Medicine®

OPEN

The Source and Credibility of Colorectal Cancer Information on Twitter

SoHyun Park, MS, Heung-Kwon Oh, MD, Gibeom Park, MS, Bongwon Suh, PhD, Woo Kyung Bae, MD, Jin Won Kim, MD, Hyuk Yoon, MD, Duck-Woo Kim, MD, PhD, and Sung-Bum Kang, MD, PhD

Abstract: Despite the rapid penetration of social media in modern life, there has been limited research conducted on whether social media serves as a credible source of health information. In this study, we propose to identify colorectal cancer information on Twitter and assess its informational credibility.

We collected Twitter messages containing colorectal cancer-related keywords, over a 3-month period. A review of sample tweets yielded content and user categorization schemes. The results of the sample analysis were applied to classify all collected tweets and users, using a machine learning technique. The credibility of the information in the sampled tweets was evaluated.

A total of 76,119 tweets were analyzed. Individual users authored the majority of tweets (n = 68,982, 90.6%). They mostly tweeted about news articles/research (n = 16,761, 22.0%) and risk/prevention (n = 14,767, 19.4%). Medical professional users generated only 2.0% of total tweets (n = 1509), and medical institutions rarely tweeted (n = 417, 0.6%). Organizations tended to tweet more about information than did individuals (85.2% vs 63.1%; $P < 0.001$). Credibility analysis of medically relevant sample tweets revealed that most were medically correct (n = 1763, 84.5%). Among those, more frequently retweeted tweets contained more medically correct information than randomly selected tweets (90.7% vs 83.2%; $P < 0.01$).

Our results demonstrate an interest in and an engagement with colorectal cancer information from a large number and variety of users.

Coupled with the Internet's potential to increase social support, Twitter may contribute to enhancing public health and empowering users, when used with proper caution.

(*Medicine* 95(7):e2775)

Abbreviations: API = Application Programming Interface, SNS = social networking services, URL = Universal Resource Locator.

INTRODUCTION

Twitter is a popular microblogging service that people use to talk about their daily activities and to seek or share information.[1] Twitter users can write (''tweet'') about any topic within the 140-character limit, repost messages by other users (''retweet''), and follow others to receive their tweets.[2] The rapid penetration of Twitter, which currently has >500 million users worldwide, has shaped the platform as an alternative means for sharing and seeking health information online.[3] However, health information mediated via social media remains unregulated and varies in quality, accuracy, and readability.[4] Many have reported a concern about the quality of information that patients and consumers may access via the Internet.[5–10] Colorectal cancer is often referred to as the world's third most threatening cancer killer, and patients often seek information online to make well-informed decisions to overcome it.[11] Yet, the organs related to colorectal cancer often cause myths and misconceptions that are refrained from being discussed in public.[12] Twitter, as one of the most vibrant online social media, can be used for an unobtrusive monitoring and discovering of the public awareness of the disease. In this study, we investigate the information sources and attempt to evaluate the credibility of colorectal cancer information in tweet content.

METHODS

This study has examined the following data to analyze the source and credibility of colorectal cancer information on Twitter: Twitter users; tweet content; Universal Resource Locators (URLs) in tweets; and information credibility.

Data Collection

Twitter offers public Application Programming Interface (API) endpoints through which anyone with an authenticated connection may access publicly available Twitter messages (''tweets''). This enables any interested parties to access the data without violating any ethical guidelines, but one needs such a connection to a streaming endpoint to reach a feed of tweets. The Twitter streaming API enables the API's user to receive real-time updates for a large number of users. Events may be streamed for any user who has granted Open Authorization access to the application.[13] We collected tweets containing colorectal cancer keywords for 3 months, from August 1, 2014, to October 31, 2014. The keywords used are ''colorectal,''

"rectal," "colon," and "colonic," each paired with "cancer," "tumor," "tumour," and "neoplasm." In this way, an entire collection of Twitter messages, containing one or more of the keywords, was acquired. The program created to connect to the API and collect the tweet data was scripted in the Python language (Python 2.7.2.).

Data Sampling

We needed to sample the data to prepare human-labeled ground truth to automatically classify the user and content of the collected tweets. Although a random sampling could suffice the need for classifications, we prepared another sample set for credibility analysis. First, a 3% of entire tweets were randomly sampled. This was entitled the "random" sample group. Second, we selected tweets based on the popularity of a tweet (as measured in the number of its being "retweeted"). A subset of 1000 most frequently retweeted tweets was entitled the "popular" sample group. On the basis of the recurrence, the "popular" tweets covered approximately the top 30% of all collected tweets.

User Categorization

To classify users, we collected the profile descriptions of Twitter users who authored the sampled tweets. The user profile descriptions were the only source of information that one may use in a social media analysis, as it is least feasible to contact individual Twitter users for their demographic information. To determine whether a tweet was authored by a medical professional or a medical institution, one of the researchers reviewed the profile description texts for each user account in the sample. Each user was marked as either an individual or an organization. If the description was left blank, that user was categorized as unverifiable.

Individual users were subcategorized into professionals and nonprofessionals. A user is free to enter any text he or she chooses in the user profile, so the length and content of these profiles varied. Because the profile does not stipulate the inclusion of a user's occupation, it was assumed that a user profile including one or more medical professions would be a user account of a certified professional in medicine. A list of 89 medical professional titles obtained from Wikipedia[14] was used to filter out professionals from nonprofessionals.

Organizational users were manually subcategorized via an inductive coding process.[15] Their profile descriptions were exported to a Microsoft Excel spreadsheet and 2 rounds of coding were undertaken to finalize the subcategories. The subcategories used are "news/media," "foundations/communities," "medical institutions," "government affiliates," and "other." Each coded user profile description was reviewed by the authors, and their inter-rater reliability was measured with Cohen κ[16] (κ = 0.78).

Content Categorization

Because we aimed to identify the medical relevance of a tweet's content, a tweet was to be categorized as either "medically relevant" or "medically irrelevant." If a tweet lacked sufficient content to make a determination, it would be categorized as "other." A medically relevant tweet was defined as a tweet that educates the reader about scientific or public information about colorectal cancer. For example, "Ulcerative colitis, Crohn's Disease or other inflammatory bowel diseases are risk factors for colon cancer. #GetScreened" was marked as "medically relevant." On

the contrary, a tweet that said "you just want to give me colon cancer huh," was marked as "medically irrelevant." One of the authors reviewed all tweet content included in the sample, again by inductive coding.[14] All tweets were marked on a Microsoft Excel spreadsheet, then reviewed by a board-certified colorectal surgeon at Seoul National University Bundang Hospital. The inter-rater reliability of the coding results was sufficiently high (κ = 0.89).

A total of 11 categories were finalized; 7 belonged to the "medically relevant" meta-category and the rest to "medically irrelevant." The medically relevant categories are "news/research," "risk/prevention," "symptoms/diagnosis," "treatments/prognosis," "epidemiology," "screening," and "other." The medically irrelevant categories included "chat," "commercial/fundraising," "celebrities," and "other." We confirmed that these categories showed significant correspondence with the result of unsupervised classification of tweet content by the MAchine Learning for LangaugE Toolkit implementation of Latent Dirichlet Allocation, a technique used in a previous Twitter study on breast cancer awareness.[17]

Automatic Classification of Users and Content

The manually coded tweets and users in the sample were used to apply a machine learning technique to categorize all tweet content and users. For this purpose, we used a naïve Bayesian machine learning technique. The naïve Bayesian classifier is considered an effective method for classifying documents.[17] The process of using the classifier is shown in Figure 1. Here, a word is stemmed to prevent unnecessary bias introduced by duplicative text. The categorized sample data serve as the trained data, and the rest are automatically classified by calculating the assumed probabilities of a word in the text. The conditional probability for a particular word's appearance in a category is calculated given a classification from the trained data.[18] The accuracy of the classifier was measured by running a cross-validation check on the coded data in the sample. The accuracy scores for both the tweet content and the users were significantly high: 96.0% for the tweet content and 91.5% for the users.

Identifying Exogenous Sources of Information

Twitter users often use URL shortening services to refer to an exogenous source of content; that is, one that is other than the Twitter website. URL shortening is often necessary because a tweet only allows 140 characters. When shortened, a URL consists of a random set of English and numeric characters, for example http://t.co/sjv9wKi1Rb. We were interested in identifying what kinds of information sources Twitter users use in their tweets. Hence, we unshortened all URLs embedded in our data. The unshortened URLs pointed to their original absolute website addresses, from which we extracted website domains. The frequency of a particular URL domain was calculated to reveal what types of content on colorectal cancer were shared the most by Twitter users.

Evaluating Information Credibility

Four board-certified physicians, specializing in gastroenterology, family medicine, colorectal surgery, and oncology, examined and evaluated the medically relevant tweets in the sampled tweets. After a careful review of the previous literature[5,19] and quality standards in medicine (eg, the DISCERN questionnaire[20]), the physician committee determined the medical correctness of the tweet content. Because few studies
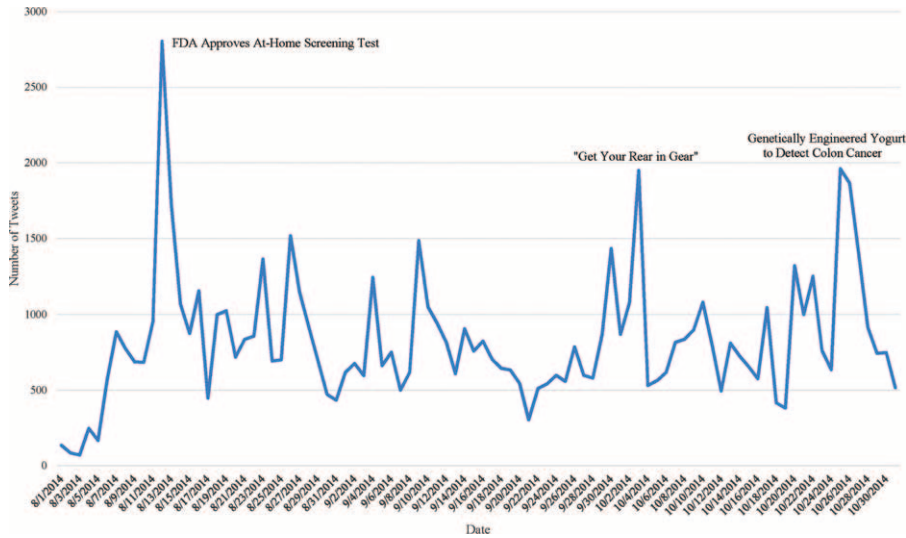
FIGURE 1. A graph showing daily tweet counts for colorectal cancer-related content on Twitter. In mid-August, the flow peaked when the FDA approved an at-home screening test for colorectal cancer. The second peak happened in early October when the ''Get Your Rear in Gear'' marathon was held. The last peak is at the end of October, when news about genetically engineered yogurt that could detect colorectal cancer was announced.

have examined the quality of information in tweets, the classification scheme of the recent Ebola-related tweet study[10] was adopted. All members of the committee reviewed each tweet and agreed upon the category to which the tweet belonged: ''medically correct,'' ''medical misinformation,'' and ''unverifiable.''

For the medically relevant tweets in the sample data, each of physicians was assigned a quarter share. The physicians, using their medical expertise, determined medical correctness of each tweet in their share if a quick judgment was possible. For the undecided tweets, they cross-examined the information against published peer-reviewed journals, preferably of reputation. For the tweets that accompanied no academic evidence, the physicians gathered as a committee and judged upon its verifiability.

Statistical Analysis

A comparison between categorical variables was made using the χ^2 test or Fisher exact test, as appropriate. All reported *P* values are 2-tailed, with a *P* value of 0.05 indicating statistical significance. Analyses were conducted using Statistical Package for Social Sciences version 18.0 (IBM Inc, Armonk, NY).

RESULTS

A total of 76,119 tweets authored by 43,365 unique users were collected and analyzed. An average of 827.4 tweets related to colorectal cancer were authored per day (Figure 2). Tweets peaked on certain days when there were important events or news related to the topic, such as the Food and Drug Administration giving approval of an at-home screening test, or the ''Get Your Rear in Gear'' marathon by the Colon Cancer Coalition.[12]

User Analysis

Table 1 shows a comparison of the contents of tweets by individual and organizational users. A total of 68,982 tweets (90.6%) were produced by individual users, whereas Twitter activity by organizational users was minute (n = 7137, 9.4%). Among tweets by individual users, tweets produced by medical professionals only comprised 2.0% (n = 1509), as shown in Table 2. In terms of content, professional users tweeted a little more about treatments/prognosis than symptoms/diagnosis. Table 3 lists tweets by organizational subcategories. News/media user accounts showed the most vibrant tweeting activities (n = 4913, 68.8%). Foundations/communities such as local
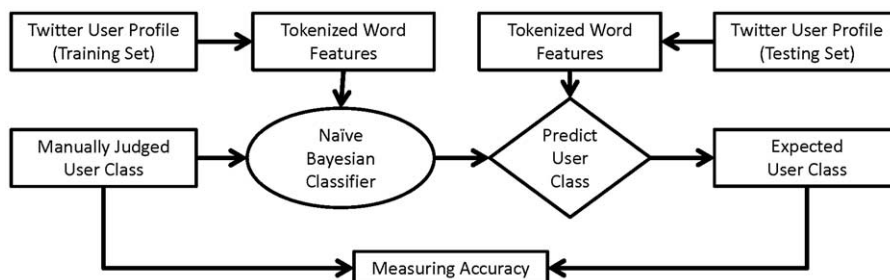

FIGURE 2. A naïve Bayesian framework for automatic classification of users and content.

TABLE 1. Comparison of Tweet Contents by Individual and Organizational Users

Content/Users	Individuals N (%)	Organizations N (%)	Total N (%)
Medically relevant	43,552 (63.1)	6079 (85.2)	49,631 (65.2)
Medically irrelevant	25,430 (36.9)	1058 (14.8)	26,488 (34.8)
Total	68,982	7137	76,119 (100.0)

The difference in the amount of medically relevant tweets between the 2 user categories was statistically significant ($P < 0.001$).

TABLE 2. Individual Twitter Users and Their Tweet Content

All Individuals

		Nonprofessionals		Professionals	
		N	%	N	%
Medically relevant	News/research	16,297	24.1	464	30.7
	Risk/prevention	14,401	21.3	366	24.3
	Symptoms/diagnosis	4187	6.2	118	7.8
	Treatments/prognosis	2697	4.0	175	11.6
	Epidemiology	2785	4.1	52	3.4
	Screening	1897	2.8	91	6.0
	Other	21	0.0	1	0.1
Medically irrelevant	Chat	13,689	20.2	155	10.3
	Commercial/fundraising	4903	7.2	25	1.7
	Celebrities	4279	6.3	45	3.0
	Other	2317	3.4	17	1.1
Total		67,473	100	1509	100

Individual users were subdivided into medical professionals and nonprofessionals.

TABLE 3. Organizational Twitter Users and Their Tweet Content Categories

All Organizations

		News/Media		Foundations/ Communities		Medical Institutions		Others	
		N	%	N	%	N	%	N	%
Medically relevant	News/research	2174	44.2	135	13.5	133	31.9	150	18.6
	Risk/prevention	1092	22.2	227	22.7	88	21.1	195	24.2
	Symptoms/diagnosis	401	8.2	43	4.3	28	6.7	43	5.3
	Treatments/prognosis	651	13.3	9	0.9	72	17.3	103	12.8
	Epidemiology	80	1.6	98	9.8	12	2.9	5	0.6
	Screening	138	2.8	101	10.1	69	16.5	30	3.7
	Other	0	0.0	1	0.1	0	0.0	1	0.1
Medically irrelevant	Chat	94	1.9	246	24.6	13	3.1	57	7.1
	Commercial/fundraising	27	0.5	106	10.6	2	0.5	205	25.4
	Celebrities	249	5.1	33	3.3	0	0.0	16	2.0
	Other	7	0.1	1	0.1	0	0.0	2	0.2
Total		4913	100	1000	100	417	100	807	100

Organization users were subdivided into 4 different categories.

TABLE 4. Top 10 URLs Shared Most Frequently by Users in Tweets With URL Links

	Domain of URL	Domain Type	Tweet Count N	%
1	sqqr.us	News/magazine	1558	2.1
2	expatinc.com	Information	1496	2.0
3	webmd.com	Information	1280	1.7
4	ecannabis.com	Shopping	1208	1.6
5	sunoticiero.com	News/magazine	1089	1.4
6	facebook.com	SNS	1060	1.4
7	doctorozadvices.com	SNS	999	1.3
8	ccalliance.org	Organization	953	1.3
9	indiegogo.com	Crowdfunding	870	1.1
10	stopcoloncancernow.com	Organization	814	1.1

The top 50 most frequently used website domains comprised approximately 70% of all tweets with URLs. News and medical information resources made up to 30%, and other websites included social networking sites and colorectal cancer-related organizations.

colon cancer surgeon groups followed next (n = 1000, 1.3%). Medical or research institutions comprised only 0.6% of total tweets (n = 417).

Content Analysis

As shown in Table 1, more than half of the tweets were classified as medically relevant (n = 49,631, 65.2%). When tracing the origins of medically relevant content, we find that organizations tend to produce much more medically relevant tweets about colorectal cancer than individuals, despite the far smaller volume of tweets from organizations. The difference in the amount of medically relevant tweets between the 2 user categories was statistically significant (85.2% vs 63.1%; $P < 0.001$).

Table 2 and Table 3 list both medically relevant and medically irrelevant tweets and their subcategories for individuals and organizations. News articles/research findings were tweeted the most by both individual and organizational users. Risk factors/prevention measures followed, comprising approximately 20% of the total tweets (n = 16,369). Although the rest of the subcategories had comparable tweet shares, a different trend between individuals and organizations was observed. Individual users tweeted more about symptoms/diagnosis than treatments/prognosis by 1.9%, but this was reversed

for organizational users by 0.4%. In addition, for epidemiology and screening, individual users seemed to be more interested in epidemiology (n = 2837, 3.7%) than screening (n = 1988, 2.6%). This trend was also reversed for organizational users, who tweeted slightly more about screening than epidemiology.

Frequently Shared URLs in Tweets

Of 76,119 tweets, 56,555 tweets (74.3%) had one URL or more embedded in them. Table 4 shows the top 10 domains. News websites and medical resource websites had the most shares on Twitter. For example, ExpatInc.com, a medical information resources website, was one of the most frequently referred to sources of information in our data (n = 1496, 2%). This result corresponded with the aforementioned finding that most medically relevant tweets were related to disease-related news articles and research. Other websites often linked to in the sampled tweets included social networking sites and colorectal cancer-related organizations.

Information Credibility Analysis

We identified 2086 medically relevant tweets (63.4%) from our collected sample data (n = 3290). Table 5 shows that almost 84.5% (n = 1763) of the reviewed medically relevant tweets in the sample contain medically correct information, whereas medical misinformation comprised only 1.6% (n = 33). The medically relevant tweets lacking academic evidence from peer-reviewed journals were categorized as unverifiable (n = 290, 13.9%). Medically correct tweets include the ones such as ''Info: Studies suggest high intake of raw/cooked #garlic can significantly lower risk of colorectal/stomach #cancer.'' Medically incorrect ones included: ''Study shows curcumin in #turmeric is as effective as oxaliplatin 4 treating colorectal #cancer.'' Unverifiable tweets were often too vague or without concrete evidence for analysis: ''Increased vitamin D in blood adds years to life and helps prevent colon cancer.'' We also compared information credibility between the popular group and the random group, as shown in Table 5. The top 1000 retweets data set (popular group) showed a statistically significant difference in its share of medically correct information from the randomized 3% of tweets (random group) ($P < 0.01$). More popular tweets, those retweeted by many users, included more medically correct information.

DISCUSSION

The Internet has become an important mass medium for consumers seeking health information.[5,21–23] The evolution of social networking services (SNS) has revolutionized public

TABLE 5. Information Credibility Assessment of the Sampled Tweets Reviewed by the Physicians Committee

	Sampled Tweets		
	Top 1000 (Popular Group)	Random 3% (Random Group)	Total
	N (%)	N (%)	N (%)
Medically Correct	331 (90.7%)	1432 (83.2%)	1763 (84.5%)
Medical Misinformation	5 (1.4%)	28 (1.6%)	33 (1.6%)
Unverifiable	29 (8.0%)	261 (18.2%)	290 (13.9%)

A total of 2086 medically relevant tweets have been retrieved from 3290 sample tweets. Note that the sampled data set of the top 1000 tweets had more medically correct information than the other randomized 3% of tweet data ($\chi^2 = 13.4$, $P < 0.01$).

health communications.[24–26] A recent survey shows that health information is steadily making presence in the social scene. Eight percent of Internet users reported that they have posted a health-related question online or shared their health experiences online.[27] In academia, many have paid heed to social discussions of different medical conditions or diseases on a major SNS platform, Twitter.[6,7,10,17,28–34]

We have taken a data-driven approach to analyze the source and credibility of colorectal cancer information on Twitter. Our results show who talks about what kinds of information on colorectal cancer and whether such information is verifiable. From our analysis, we gleaned the following: news articles and/or research findings related to colorectal cancer are most frequently shared; the push of information about colorectal cancer is mainly from users who are ordinary individuals; but organizations tend to tweet information, more so than individuals. The users frequently use shortened URLs in tweets, which lead to external links that are often news and medical resource websites. As for credibility, more medically correct information was found in frequently shared (''retweeted'') tweets than in randomly selected tweets. Although this may suggest the wisdom of the crowd, because there were far fewer medical professional users than ordinary individuals, tweets seem to carry an inherent risk of unverified knowledge.

Although tweets are limited to 140 characters, they contain a significant piece of information. Our results show that 65.2% of the tweets contain medically relevant information, and the physicians committee was able to determine the medical correctness of 86.1% of the reviewed information. Furthermore, users often embed a shortened URL links in tweets, indicating and supplementing further information in detail. Hence, we suggest that tweets may carry an information value as themselves, and that they may be worth monitoring for public health purposes.

This study finds that tweeting health information, and colorectal cancer information in particular, offers both promises and risks. A piece of well-structured health information in a tweet may educate a patient without medical knowledge or background. Moreover, tweets can help promote positive health changes, as they facilitate social networking. Recent studies are looking into the effect of SNS on promoting health behavior changes, and find that SNS interventions are positively related with health behavioral outcomes, although with heterogeneity.[35] Therefore, Twitter may have the potential to increase access to health information and to encourage engagement.

Nonetheless, Twitter is not a health information broadcasting medium. Medical professionals often disagree on whether Twitter is an effective medium for delivering health information, and prior research shows both support and caution.[10,36] In addition, studies have repeatedly found that users may not interpret health information on the Internet in the most accurate way.[37] Many with low health literacy may be unable to determine the qualifications of online authors and trust in online health information.[38] Our results suggest a possibility of collective intelligence in more frequently retweeted information. Frequently retweeted information, at an aggregate level, might supplement judging of health information on the Internet.

We acknowledge limitations with our work. Twitter is not without selection bias, although a large-scale data analysis is possible.[39] In addition, full transparency of user demographics is impossible on the Internet, and Twitter user profile descriptions are subject to change at the user's whim. Future work should address additional user features such as user activities and follower networks.

The present study has embarked on a social media analysis on the quality of consumer health information about colorectal cancer, a research area that is still not fully unearthed. By addressing health and illness, social networking services such as Twitter are reshaping health care, acting as a powerful new way for doctors and patients to interact and share information.[40] Coupled with its potential to increase social support and feelings of connectedness,[41,42] Twitter may lead consumers to a sense of empowerment[43] and enhance public health, so long as it is used with proper caution. Our work suggests Twitter as a promising venue for monitoring public discourse on health. The present study has informed both research and health professionals about the currency of information shared on colorectal cancer. We believe this research could play an important role in promoting public awareness of colorectal cancer and providing better patient education by analyzing the source and credibility of real-time public information.

REFERENCES

1. Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng. ''Why we twitter: understanding microblogging usage and communities.'' In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56–65. ACM, 2007.

2. Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. ''What is Twitter, a social network or a news media?.'' In Proceedings of the 19th international conference on World wide web, pp. 591-600. ACM, 2010.

3. De Choudhury, Munmun, Meredith Ringel Morris, and Ryen W. White. "Seeking and sharing health information online: Comparing search engines and social media." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1365-1376. ACM, 2014.

4. Kaicker J, Debono VB, Dang W, et al. Assessment of the quality and variability of health information on chronic pain websites using the DISCERN instrument. *BMC Med.* 2010;8:59.

5. Eysenbach G, Powell J, Kuss O, et al. Empirical studies assessing the quality of health information for consumers on the World Wide Web: a systematic review. *JAMA.* 2002;287:2691–2700.

6. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PloS One.* 2010;5:e14118.

7. Scanfeld D, Scanfeld V, Larson EL. Dissemination of health information through social networks: Twitter and antibiotics. *Am J Infect Control.* 2010;38:182–188.

8. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication, and publication behavior on the Internet. *J Med Internet Res.* 2009:e11.

9. Rajagopalan MS, Khanna VK, Leiter Y, et al. Patient-oriented cancer information on the Internet: a comparison of Wikipedia and a professionally maintained database. *J Oncol Pract.* 2011;7:319–323.

10. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? *BMJ.* 2014;349:g6178.

11. Wasserman M, Baxter N, Rosen B, et al. Systematic review of Internet patient information on colorectal cancer surgery. *Dis Colon Rectum.* 2014;57:64–69.

12. Colon Cancer Coalition. Get your rear in gear. Colon Cancer Coalition. http://www.coloncancercoalition.org/. Last Accessed on 12/28/2015.

13. Twitter. The streaming APIs. 2015. *Twitter Developers.* https://dev.twitter.com/streaming/overview. Last Accessed on 12/28/2015.]

14. List of healthcare occupations. *Wikipedia.* https://en.wikipedia.org/wiki/List_of_healthcare_occupations. Last Accessed on 12/28/2015.

15. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. *Am J Eval.* 2006;27:237–246.

16. Carletta J. Assessing agreement on classification tasks: the kappa statistic. *Comput Ling.* 1996;22:249–254.

17. Thackeray R, Burton SH, Giraud-Carrier C, et al. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC Cancer.* 2013;13:508.

18. Segaran T. Programming Collective Intelligence: Building Smart Web 2.0 Applications. Boston, MA: O'Reilly Media; 2007.

19. Berland GK, Elliott MN, Morales LS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. *JAMA.* 2001;285:2612–2621.

20. Khazaal Y, Chatton A, Cochand S, et al. Brief DISCERN, six questions for the evaluation of evidence-based content of health-related websites. *Patient Educ Couns.* 2009;77:33–37.

21. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ.* 2002;324:573–577.

22. Eysenbach G. The impact of the Internet on cancer outcomes. *CA Cancer J Clin.* 2003;53:356–371.

23. Narimatsu H, Matsumura T, Morita T, et al. Detailed analysis of visitors to cancer-related web sites. *J Clin Oncol.* 2008;26:4219–4223.

24. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *J Med Internet Res.* 2008;10:e22.

25. Chou W-yS, Hunt YM, Beckjord EB, et al. Social media use in the United States: implications for health communication. *J Med Internet Res.* 2009;11:e48.

26. Dredze M. How social media will change public health. *IEEE Intell Syst.* 2012;27:81–84.

27. Fox S, Duggan M. Health Online 2013. Pew Internet & American Life Project. Washington, DC: Pew Research; 2013.

28. De la Torre-Díez I, Díaz-Pernas FJ, Antón-Rodríguez M. A content analysis of chronic diseases social groups on Facebook and Twitter. *Telemed J E Health.* 2012;18:404–408.

29. Sugawara Y, Narimatsu H, Hozawa A, et al. Cancer patients on Twitter: a novel patient community on social media. *BMC Res Notes.* 2012;5:699.

30. Himelboim I, Han JY. Cancer talk on Twitter: community structure and information sources in breast and prostate cancer social networks. *J Health Commun.* 2014;19:210–225.

31. Lyles CR, López A, Pasick R, et al. ''5 mins of uncomfyness is better than dealing with cancer 4 a lifetime'': an exploratory qualitative analysis of cervical and breast cancer screening dialogue on Twitter. *J Cancer Educ.* 2013;28:127–133.

32. Murthy, Dhiraj, Alexander Gross, and Daniela Oliveira. ''Understanding cancer-based networks in Twitter using social network analysis." In Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on, pp. 559-566. IEEE, 2011.

33. Sullivan SJ, Schneiders AG, Cheang C-W, et al. ''What's happening?'' A content analysis of concussion-related traffic on Twitter. *Br J Sports Med.* 2012;46:258–263.

34. Shive M, Bhatt M, Cantino A, et al. Perspectives on acne: what Twitter can teach health care providers. *JAMA Dermatol.* 2013;149:621–622.

35. Laranjo L, Arguel A, Neves AL, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. *J Am Med Inform Assoc.* 2015;22:243–256.

36. Robillard JM, Johnson TW, Hennessey C, et al. Aging 2.0: health information about dementia on Twitter. *PloS One.* 2013;8:e69861.

37. Fiksdal AS, Kumbamu A, Jadhav AS, et al. Evaluating the process of online health information searching: a qualitative approach to exploring consumer perspectives. *J Med Internet Res.* 2014;16:e224.

38. Diviani N, van den Putte B, Giani S, et al. Low health literacy and evaluation of online health information: a systematic review of the literature. *J Med Internet Res.* 2015;17:e112.

39. Janssens ACJ, Kraft P. Research conducted using data obtained through online communities: ethical implications of methodological limitations. *PLoS Med.* 2012;9:e1001328.

40. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. *Health Aff.* 2009;28:361–368.

41. Wangberg SC, Andreassen HK, Prokosch H-U, et al. Relations between Internet use, socio-economic status (SES), social support, and subjective health. *Health Promot Int.* 2008;23:70–77.

42. Idriss, Shereene Z., Joseph C. Kvedar, and Alice J. Watson. ''The role of online support communities: benefits of expanded social networks to patients with psoriasis.'' Archives of Dermatology 145, no. 1 (2009): 46–51.

43. Van Uden-Kraan CF, Drossaert CH, Taal E, et al. Participation in online patient support groups endorses patients' empowerment. *Patient Educ Couns.* 2009;74:61–69.