GT) for capturing the Internet behavior worldwide in the years 2004 ± 2015	PLos One	4	Silicosis	Time series	A peak in silicosis-related web searches was noticed in 2010–2011
Bragazzi et al. (2016 f), Italy	Public health awareness of autoimmune diseases after the death of a celebrity	Descriptive	Google Trends Wikitrends, Google News, YouTube, Twitter	Demand + Supply based	Vasculitis	To investigate the relation between interests and awareness of an autoimmune disease after a relevant event such as the death of a celebrity	Clin Rheumatol	3	Vasculitis	Time series	Harold Ramis' death of vasculitis resulted into an increase in vasculitis-related Google searches, Wikipedia page access, news consumption, and, even though to a Lesser extent, tweet production

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Brigo et al. (2014), Italy	Web search behavior for multiple sclerosis: An infodemiological study	Descriptive	Google Trends	Demand- based	Multiple sclerosis	Evaluate changes in Web search behavior occurring in English-speaking countries over time for the term 'multiple sclerosis' (MS)	Mult Scler Relat Disord	15	Multiple Sclerosis	Time series	Over time there was a reduction in tendency to search for the term 'multiple sclerosis'. Most terms associated with the search queries for MS were related to causes and symptoms (including pain) of the disease Most peaks in search volume over the period studied corresponded to news of celebrities having MS
Brigo et al. (2015), Italy	Information-seeking behavior for epilepsy: an infodemiological study of searches for Wikipedia articles	Descriptive	Wikipedia Trends	Supply – based	Epilepsy	Evaluate information-seeking behavior of English-speaking internet users searching Wikipedia for articles related to epilepsy and epileptic seizures	Epileptic Disord	6	'epilepsy and driving' 'epilepsy and employment', 'epilepsy in children' 'epileptic seizure' 'seizure types'	Time series	Fears and worries about epileptic seizures, their impact on driving and employment, and news about celebrities with epilepsy might be major determinants in searching Wikipedia for information
Brigo and Erro (2016), Italy	Why do people google movement disorders? An infodemiological study of information seeking behaviors	Descriptive	Google Wikipedia	Demand + supply based	Movement disorders	Evaluate and interpret web search queries for terms related to movement disorders (MD) in English-speaking countries and their changes over time.	J Neurol Sci	2	Parkinson's disease, Huntington disease, Wilson's disease Chorea, Dystonia	Time series	The highest peaks of MD search queries were temporally related to news about celebrities suffering from MD, to specific mass-media events or to news concerning pharmaceutic companies or scientific discoveries on MD

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Burton et al. (2012), USA	'Right Time, Right Place' health communication on Twitter: Value and accuracy of location information	Surveillance	Twitter	Supply - based	Health communication	Document the prevalence of the location identification options available through Twitter and to present an estimate of the usability of each option	J Med Internet Res	56	N/A	Cross sectional	Majority of tweets are not currently associated with an identifiable geographic location
Chan et al. (2013), China	Infodemiology of alcohol use in Hong Kong mentioned on blogs: Infoveillance study	Surveillance	Google Blog Search	Supply – based	Alcohol use	To assess and explain the online use of alcohol-related Chinese keywords and to validate blog searching as an infoveillance method for surveying changes in drinking patterns in Hong Kong people in 2005–2010	J Med Internet Res	N/A	Alcohol beer or wine spirit	Time series	The online use of alcohol-related concepts increased noticeably for 'alcohol' in 2008 and 'spirit' in 2008–2009 but declined for 'beer or wine' over the years
Chen et al. (2014), China	Development and application of a Chinese webpage Suicide Information Mining System (Sims)	Descriptive	N/A	Other studies	Suicide information system	Designing and piloting a convenient Chinese webpage suicide information mining system (SIMS) to help search and filter required data from the internet and discover potential features and trends of suicide	J Med Sys	N/A	N/A	N/A	Designing and piloting a convenient Chinese webpage suicide information mining system (SIMS) to help search and filter required data from the internet and discover potential features and trends of suicide
Chew and Eysenbach (2010), Canada	Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak	Descriptive	Twitter	Supply – based	H1N1	Monitor the use of the terms 'H1N1' versus 'swine flu' over time Conduct a content analysis of 'tweets' Validate Twitter as a real-time content, sentiment, and public attention trend-tracking tool	PLOS ONE	708	Swine flu, and/or H1N1	Time series	Tweets can be used for real-time content analysis and knowledge translation research, allowing health authorities to respond to public concerns
Curtis et al. (2015), USA	Using web searches to track interest in synthetic cannabinoids	Descriptive	Google Trends	Demand – based	Synthetic cannabinoids	Determine the availability of websites offering to sell synthetic cannabinoids in the United States	Drug Alcohol Rev	8	'synthetic marijuana,' 'synthetic weed,' 'K2 Spice', 'herbal incense'	Both	From its emergence in the US in 2009, the use of synthetic cannabinoids has burgeone between 2010 and 2011 there was a substantial increase

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Domnich et al. (2014), Russia	Demand-based web surveillance of sexually transmitted infections in Russia	Surveillance	Yandex	Demand – based	Sexually transmitted infections	Investigate the possibility of using HIV- and syphilis-related web queries to predict incident diagnosis rates of sexually transmitted infections in Russia	Int J Public Health	2	'HIV' and 'syphilis' (in Russian)	Cross sectional	A high positive correlation between notification rates and search volume was observed
Eysenbach G (2006), Canada	Infodemiology: Tracking flu-related searches on the web for syndromic surveillance	Surveillance	Google Adsense	Demand – based	Flu	Using prospectively gathered data on Internet search trends for	AMIA Annu Symp Proc	268	'flu' or 'flu symptoms'	Time series	There is an excellent correlation between the number of clicks on a keyword-triggered link in Google with epidemiological data from the flu season 2004/2005 in Canada
Fung et al. (2013), USA	Chinese social media reaction to the MERS-CoV and avian influenza A (H7N9) outbreaks	Descriptive	Weibo (Chinese version of Twitter)	Supply- based	Flu	Use Weibo as a measure of the Chinese people's reactions to two different outbreaks : MERS-CoV, H7N9	Infect Dis Poverty	25	H7N9 SARS Avian flu	Time series	The reaction to the H7N9 outbreak was two orders of magnitude stronger than the reaction to the MERS-CoV outbreak
Guy et al. (2011), USA	Social media: A systematic review to understand the evidence and application in infodemiology	Descriptive	Social media	Supply- based	Social media	Systematically review the literature utilizing social media as a source for disease prediction and surveillance	eHealth	7	Pandemic Epidemic Communicable disease Social Media, Disease surveillance	Time series	Open-source micro-blogging sites use to inform influenza-like-illness monitoring
Hanson et al. (2013), USA	An exploration of social circles and prescription drug abuse through Twitter	Descriptive	Twitter	Supply – based	Prescription drug abuse	Investigate the social circles of prescription drug abusers on Twitter Observe the discussion and engagement of these users regarding prescription drug abuse	J Med Internet Res	53	Adderall Xanax Valium Ritalin Ativan	Time series	Twitter users who discuss prescription drug abuse online are surrounded by others who also discuss it—potentially reinforcing a negative behavior and social norm

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Hill et al. (2011), USA	Natural supplements for H1N1 influenza: retrospective observational infodemiology study of information and search activity on the internet	Descriptive	Google Trends	Demand – based	H1N1	Identify and characterize websites that provide information about herbal and natural supplements with information about H1N1 To examine trends in the public's behavior in searching for information about supplement use in preventing or treating H1N1	J Med Internet Res	11	Herb, Natural, Flu, Cold, Swine flu, H1N1	Time series	A large number of websites support information about supplements and H1N1
Hua et al. (2013), USA	Health-related effects reported by electronic cigarette users in online forums	Descriptive	On line forums	Supply – based	E – cigarette	Document the positive and negative short-term health effects produced by e-cigarette use through an analysis of original posts from three online e-cigarettes forums	J Med Internet Res	98	Electronic cigarette forum	Time series	E-cigarette use can have wide ranging positive and negative effects and that online forums provide a useful resource for examining how e-cigarette use affects health
Kamenova et al. (2014), Canada	Representations of stem cell clinics on Twitter	Descriptive	Twitter	Supply - based	Stem cell	Provide insights into designing health marketing interventions to promote physical activity on Twitter	Stem Cell Rev	7	Nine stem cell treatment providers	Both	Explicit claims or suggestions of benefits associated with unproven stem cell treatments in approximately one third of the tweets
Katsuki et al. (2015), USA	Establishing a link between prescription drug abuse and illicit online pharmacies: Analysis of Twitter data	Surveillance	Twitter	Supply - based	Prescription drug abuse	Surveillance and analysis of Twitter data to characterize the frequency of NUPM-related tweets Identifies illegal access to drugs of abuse via online pharmacies	J Med Internet Res	13	Nonproprietary name + Brand name+ Common 'street' or 'slang' names of drugs	Time series	Over 45 000 tweets that directly promoted NUPM by providing a URL that actively marketed the illegal online sale of prescription drugs of abuse
Kim et al. (2014), Sought Korea	Investigating the congruence of crowd sourced Information with official government data: The case of pediatric clinics	Casual inference	Official government Data Crowd-sourced data from online communities	Other studies	Pediatric clinics	Investigates the congruence between crowd sourced information and official government data in the health care domain and identifies the determinants of low congruence where it exists	J Med Internet Res	3	N/A	Cross sectional	The congruence of crowd-sourced information on health care services varied across regions These variations could be explained by geo-specific demographic factors

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Kim et al. (2015), USA	Using Twitter data to gain insights into e-cigarette marketing and locations of use: An infoveillance study	Surveillance	Twitter	Supply – based	E – cigarette	Gain insights into marketing trends for selling and promoting e-cigarettes & locations where people use e-cigarettes	J Med Internet Res	16	Electronic cigarette eCig Green smoke Vaping smoking Vaping	Both	Twitter data can provide new insights on e-cigarettes to help inform future research, regulations, surveillance, and enforcement efforts
Kind et al. (2004), USA	Do the leading children's hospitals have quality web sites? A description of children's hospital web sites	Descriptive	Web sites of 26 nationally prominent children's hospitals in the United States	Other studies	Characteristics of the Web sites of leading children's hospitals	Describe technical characteristics of the Web sites of leading children's hospitals	J Med Internet Res	27	N/A	Cross sectional	These Web sites are a potentially useful source of patient information although children's hospitals need to keep up with increasingly high standards and demands of health-care consumers
Lampos et al. (2015), UK	Assessing the impact of a health intervention via user-generated Internet content	Casual inference	Twitter Search query logs from Bing search engine	Demand + supply based	Flu vaccination program	Introduce a complementary framework for evaluating the impact of a targeted intervention, such as a vaccination campaign against an infectious disease, through a statistical analysis of user-generated content submitted on web platforms	Data Min Know Disc	9	Chills, Cough, Fatigue, Fever, Flu	Cross sectional	The impact estimates derived from the application of the proposed statistical framework support conventional assessments of the campaign
Liang and Scammon (2013), USA	Incidence of online health information search: A useful proxy for public health risk perception	Surveillance	Google insight	Demand – based	ILI	Investigate the extent to which public health risk perception as demonstrated by online information searches related to a health risk can be explained by the incidence of the health risk and social components of a specific population's environment	J Med Internet Res	11	Flu shots, Flu prevention, Flu symptoms, Flu fever, Treatment of the flu	Both	The social environment influences public risk perception regardless of disease incidence Monitoring the social variables can be very helpful in being ready to respond to the public's behavior in dealing with public health threats

(continued)

© 2018 Health Libraries Group

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Liu et al. (2011), China	The quality and characteristics of leading general hospitals' websites in China	Descriptive	Hospital websites of China	Other studies	Quality of hospital websites	Evaluation of quality of hospital websites in China	J Med Sys	14	N/A	Cross sectional	Websites of most hospitals showed a good performance in website content, showed a normal performance in website function and design, but showed a bad performance in website management & usage
Mnadla et al. (2016), Italy	Infodemiological data of Ironman Triathlon in the study period 2004–2013	Descriptive	Google Trends	Demand – based	Ironman Triathlon	Reports data concerning the Internet-related activities and interest for Ironman Triathlon competition	Data brief	N/A	Ironman Triathlon	Both	The interest for Ironman Triathlon was found to be cyclic over time
Mishori et al. (2014), USA	Mapping physician Twitter networks: Describing how they work as a first step in understanding connectivity, information flow, and message diffusion	Descriptive	Twitter	Supply – based	Physician networks	-Describes the characteristics of four medical networks Analyzes their theoretical dissemination potential, their actual dissemination, and the propagation and distribution of tweets.	J Med Internet Res	16	Medical networks : AMA AAFP AAP ACP	Cross sectional	-To increase the dissemination potential, medical groups should develop a more cohesive community of shared followers. Tweet content must be engaging to provide a hook for retweeting and reaching potential audience
Nagar et al. (2014), USA	A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives	Surveillance	Twitter Google Trends Search query data	Demand + supply based	Flu	Validate the temporal predictive strength of daily Twitter data for influenza-like illness emergency department (ILI-ED) visits during the New York City 2012–2013 influenza season against other available and established datasets (Google search query, or GSQ)	J Med Internet Res	36	'flu', 'influenza', 'gripe' 'high fever'	Both	Twitter's strength both qualitatively and quantitatively for ILI-ED prediction compared to alternative daily datasets mixed with awareness-based data such as GSQ
Nagel et al. (2013), USA	The complex relationship of real space events and messages in cyberspace: Case study of influenza and pertussis using Tweets	Surveillance	Twitter	Supply – based	Flu	Explore the interaction between cyberspace message activity, measured by keyword-specific tweets, and real world occurrences of influenza and pertussis	J Med Internet Res	34	Flu, Influenza, Pertussis, Whooping cough	Both	Tweets can function as a signal of disease activity and public interest

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Nakada et al. (2014), Japan	Development of a national agreement on Human Papillomavirus Vaccination in Japan: An infodemiology study	Descriptive	The promotion of HPV vaccination articles appearing in major Japanese newspapers and Web pages	Demand – based	Human Papilloma virus Vaccination	Identify the role of print and online media references, including references to celebrities or other informants, as factors potentially responsible for the relatively rapid national acceptance of HPV vaccination in Japan	J Med Internet Res	3	Recruitment Vaccine, Screening, Cervarix, Prevention, Vaccination Side effect, Infertility danger	Time series	The rapid development of a national agreement regarding HPV vaccination in Japan may be primarily attributed to the advocacy of vaccine beneficiaries, supported by advocacy by celebrities and positive reporting by print and online media
Nascimento et al. (2014), USA	Real-time sharing and expression of migraine headache suffering on Twitter: A cross-sectional infodemiology study	Descriptive	Twitter	Supply –based	Migraine	Using social media to assess migraine headache impact in real time to avoid memory bias Identify a set of current suffering descriptors that were not prompted by an experimenter	J Med Internet Res	20	Migraine	Cross sectional	The modern characteristics and broad impact of migraine headache suffering on patients' lives as it is spontaneously shared via social media
Nuti et al. (2014), Austria	The use of Google Trends in health care research: A systematic review	Descriptive	Google Trends	Demand – based	Health care	Systematically reviewed health care literature using Google Trends to classify articles by topic and study aim	PLOS ONE	62	Google Trends Google Insights	Time series	Google Trends is being used to study health phenomena in a variety of topic domains in myriad ways
Ocampo et al. (2013), Thailand	Using search queries for malaria surveillance, Thailand	Surveillance	Google Correlate	Demand – based	Malaria	Determine whether query data from malaria-related Google searches can be used to predict existing malaria surveillance trends for Thailand	Malar J	17	Malaria, Mosquito, Anemia, Common house mosquito, Eradicate mosquito	Time series	This methodology provides a cost-effective description of malaria prevalence that can act as a complement to traditional public health surveillance
Priest et al. (2016), USA	Finding the patient's voice using big data: Analysis of users' health-related concerns in the ChaCha question-and-answer service (2009–2012)	Descriptive	ChaCha Question-and-Answer Service	Demand – based	Sexual health	Describe user characteristics and health-related queries of the ChaCha question-and-answer platform	J Med Internet Res	3	Health queries	Time series	The private nature of the ChaCha service provided a perfect environment for maximum frankness among users, especially among adolescents posing sensitive health questions

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Santos and Matos (2014), Portugal	Analyzing Twitter and web queries for flu trend prediction	Surveillance	Twitter Web queries	Demand + supply based	Flu	Evaluates the use of Twitter messages and search engine query logs to estimate and predict the incidence rate of influenza like illness in Portugal	Theor Biol Med Model	21	Gripe Constipação body/throat pains, Headache, Fever	Time series	By changing the initial steps of data preprocessing and feature extraction and selection, the proposed approaches can be adapted to other languages
Siri et al. (2016), Canada	Infodemiological data of high-school drop-out related web searches in Canada correlating with real-world statistical data in the period 2004–2012	Descriptive	Google Trends	Demand- based	High-school drop-out	Describes high-school drop-out related web activities in Canada, 2004–2012	Data Brief	N/A	High-school drop-out	Both	The peak in the Internet-related activities occurs in 2004
Solano et al. (2016), Italy	A Google-based approach for monitoring suicide risk	Descriptive	Google Trends (Italian version)	Demand- based	Suicide	Investigate the relationship between suicide-rates and Google suicide-related search volumes in the Italian population (2008–2012) using the Italian mortality database that provided monthly national data concerning suicides (2008–2012)	J Psychiatr Res	17	'suicide', 'commit suicide', 'how to commit suicide'	Time series	Online searches for suicide-related terms in Italy are more likely to be linked to factors other than suicidality such as personal interest and suicide bereavement
Showghi and Williams (2012), UK	Information about male chronic pelvic and urogenital pain on the internet: An evaluation of internet resources	Descriptive	Hospital Websites	Other studies	Urogenital pain	Describe and evaluate the Internet resources available to patients searching for information about chronic urogenital/pelvic pain	Pain Med	6	Urogenital pain, Chronic pelvic pain, Chronic prostatitis, Penile pain, Pudendal neuralgia	Time series	There is a wealth of information available online, but much is of poor quality, and taken together, is likely to confuse more than enlighten patients attempting to understand male urogenital/pelvic pain symptoms or supplement information from health care professionals

(continued)

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Tilston et al. (2010), UK	Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic	Surveillance	UK flu survey	Demand – based	H1N1	Trends in incidence can be captured by Internet-based surveillance perhaps even more reliably than standard GP-based systems	BMC Public Health	62	N/A	Both	This crude dataset considerably overestimates the relative height of the first peak when compared with the HPA case estimates
Wang et al. (2015), Taiwan	Forecasting the incidence of dementia and dementia-related outpatient visits with Google Trends: evidence from Taiwan	Causal inference	Google Trends	Demand – based	Dementia	Determine whether Google Trends could provide insight into trends in dementia incidence and related outpatient visits in Taiwan	J Med Internet Res	10	Dementia; Alzheimer's disease; Parkinson's disease; Vascular dementia, Geriatric dementia	Time series	
Wong et al. (2013), Hong Kong	Accessing suicide-related information on the internet: A retrospective observational study of search behavior	Descriptive	America Online (AOL) subscribers' web queries	Demand – based	Suicide	Investigate what webpages people actually clicked on after searching with suicide-related queries on a search engine Examine what queries people used to get access to pro-suicide websites	J Med Internet Res	20	Suicide	Time series	
Yin et al. (2015), USA	A scalable framework to detect personal health mentions on Twitter	Descriptive	Twitter	Supply – based	Personal health	Develop a scalable framework to detect personal health status mentions on Twitter and assess the extent to which such information is disclosed	J Med Internet Res	9	Cancer, Depression, Hypertension, Leukemia, SYND	Time series	It is possible to automatically detect personal health status mentions on Twitter in a scalable manner.
Yom-Tov and Gabrilovich (2013), USA	Post market drug surveillance without trial costs: Discovery of adverse drug reactions through large-scale analysis of web search queries	Surveillance	Web queries submitted to the Yahoo US Web search engine	Demand – based	Drug surveillance	Propose a low cost, fully automated method for continuous monitoring of adverse drug reactions in single drugs and in combinations thereof	J Med Internet Res	44	20 top-selling drugs in the United States by revenue	Time series	

(continued)

© 2018 Health Libraries Group

Table 5. (continued)

Author/s (year), Country	Title	Aim of study	Data sources	Study type	Topic	Purpose	Journal	citations	Search terms used	Analysis	Primary findings
Yom-Tov et al. (2015), UK	Estimating the secondary attack rate and serial interval of influenza-like illnesses using social media Estimating the secondary attack rate and serial interval of influenza-like illnesses using social media Estimating secondary attack rate and serial interval of influenza-like illness using social media	Surveillance	Twitter	Supply – based	ILI	Propose a method for estimating the intra-family the secondary attack rate (SAR) and serial interval (SI) from posting on the twitter social network secondary attack rate serial interval	Influenza Other Respir Viruses	3	Bad cough Bed flu Chest infection Chesty cough Cold flu Cough	Time series	The proposed method can assist health authorities by providing near real time of SAR and SI and especially in alerting to sudden increases thereof.
Zhang et al. (2013), USA	Electronic word of mouth on Twitter about physical activity in the United States: Exploratory infodemiology study	Descriptive	Twitter	Supply – based	Physical activity	Provide insights into designing health marketing interventions to promote physical activity on Twitter	J Med Internet Res	17	Aerobics Badminton Baseball Basketball Bicycling bike	Time series	People with more followers were more likely to post neutral tweets about physical activity Users with fewer followers and with fewer followings were more likely to talk positively about physical activity on Twitter
Zheluk et al. (2013), Russia	Internet search patterns of Human Immunodeficiency Virus and the digital divide in the Russian federation: Infoveillance study	Surveillance	Google Trends Yandex	Demand – based	HIV	- Validate Internet search patterns against national HIV prevalence data Investigate the relationship between search patterns and the determinants of Internet access	J Med Internet Res	8	HIV" (VICh), 'AIDS' (SPID) in Russian	Cross sectional	
Zheluk et al. (2014), Russia	Internet search and Krokodil in the Russian federation: An infoveillance Study	Surveillance	Google Trends Yandex	Demand – based	Krokodil	Determine if Internet search patterns could detect regularities in behavioral responses to Russian CCM policy at the population level Determine if complementary data sources could explain the regularities we observed.	J Med Internet Res	3	Desomorphine, How to prepare desomorphine, Krokodil Desomorphine, Krokodil Drug	Both	

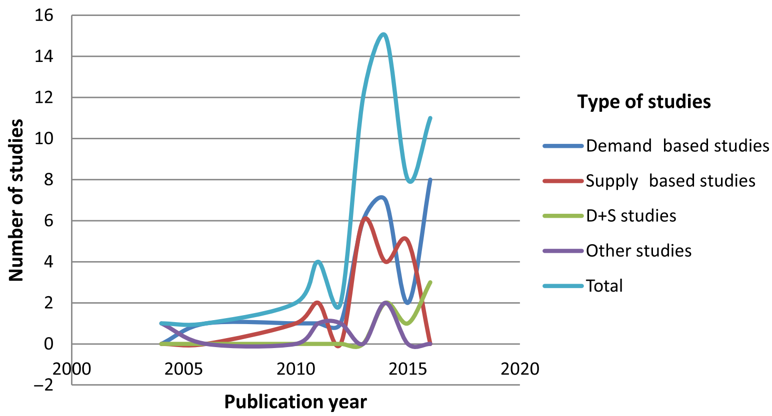

Chart 1 Frequency of studies based on publication year

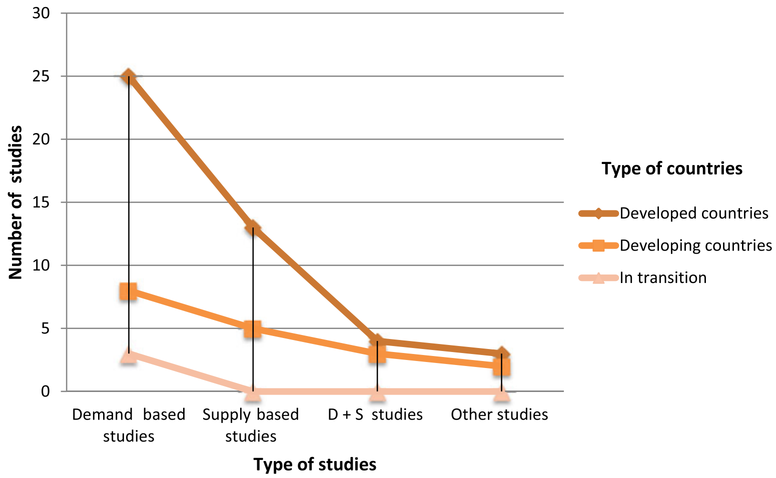

Chart 2 Frequency of studies based on developed, developing & in transition countries.

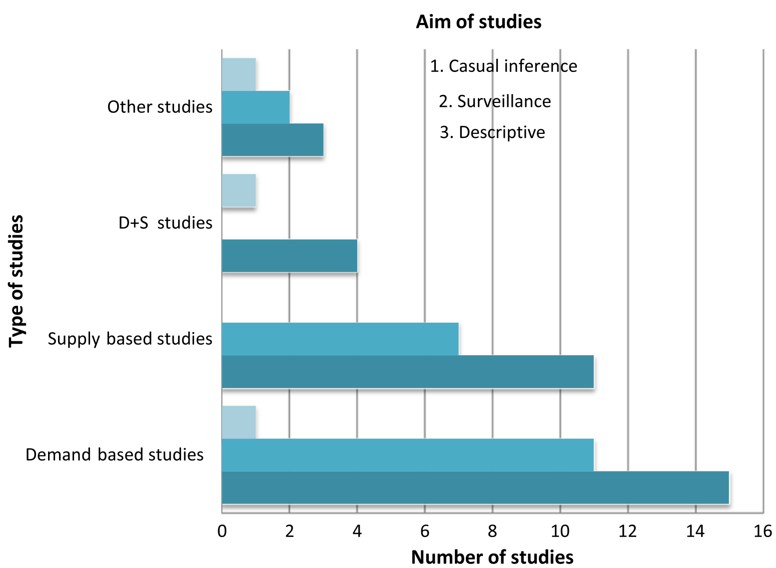

Chart 3 Frequency of studies based on their aim

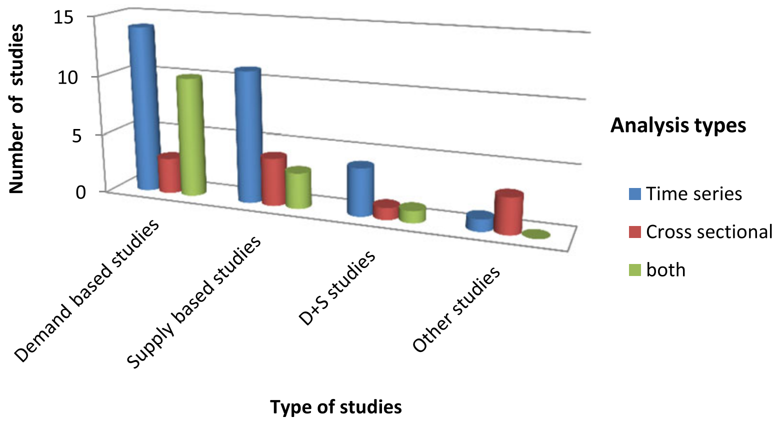

Chart 4 Frequency of studies based on their analysis type

(n=33 (60%)), some papers have a surveillance aim (n=20 (34%)), and the least studies have a casual inference aim (n=3 (6%)) (Chart 3).

According to Table 5, Google Trends (17 articles, 63%) was the most used data source in demand based studies. Although in the supply based studies, Twitter (13 articles, 72%) was the most used data source (Aslam et al., 2014; Burton, Tanner, Giraud-Carrier, West & Barnes, 2012; Chew & Eysenbach, 2010; Fung et al., 2013; Hanson et al., 2013; Kamenova, Reshef & Caulfield, 2014; Katsuki et al., 2015; Kim et al., 2015; Mishori, Oberoi Singh, Levy & Newport, 2014; Nagel et al., 2013; Nascimento et al., 2014; Yin, Fabbri, Rosenbloom & Malin, 2015; Yom-Tov et al., 2015; Zhang et al., 2013).

Even in the demand + supply studies, the most used data sources were Google Trends and Twitter. The analysis types most used in the studies were the time series analysis with a frequency rate of 55% and cross-sectional analysis with 20% frequency rate. Both analysis types were simultaneously used in 25% of the studies (Chart 4).

With respect to the validation, 33% (nine articles) of the demand based studies made comparisons against comparison data sets to validate their output (Bragazzi, Bacigaluppi, Robba, Siri, et al., 2016; Bragazzi, Barberis, et al., 2016; Domnich et al., 2014; Eysenbach, 2006; Mnadla et al., 2016; Ocampo et al., 2013; Siri et al., 2016; Wang, Chen, Yu & Chen, 2015; Yom-Tov & Gabrilovich, 2013).

The examples of the comparison/real world datasets include Canada's national statistical agency website, WHO - Regional Office for South-East Asia website, Ironman official website, Centers for Diseases Control and prevention website, IZSAM G.Caporale Teramo website and the EpiCentro website of the Higher Institute of Health (ISS).

Discussion

The term 'infodemiology' was first used for analysing the demand side of whatever is published in the Web, and then, it was used in the supply side for analysing the people's needs and monitoring their health information seeking behaviour. In this scoping review of infodemiology studies, it was found that the Web (1.0 & 2.0) tools are utilised widely by researchers in different topics. Pelat, Turbelin, Bar-Hen, Flahault and Valleron (2009) demonstrated that the use of search queries for disease detection could be applied to different diseases. Like Bernardo et al. (2013) findings, the most disease commonly evaluated using infodemiology approaches was flu. In accordance with Bernardo et al. (2013), the demand based studies were generated using Web 1.0 tools, whereas Web 2.0 tools were used for the supply based studies.

Although there was an increase in the publication of infodemiology studies over time, both the demand based studies and supply based studies remained stable in 2015 and 2016, respectively. The demand based studies were the most published; however, the supply based studies were cited more often in comparison with the

© 2018 Health Libraries Group

other three groups. A subanalysis of the studies based on some defined variables showed that the median citation rate (34 per article) is more than the average for all the scientific articles (7.64 per article) (Nagar et al., 2014) and other infodemiology studies (Nuti et al., 2014).

The high citation rate of a few studies included in the review could be the reason (Eysenbach, 2006; Wyrwoll, 2014). Almost half the studies were demand based, and the data source most used was Google Trends. Most of the demand based studies were conducted in developed countries using Google Trends; the highest numbers were conducted in Italy.

Until 2009, researchers believed that using Google Trends is effective only in developed countries for doing disease surveillances (Carneiro & Mylonakis, 2009). Around 2009, some researchers from South-East Asian countries, Latin America, Russia and China indicated that conducting researches about the search patterns can be considered as a valid and reliable method to perform disease surveillances in developing countries (Ayers et al., 2012; Ritterman, Osborne & Klein, 2009). Although in some developing countries, the scientific interests and skills of the researchers help in overcoming the resource intensive and challenging data collection of infodemiology studies, but regarding our finding, as shown in Table 5, most infodemiology studies were conducted in developed countries.

The results of the present study showed that Twitter was the most used data source in the supply based studies. Most of the supply based studies used Twitter and were performed in developed countries; at the top of them was USA. According to Bernardo et al. (2013), Google Trends and Twitter were the most data sources used, and this is in accordance with our findings. In confirmation with our findings, some data sources like Facebook or news aggregates were used at least (Aramaki, Maskawa & Morita, 2011; Olson, Konty, Paladini, Viboud & Simonsen, 2013).

Some of the demand + supply based studies in this review used Web 1.0, Web 2.0 technologies spontaneously for disease surveillance, but none of them used Web technologies alongside epidemiological data, as used by Sharpe, Hopkins, Cook and Striley (2016) and Woo et al. (2016).

The data collection techniques used in the studies included in this review consisted of aggregating data from Web 1.0, Web 2.0 technologies or epidemiological data. This is in confirmation of Wongkoblap, Vadillo and Curcin (2017); however, they directly collected data from Facebook participants. One reason for this might be that getting data from Facebook requires users' consent and Facebook does not allow access to interactions between users. None of the studies included in this review used a direct data collection technique. It might be, in the case of supply based studies, because Web 2.0 technologies like Twitter provide API's that allow developers to get information about followers and followees.

The Web tools used for conducting infodemiology studies have some limitations. The first limitation of these Web tools, such as Google Trends or Twitter, is that they track only the segment of population that uses and surfs the Web and monitors their health information behaviour (Alicino et al., 2015; Bragazzi, Dini, Toletone, Brigo & Durando, 2016b; Bragazzi, Barberis, et al., 2016; Bragazzi, Watad, et al., 2016; Nuti et al., 2014). However, as mentioned in other studies, the major limitation of these Web tools, especially Google Trends, is the lack of detailed information on the method used for the search and analysis of new big data streams (Alicino et al., 2015; Bragazzi, Barberis, et al., 2016; Bragazzi et al., 2016b; Nuti et al., 2014).

Twitter also has another limitation: it is almost US centric. Over 67 million Twitter users are in the United States (Aslam, 2017). Hence, it poses limitations on the mining of health information from other countries. The ban of the use of Twitter in some developing countries could be another limitation. This issue deprived the affected developing countries from the benefits provided by infodemiology studies.

The first strength point of the study is its scoping review methodology because it is the first attempt to systematically map the published literature on infodemiology studies. The other strength of this scoping review includes a broad search of the literature using multiple databases. Each article was reviewed by two independent reviewers who met at regular intervals to resolve conflicts.

© 2018 Health Libraries Group

The present study also has some limitations. Firstly, although we tried to do a comprehensive search in the databases, it is possible that there were some studies that we could not retrieve and include in the review; limiting our search to English language may have excluded some of the studies in other languages.

The other limitation was the classification of the studies based on their aim and analysis type; in this case, both authors tried to resolve their disagreements by consensus. The fourth limitation is that although more demand and supply based studies in the review were carried out using Google Trends and Twitter, there were other data sources such as online forums, Wikipedia trends, YouTube, search and engine queries, which were used in the studies with a lower frequency rate. We reported these Web tools in Table 4, but it seems that reporting and analysing the benefits and limitations for each of these data sources may cause the scattering of the results. This exceeds the scope of this discussion.

Infodemiology studies are observational in nature and do not involve individual research participants, so the conventional tools fit for assessing their bias cannot be used (Nuti et al., 2014; Stroup et al., 2000; Viswanathan et al., 2012). In this review, 3 kinds of biases were considered. The first bias of this review which was unavoidable was the language bias as the inclusion and exclusion criteria considered only the studies in English language in this review.

The selection bias was the second bias which we tried to address in this study. For avoiding the selection bias, the title and abstract of each identified study were assessed by the first reviewer (KZ) blinded to authors, affiliations and the publishing journal. Then, the second reviewer independently assessed the title and abstracts of the studies.

The third bias of this review was the publication bias as the numbers of studies that reported positive findings were more than those that reported neutral or negative findings. Hence, the present study had a positive publication bias; it was in accordance with the other studies (Nuti et al., 2014). Almost 37% of the demand based surveillance studies validated the Google Trends output against the real world (comparison) data sets. In both the demand based and supply based studies, different search terms and search dates were used, but no rational for these selections was provided to further understand the search method and increase the face validity of the review; this was in line with the findings of Nuti et al., (2014).

For enhancing the transparency of such infodemiology studies, researchers can get a screenshot from the raw data. The corporations responsible for Web tools such as Google Inc. can provide some precise instructions and guidelines regarding the capabilities of their tools, their changes over time and the standard methodologies for conducting infodemiology studies. Some cooperation between the researchers at the universities and Web tools corporations for the benchmarking of the best practices could be another useful step in this area.

Although our aim was not to investigate the reproducibility of the studies, but it seems that the search methodologies were not documented completely by the researchers. Having no methodological standards or guidelines for the use of Web tools in infodemiology studies and for a proper reporting of their use may be the reason for the incomplete documentation. It seems that Research-Embedded Health Librarians (REHLs) can best express the role of health information professionals in this regard (Greyson, Surette, Dennett & Chatterley, 2013; Henderson, 2014).

As part of a team of researchers, REHLs worked alongside the infodemiology research team from the inception of the research process by providing not only tailored, intensive information services to infodemiology research teams within which they are integrated but also supported evidence based practice and knowledge syntheses, such as systematic reviews and scoping reviews on infodemiology studies; conducting scoping reviews on infodemiology will result to identifying potential research gaps and conducting systematic reviews will result to summing up the best available research on infodemiology methodologies. Therefore, methodological standards or guidelines for the use of Web tools in the infodemiology studies will be produced.

A more proactive role might be conducting infodemiology studies by health information

© 2018 Health Libraries Group

specialists themselves; therefore, opportunities must be available for health librarians to structure individual training efforts to develop new knowledge, skills and expertise in taking on infodemiology studies. It is important also for health information specialists to set up personal goals and a continuing professional development plan to gain new knowledge, skills and abilities to take on infodemiology studies (Lawton & Burns, 2015).

Therefore, they should consider a baseline of competency areas such as data mining when drawing up personal professional development plans. Health information specialists in emerging roles need to keep their skills up to date to remain competitive and reply to the changing requirements of their profession. It is worth mentioning that using infodemiology approaches for mining peer generated contents on social media, being aware of peers' information needs in real time and replying to their information needs with evidence based, reliable and appropriate information will cause peers' professional empowerment and profession empowerment.

New opportunities and challenges are emerging for health information specialists as informationists for providing evidence based and reliable health information in real time in infodemiology studies too. As health information has the ability to influence health outcomes (Coulter & Ellins, 2006; Coulter et al., 2006; Keselman, Browne & Kaufman, 2008; Zhao & Zhang, 2017), the delivery of high quality and appropriately targeted consumer health information is central to any achievement of health literacy. The health literacy of the population will promote understanding, decision-making and the application of knowledge and health advocacy by health information specialists (Smith & Duman, 2009).

According to international trends of health science librarianship in some developed, developing and countries in transition (Lappa et al., 2012; Murphy & Jargin, 2017; Wales, Bruch, Foster, Gorman & Peters, 2014; Xie, Chan, Lam & Chiu, 2014), formal education courses in health information science do not focus on data mining; therefore, fundamental revision of the curricula seems to be required to include the baseline competencies for conducting infodemiology studies.

On the other hand, professional library associations representing health librarians need to hold up data mining courses for those interested graduates and develop education policies that include specific competencies for conducting infodemiology studies. These associations should consider collaborating internationally to formulate education policies and standards tailored specifically to conduct infodemiology studies by health information specialists (Viswanathan et al., 2012).

Conducting infodemiology studies using the other Web tools, trying to run more studies using the Web 2.0 tools in developing countries, conducting practical workshops on data mining, text mining and infodemiology are suggestions for the future.

Conclusion

To our knowledge, this is the first study to investigate infodemiology studies from the inception of the infodemiology concept back in 2002. The initial infodemiology studies pertain to the quality assessment of the hospital's websites. Most studies belong to developed countries, based on flu, and published in the Journal of Medical Internet Research. The publication rate of demand based studies was 3–2, in comparison with supply based studies. The most used data source in the demand based and the supply based studies were Google Trends and Twitter, respectively.

Although the contribution of developing countries to infodemiology studies has been negligible, this methodology can be a potential for these countries; Web technologies have provided new challenges and opportunities for health information specialists for analysing UGC by data mining methods in real time, providing evidence based, reliable and appropriate heath information for health information consumers and health literacy promotion.

Acknowledgements

This study is a part of a PhD thesis supported by Iran University of Medical Sciences (IUMS/SHMIS-2015/16). The authors would like to thank all of the participants.

© 2018 Health Libraries Group

Conflict of interest

The authors declared no conflict of interest.

Source of funding

Although our work is a part of PhD thesis which approved by Iran University of Medical Sciences, but we did not receive any funding from the university.

References

Alicino, C., Bragazzi, N. L., Faccio, V., Amicizia, D., Panatto, D., Gasparini, R., … Orsi, A. (2015). Assessing Ebola-related web search behavior: Insights and implications from an analytical study of Google Trends-based query volumes. *Infectious Diseases of Poverty*, *4*, 54. https://doi.org/10.1186/s40249-015-0090-9

Aramaki, E., Maskawa, S., & Morita, M. (2011). Twitter catches the flu: detecting influenza epidemics using Twitter. *Proceedings of the 2011 Conference on Empirical Natural Language* (pp. 1568–1576). Edinburgh, UK: Association for Computational Linguistics. Available from: http://www.aclweb.org/anthology/D11-1145 [last accessed 20 July 2017].

Aslam, S. (2017). Twitter by the numbers: Stats, demographics & fun facts. Available from: https://www.omnicoreagency.com/twitter-statistics/ [last accessed 20 July 2017].

Aslam, A. A., Tsou, M. H., Spitzberg, B. H., An, L., Gawron, J. M., Gupta, D. K., … Lindsay, S. (2014). The reliability of tweets as a supplementary method of seasonal influenza surveillance. *Journal of Medical Internet Research*, *16*, e250. https://doi.org/10.2196/jmir.3532

Ayers, J. W., Althouse, B. M., Allem, J. P., Ford, D. E., Ribisl, K. M., & Cohen, J. E. (2012). A novel evaluation of World No Tobacco day in Latin America. *Journal of Medical Internet Research*, *14*, e77. https://doi.org/10.2196/jmir.2148

Bernardo, T. M., Rajic, A., Young, I., Robiadek, K., Pham, M. T., & Funk, J. A. (2013). Scoping review on search queries and social media for disease surveillance: A chronology of innovation. *Journal of Medical Internet Research*, *15*, e147. https://doi.org/10.2196/jmir.2740

Bragazzi, N. L. (2013). Infodemiology and infoveillance of multiple sclerosis in Italy. *Multiple Sclerosis International*, *2013*, 924029. https://doi.org/10.1155/2013/924029

Bragazzi, N. L. (2014). Google Trends-based approach for monitoring NSSI. *Journal of Psychology Research & Behavior Management*, *7*, 1–8. https://doi.org/10.2147/PRBM.S44084

Bragazzi, N. L., Bacigaluppi, S., Robba, C., Nardone, R., Trinka, E., & Brigo, F. (2016). Infodemiology of status epilepticus: A systematic validation of the Google Trends-based search queries. *Epilepsy & Behavior*, *55*, 120–123. https://doi.org/10.1016/j.yebeh.2015.12.017

Bragazzi, N. L., Bacigaluppi, S., Robba, C., Siri, A., Canepa, G., & Brigo, F. (2016). Infodemiological data of West-Nile virus disease in Italy in the study period 2004–2015. *Data in Brief*, *2*, 839–845. https://doi.org/10.1016/j.dib.2016.10.022

Bragazzi, N. L., Barberis, I., Rosselli, R., Gianfredi, V., Nucci, D., Moretti, M., … Martini, M. (2016). How often people Google for vaccination: Qualitative and quantitative insights from a systematic search of the web-based activities using Google Trends. *Human Vaccines & Immunotherapeutics*, *13*, 464–469. https://doi.org/10.1016/j.dib.2016.10.022

Bragazzi, N. L., Dini, G., Toletone, A., Brigo, F., & Durando, P. (2016a). Infodemiological data concerning silicosis in the USA in the period 2004–2010 correlating with real-world statistical data. *Data in Brief*, *10*, 457–464. https://doi.org/10.1016/j.dib.2016.11.021

Bragazzi, N. L., Dini, G., Toletone, A., Brigo, F., & Durando, P. (2016b). Leveraging big data for exploring occupational diseases-related interest at the level of scientific community, media coverage and novel data streams: The example of silicosis as a pilot study. *PLoS ONE*, *11*, e0166051. https://doi.org/10.1371/journal.pone.0166051

Bragazzi, N. L., Watad, A., Brigo, F., Adawi, M., Amital, H., & Shoenfeld, Y. (2016). Public health awareness of autoimmune diseases after the death of a celebrity. *Clinical Rheumatology*, *36*, 1911–1917. https://doi.org/10.1007/s10067-016-3513-5

Brigo, F., & Erro, R. (2016). Why do people Google movement disorders? An infodemiological study of information seeking behaviors. *Neurological Sciences*, *37*, 781–787. https://doi.org/10.1007/s10072-016-2501-5

Brigo, F., Lochner, P., Tezzon, F., & Nardone, R. (2014). Web search behavior for multiple sclerosis: An infodemiological study. *Multiple Sclerosis and Related Disorders*, *3*, 440–443. https://doi.org/10.1016/j.msard.2014.02.005

Brigo, F., Otte, W. M., Igwe, S. C., Ausserer, H., Nardone, R., Tezzon, F., … Trinka, E. (2015). Information-seeking behavior for epilepsy: An infodemiological study of searches for Wikipedia articles. *Epileptic Disorders*, *17*, 460–466. https://doi.org/10.1684/epd.2015.0772

Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandl, K. D. (2008). Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the Health Map project. *PLOS Medicine*, *5*, e151. https://doi.org/10.1371/journal.pmed.0050151

Bruns, A. (2016). User-generated content. In K. Bruhn Jensen, R. T. Craig, J. D. Pooley & E. W. Rothenbuhler (Eds.), *The international encyclopedia of communication theory & philosophy* (1st edn, pp. 1–5). London: John Wiley & Sons https://doi.org/10.1002/9781118766804.wbiect085

Burton, S. H., Tanner, K. W., Giraud-Carrier, C. G., West, J. H., & Barnes, M. D. (2012). "Right Time, Right Place" health communication on Twitter: Value and accuracy of location information. *Journal of Medical Internet Research*, *14*, e156. https://doi.org/10.2196/jmir.2121

Carneiro, H. A., & Mylonakis, E. (2009). Google Trends: A web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases*, *49*, 1557–1564. https://doi.org/10.1086/630200

© 2018 Health Libraries Group

Casey, S. (2016). Nielsen social media report. Available from: http://www.nielsen.com/content/dam/corporate/us/en/reports-downloads/2017-reports/2016-nielsen-social-media-report.pdf [last accessed 20 July 2017].

Chan, K. L., Ho, S. Y., & Lam, T. H. (2013). Infodemiology of alcohol use in Hong Kong mentioned on blogs: Infoveillance study. *Journal of Medical Internet Research*, *15*, e192. https://doi.org/10.2196/jmir.2180

Chan, B., Lopez, A., & Sarkar, U. (2015). The canary in the coal mine tweets: Social media reveals public perceptions of non-medical use of opioids. *PLoS ONE*, *10*, e0135072. https://doi.org/10.1371/journal.pone.0135072

Chen, P., Chai, J., Zhang, L., & Wang, D. (2014). Development and application of a Chinese webpage Suicide Information Mining System (Sims). *Journal of Medical Systems*, *38*, 88. https://doi.org/10.1007/s10916-014-0088-z

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PLoS ONE*, *5*, e14118. https://doi.org/10.1371/journal.pone.0014118

Cole-Lewis, H., Pugatch, J., Sanders, A., Varghese, A., Posada, S., Yun, C., ... Augustson, E. (2015). Social listening: A content analysis of e-cigarette discussions on Twitter. *Journal of Medical Internet Research*, *17*, e243. https://doi.org/10.2196/jmir.4969

Collier, N., Son, N. T., & Nguyen, N. M. (2011). OMG U got flu? Analysis of shared health messages for bio-surveillance *Journal of Biomedical Semantics*, *2*(Suppl. 5), S9. https://doi.org/10.1186/2041-1480-2-s5-s9

Coulter, A., & Ellins, J. (2006). Patient-focused interventions: A review of evidence. Available from Picker Institute Europe website: http://www.health.org.uk/sites/health/files/PatientFocusedInterventions_ReviewOfTheEvidence.pdf [last accessed 20 July 2017].

Coulter, A., Ellins, J., Swain, D., Clarke, A., Heron, P., Rasul, F., ... Sheldon, H. (2006). Assessing the quality of information to support people in making decisions about their health and healthcare. Available from Picker Institute Europe website: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.4236&rep=rep1&type=pdf [last accessed 20 July 2017].

Culotta, A. (2010). Towards detecting influenza epidemics by analyzing Twitter messages. *Proceedings of the First Workshop on Social Media Analytics (SOMA)* (pp. 115–122). Washington, D.C.: Association for Computing Machinery. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.167.5375&rep=rep1&type=pdf [last accessed 20 July 2017].

Curtis, B., Alanis-Hirsch, K., Kaynak, O., McLellan, A. T., & Meyers, K. (2015). Using web searches to track interest in Synthetic Cannabinoids (aka 'herbal incense'). *Drug & Alcohol Review*, *34*, 105–108. https://doi.org/10.1111/dar.12189

Domnich, A., Arbuzova, E. K., Signori, A., Amicizia, D., Panatto, D., & Gasparini, R. (2014). Demand-based web surveillance of sexually transmitted infections in Russia. *International Journal of Public Health*, *59*, 841–849. https://doi.org/10.1007/s00038-014-0581-7

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ... Seligman, M. E. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, *26*, 159–169. https://doi.org/10.1177/0956797614557867

Eysenbach, G. (2002). Infodemiology: The epidemiology of (mis) information. *American Journal of Medicine*, *113*, 763–765. https://doi.org/10.1016/S0002-9343(02)01473-0

Eysenbach, G. (2006). Infodemiology: Tracking flu-related searches on the web for syndromic surveillance. *AMIA Annual Symposium Proceedings* (pp. 244–248).

Eysenbach, G. (2008). Medicine 2.0: Social networking, collaboration, participation, Apo mediation, and openness. *Journal of Medical Internet Research*, *10*, e22. https://doi.org/10.2196/jmir.1030

Eysenbach, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *Journal of Medical Internet Research*, *11*, e11. https://doi.org/10.2196/jmir.1157

Fung, I. C., Fu, K. W., Ying, Y., Schaible, B., Hao, Y., Chan, C. H., & Tse, Z. T. (2013). Chinese social media reaction to the MERS-CoV and avian influenza A (H7N9) outbreaks. *Infectious Diseases of Poverty*, *2*, 31. https://doi.org/10.1186/2049-9957-2-31

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*, 1012–1014. https://doi.org/10.1038/nature07634

Greyson, D., Surette, S., Dennett, L., & Chatterley, T. (2013). "You're just one of the group when you're embedded": Report from a mixed-method investigation of the research-embedded health. *Journal of the Medical Library Association*, *101*, 287–297. https://doi.org/10.3163/1536-5050.101.4.010

Guy, S., Ratzki-Leewing, A., Bahati, R., & Gwadry-Sridhar, F. (2012). Social media: A systematic review to understand the evidence and application in infodemiology. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, *91*, 1–8. https://doi.org/10.1007/978-3-642-29262-0_1

Hanson, C. L., Cannon, B., Burton, S., & Giraud-Carrier, C. (2013). An exploration of social circles and prescription drug abuse through Twitter. *Journal of Medical Internet Research*, *15*, e189. https://doi.org/10.2196/jmir.2741

Heaivilin, N., Gerbert, B., Page, J. E., & Gibbs, J. L. (2011). Public health surveillance of dental pain via Twitter. *Journal of Dental Research*, *90*, 1047–1051. https://doi.org/10.1177/0022034511415273

Henderson, M. E. (2014). New roles and new horizons for health sciences librarians and libraries. In M. S. Wood (Ed.), *Health sciences librarianship* (pp. 405–420). Chicago: Rowman & Littlefield and the Medical Library Association. Available from: https://scholarscompass.vcu.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1021&context=libraries_pubs [last accessed 20 July 2017].

© 2018 Health Libraries Group

Hill, S., Mao, J., Ungar, L., Hennessy, S., Leonard, C. E., & Holmes, J. (2011). Natural supplements for H1N1 influenza: Retrospective observational infodemiology study of information and search activity on the Internet. *Journal of Medical Internet Research*, *13*, e36. https://doi.org/10.2196/jmir.1722

Hua, M., Alfi, M., & Talbot, P. (2013). Health-related effects reported by electronic cigarette users in online forums. *Journal of Medical Internet Research*, *15*, e59. https://doi.org/10.2196/jmir.2324

Jacobs, W., Amuta, A. O., & Jeon, K. C. (2017). Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. *Cogent Social Sciences*, *3*, 1302785. https://doi.org/10.1080/23311886.2017.1302785

Kamenova, K., Reshef, A., & Caulfield, T. (2014). Representations of stem cell clinics on Twitter. *Stem Cell Reviews and Reports*, *10*, 753–760. https://doi.org/10.1007/s12015-014-9534-z

Katsuki, T., Mackey, T. K., & Cuomo, R. (2015). Establishing a link between prescription drug abuse and illicit online pharmacies: Analysis of Twitter data. *Journal of Medical Internet Research*, *17*, e280. https://doi.org/10.2196/jmir.5144

Keselman, A., Browne, A. C., & Kaufman, D. R. (2008). Consumer health information seeking as hypothesis testing. *Journal of the American Medical Informatics Association*, *15*, 484–495. https://doi.org/10.1197/jamia.M2449

Kim, A. E., Hopper, T., Simpson, S., Nonnemaker, J., Lieberman, A. J., Hansen, H., ... Porter, L. (2015). Using Twitter data to gain insights into E-cigarette marketing and locations of use: An infoveillance study. *Journal of Medical Internet Research*, *17*, e251. https://doi.org/10.2196/jmir.4466

Kim, M., Jung, Y., Jung, D., & Hur, C. (2014). Investigating the congruence of crowd sourced information with official government data: The case of pediatric clinics. *Journal of Medical Internet Research*, *16*, e29. https://doi.org/10.2196/jmir.3078

Kind, T., Wheeler, K. L., Robinson, B., & Cabana, M. D. (2004). Do the leading children's hospitals have quality web sites? A description of children's hospital websites. *Journal of Medical Internet Research*, *6*, e20. https://doi.org/10.2196/jmir.6.2.e20

Koch-Weser, S., Bradshaw, Y. S., Gualtieri, L., & Gallagher, S. S. (2010). The Internet as a health information source: Findings from the 2007 health information national trends survey and implications for health communication. *Journal of Health Communication*, *15*(Suppl. 3), 279–293. https://doi.org/10.1080/10810730.2010.522700

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International World Wide Web* (pp. 591–600). Raleigh, NC: Association for Computing Machinery. Available from: http://www.ambuehler.ethz.ch/CDstore/www2010/www/p591.pdf [last accessed 20 July 2017].

Lampos, V., Yom-Tov, E., Pebody, R., & Cox, I. J. (2015). Assessing the impact of a health intervention via user-generated Internet content. *Data Mining and Knowledge Discovery*, *29*, 1434–1457. https://doi.org/10.1007/s10618-015-0427-9

Lappa, E., Chaleplioglou, A., Cognetti, G., Della Seta, M., Napolitani Cheyne, F., Juan-Quilis, V., ... Murphy, J. (2012). International trends in health science librarianship: Part 4 – Four Southern European countries. *Health Information & Libraries Journal*, *29*, 338–343. https://doi.org/10.1111/j.1471-1842.2012.01000.x

Lawton, A., & Burns, J. (2015). A review of competencies needed for health librarians – A comparison of Irish and international practice. *Health Information & Libraries Journal*, *32*, 84–94. https://doi.org/10.1111/hir.12093

Levac, D., Colquhoun, H., & O'Brien, K. K. (2010). Scoping studies: Advancing the methodology. *Implementation Science*, *5*, 69. https://doi.org/10.1186/1748-5908-5-69

Liang, B., & Scammon, D. L. (2013). Incidence of online health information search: A useful proxy for public health risk perception. *Journal of Medical Internet Research*, *15*, e114. https://doi.org/10.2196/jmir.2401

Liu, X., Bao, Z., Liu, H., & Wang, Z. (2011). The quality and characteristics of leading general hospitals' websites in China. *Journal of Medical Systems*, *35*, 1533–1562. https://doi.org/10.1007/s10916-010-9432-0

McCully, S. N., Don, B. P., & Updegraff, J. A. (2013). Using the Internet to help with diet, weight, and physical activity: Results from the Health Information National Trends Survey (HINTS). *Journal of Medical Internet Research*, *15*, e148. https://doi.org/10.2196/jmir.2612

Mishori, R., Oberoi Singh, L., Levy, B., & Newport, C. (2014). Mapping physician Twitter networks: Describing how they work as a first step in understanding connectivity, information flow, and message diffusion. *Journal of Medical Internet Research*, *16*, e107. https://doi.org/10.2196/jmir.3006

Mnadla, S., Bragazzi, N. L., Rouissi, M., Chaalali, A., Siri, A., Padulo, J., ... Knechtle, B. (2016). Infodemiological data of Ironman Triathlon in the study period 2004–2013. *Data in Brief*, *9*, 123–127. https://doi.org/10.1016/j.dib.2016.08.040

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*, e1000097. https://doi.org/10.1371/journal.pmed1000097

Murphy, J., & Jargin, S. (2017). International trends in health science librarianship part 20: Russian. *Health Information & Libraries Journal*, *34*, 92–94. https://doi.org/10.1111/hir.12172

Myslín, M., Zhu, S., Chapman, W., & Conway, M. (2013). Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. *Journal of Medical Internet Research*, *15*, e174. https://doi.org/10.2196/jmir.2534

Nagar, R., Yuan, Q., Freifeld, C. C., Santillana, M., Nojima, A., Chunara, R., & Brownstein, J. S. (2014). A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *Journal of Medical Internet Research*, *16*, e236. https://doi.org/10.2196/jmir.3416

© 2018 Health Libraries Group

Nagel, A. C., Tsou, M. H., Spitzberg, B. H., An, L., Gawron, J. M., Gupta, D. K., … Sawyer, M. H. (2013). The complex relationship of real space events and messages in cyberspace: Case study of influenza and pertussis using tweets. *Journal of Medical Internet Research*, *15*, e237. https://doi.org/10.2196/jmir.2705

Nakada, H., Yuji, K., Tsubokura, M., Ohsawa, Y., & Kami, M. (2014). Development of a national agreement on human papillomavirus vaccination in Japan: An infodemiology study. *Journal of Medical Internet Research*, *16*, e129. https://doi.org/10.2196/jmir.2846

Nascimento, T. D., DosSantos, M. F., Danciu, T., DeBoer, M., van Holsbeeck, H., Lucas, S. R., … DaSilva, A. F. (2014). Real-time sharing and expression of migraine headache. *Journal of Medical Internet Research*, *16*, e96. https://doi.org/10.2196/jmir.3265

Neiburger, E. (2010). User-generated content. *ALA Tech Source*, 8. Available from: https://journals.ala.org/index.php/ltr/article/view/4716/5617 [last accessed 20 July 2017].

Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The use of Google Trends in health care research: A systematic review. *PLoS ONE*, *9*, e109583. https://doi.org/10.1371/journal.pone.0109583

Ocampo, A. J., Chunara, R., & Brownstein, J. S. (2013). Using search queries for Malaria surveillance, Thailand. *Malaria Journal*, *12*, 390. https://doi.org/10.1186/1475-2875-12-390

Olson, D. R., Konty, K. J., Paladini, M., Viboud, C., & Simonsen, L. (2013). Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: A comparative epidemiological study at three geographic scales. *PLoS Computational Biology*, *9*, e1003256. https://doi.org/10.1371/journal.pcbi.1003256

Paul, M. J., & Drezde, M. (2011). You are what you Tweet: Analyzing Twitter for public health. *Proceeding of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 265–272). Barcelona, Spain: AAAI Publications. Available from: http://www.cs.jhu.edu/~mpaul/files/2011.icwsm.twitter_health.pdf [last accessed 20 July 2017].

Pelat, C., Turbelin, C., Bar-Hen, A., Flahault, A., & Valleron, A. (2009). More diseases tracked by using Google Trends. *Emerging Infectious Diseases*, *15*, 1327–1328. https://doi.org/10.3201/eid1508.090299

Pham, M. T., Rajić, A., Greig, J. D., Sargeant, J. M., Papadopoulos, A., & McEwen, S. A. (2014). A scoping review of scoping reviews: Advancing the approach and enhancing the consistency. *Research Synthesis Methods*, *5*, 371–385. https://doi.org/10.1002/jrsm.1123

Priest, C., Knopf, A., Groves, D., Carpenter, J. S., Furrey, C., Krishnan, A., … Wilson, J. (2016). Finding the patient's voice using big data: Analysis of users' health-related concerns in the ChaCha question-and-answer service (2009–2012*). Journal of Medical Internet Research*, *18*, e44. https://doi.org/10.2196/jmir.5033

Ritterman, J., Osborne, M., & Klein, E. (2009). Using prediction markets and Twitter to predict a swine flu pandemic. *Proceedings of the 1st International Workshop on Mining Social Media* (pp. 9–17). Seville, Spain: Spanish Association for Artificial Intelligence. Available from: http://homepages.inf.ed.ac.uk/miles/papers/swine09.pdf [last accessed 20 July 2017].

Rutsaert, P., Pieniak, Z., Regan, A., McConnon, A., & Verbeke, W. (2013). Consumer interest in receiving information through social media about the risks of pesticide residues. *Food Control*, *34*, 386–392. https://doi.org/10.1016/j.foodcont.2013.04.030

Sanders-Jackson, A., Brown, C. G., & Prochaska, J. J. (2015). Applying linguistic methods to understanding smoking-related conversations on Twitter. *Tobacco Control*, *24*, 136–138. https://doi.org/10.1136/tobaccocontrol-2013-051243

Santos, J. C., & Matos, S. (2014). Analyzing Twitter and web queries for flu trend prediction. *Theoretical Biology & Medical Modelling*, *11*(Suppl. 1), S6. https://doi.org/10.1186/1742-4682-11-S1-S6

Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, *38*, 182–188. https://doi.org/10.1016/j.ajic.2009.11.004

Sharpe, J. D., Hopkins, R. S., Cook, R. L., & Striley, C. W. (2016). Evaluating Google, Twitter, and Wikipedia as tools for influenza surveillance using Bayesian Change Point Analysis: A comparative analysis. *JMIR Public Health and Surveillance*, *2*, e161. https://doi.org/10.2196/publichealth.5901

Showghi, N. N., & Williams, A. C. (2012). Information about male chronic pelvic and urogenital pain on the Internet: An evaluation of Internet resources. *Pain Medicine*, *13*, 1275–1283. https://doi.org/10.1111/j.1526-4637.2012.01466.x

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PLoS ONE*, *6*, e19467. https://doi.org/10.1371/journal.pone.0019467

Siri, A., Khabbache, H., Al-Jafar, A., Martini, M., Brigo, F., & Bragazzi, N. L. (2016). Infodemiological data of high-school drop-out related web searches in Canada correlating with real-world statistical data in the period 2004–2012. *Data in Brief*, *9*, 679–684. https://doi.org/10.1016/j.dib.2016.09.032

Smith, S., & Duman, M. (2009). The state of consumer health information: An overview. *Health Information & Libraries Journal*, *26*, 260–278. https://doi.org/10.1111/j.1471-1842.2009.00870.x

Solano, P., Ustulin, M., Pizzorno, E., Vichi, M., Pompili, M., Serafini, G., & Amore, M. (2016). A Google-based approach for monitoring suicide risk. *Psychiatry Research*, *246*, 581–586. https://doi.org/10.1016/j.psychres.2016.10.030

Somaiya, R. (2014). How Facebook is changing the way its users consume journalism. *The New York Times*. Available from: http://www.nytimes.com/2014/10/27/business/media/how-facebook-is-changing-the-way-its-users-consume-journalism.html [last accessed 20 July 2017].

Statista. (2017). Twitter: number of monthly active users 2010–2017. Available from: https://www.statista.com/statistics/

© 2018 Health Libraries Group

282087/number-of-monthly-active-twitter-users/ [last accessed 20 July 2017].

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., … Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA*, *283*, 2008–2012. https://doi.org/10.1001/jama.283.15.2008

The International Telecommunication Union. (2017). ICT facts and figures. Available from: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf [last accessed 20 July 2017].

The Pew Internet Project's Research. (2013). Health online 2013. Available from: http://www.pewinternet.org/2013/01/15/health-online-2013/ [last accessed 20 July 2017].

The Pew Internet Project's Research. (2014). How social media is reshaping news. Available from: http://www.pewresearch.org/fact-tank/2014/09/24/how-social-media-is-reshaping-news/ [last accessed 20 July 2017].

Tilston, N. L., Eames, K. T., Paolotti, D., Ealden, T., & Edmunds, W. J. (2010). Internet-based surveillance of influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic. *BMC Public Health*, *10*, 650. https://doi.org/10.1186/1471-2458-10-650

United Nations. (2014). The World Economic Situation and Prospects (WESP): Country classification. Available from: http://www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf [last accessed 20 July 2017].

Viswanathan, M., Ansari, M. T., Berkman, N. D., Chang, S., Hartling, L., McPheeters, M., … Treadwell, J. R. (2012). Assessing the risk of bias of individual studies in systematic reviews of health care interventions. Agency for Healthcare Research and Quality (US). Available from: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0077771/ [last accessed 20 July 2017].

Wales, A., Bruch, S., Foster, W., Gorman, M., & Peters, J. (2014). International trends in health science librarianship Part 9: The UK – Scotland and Wales. *Health Information & Libraries Journal*, *31*, 79–83. https://doi.org/10.1111/hir.12050

Wang, H. W., Chen, D. R., Yu, H. W., & Chen, Y. M. (2015). Forecasting the incidence of dementia and dementia-related outpatient visits with Google Trends: Evidence from Taiwan. *Journal of Medical Internet Research*, *17*, e264. https://doi.org/10.2196/jmir.4516

Wang, F., Zeng, D., Carley, K. M., & Mao, W. (2007). Social computing: From social informatics to social intelligence. *IEEE Intelligent Systems*, *22*, 79–83. https://doi.org/10.1109/MIS.2007.41

Wong, P. W., Fu, K. W., Yau, R. S., Ma, H. H., Law, Y. W., Chang, S. S., & Yip, P. S. (2013). Accessing suicide-related information on the Internet: A retrospective observational study of search behavior. *Journal of Medical Internet Research*, *15*, e3. https://doi.org/10.2196/jmir.2181

Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2017). Researching mental health disorders in the era of social media: Systematic review. *Journal of Medical Internet Research*, *19*, e228. https://doi.org/10.2196/jmir.7215

Woo, H., Cho, Y., Shim, E., Lee, J. K., Lee, C. G., & Kim, S. H. (2016). Estimating influenza outbreaks using both search engine query data and social media data in South Korea. *Journal of Medical Internet Research*, *18*, e177. https://doi.org/10.2196/jmir.4955

Wyrwoll, C. (2014). *Social media: Fundamentals, models and ranking of User-Generated Content*. New York: Springer Vieweg.

Xie, Z., Chan, J. L., Lam, L. M., & Chiu, T. H. (2014). International trends in health science librarianship Part 10: The Greater China area. *Health Information & Libraries Journal*, *31*, 161–166. https://doi.org/10.1111/hir.12059

Yang, Y. T., Horneffer, M., & DiLisio, N. (2013). Mining social media and web searches for disease detection. *Journal of Public Health Research*, *2*, 17–21. https://doi.org/10.4081/jphr.2013.e4

Yin, Z., Fabbri, D., Rosenbloom, T., & Malin, B. (2015). A scalable framework to detect personal health mentions on Twitter. *Journal of Medical Internet Research*, *17*, e138. https://doi.org/10.2196/jmir.4305

Yom-Tov, E., & Gabrilovich, E. (2013). Post-market drug surveillance without trial costs: Discovery of adverse drug reactions through large-scale analysis of web search queries. *Journal of Medical Internet Research*, *15*, e124. https://doi.org/10.2196/jmir.2614

Yom-Tov, E., Johansson-Cox, I., Lampos, V., & Hayward, A. C. (2015). Estimating the secondary attack rate and serial interval of influenza-like illnesses using social media. *Influenza & other Respiratory Viruses*, *9*, 191–199. https://doi.org/10.1111/irv.12321

Zhang, N., Campo, S., Janz, K. F., Eckler, P., Yang, J., Snetselaar, L. G., & Signorini, A. (2013). Electronic word of mouth on Twitter about physical activity in the United States: Exploratory infodemiology study. *Journal of Medical Internet Research*, *15*, e261. https://doi.org/10.2196/jmir.2870

Zhao, Y., & Zhang, J. (2017). Consumer health information seeking in social media: A literature review. *Health Information & Libraries Journal*, *34*, 268–273. https://doi.org/10.1111/hir.12192

Zheluk, A., Quinn, C., Hercz, D., & Gillespie, J. A. (2013). Internet search patterns of Human Immunodeficiency Virus and the digital divide in the Russian federation: Infoveillance study. *Journal of Medical Internet Research*, *15*, e256. https://doi.org/10.2196/jmir.2936

Zheluk, A., Quinn, C., & Meylakhs, P. (2014). Internet search and Krokodil in the Russian Federation: An infoveillance study. *Journal of Medical Internet Research*, *18*, e212. https://doi.org/10.2196/jmir.3203

Received *1 July 2017*; Accepted *17 March 2018*

© 2018 Health Libraries Group