

Transmissibility of the Campaign for Colorectal Cancer Awareness in Korea Among Twitter Users

Keun Chul Lee, Heung-Kwon Oh, Gibeom Park¹, SoHyun Park¹, Bongwon Suh¹, Woo Kyung Bae², Jin Won Kim³, Hyuk Yoon³, Myung Jo Kim, Sung-Il Kang, Il Tae Son, Duck-Woo Kim, Sung-Bum Kang

Department of Surgery, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam; ¹Department of Transdisciplinary Studies, Graduate School of Convergence Science and Technology, Seoul National University, Seoul; ²Health Promotion Center, Department of Family Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam; ³Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Korea

Purpose: The Korean Society of Coloproctology holds its annual colorectal awareness month every September. This study analyzed the users and the contents of Korean tweets regarding colorectal cancer and estimated the transmissibility of the awareness campaign among Twitter users.

Methods: Prospective data collection was employed to accumulate Korean tweets containing the keywords “colorectal cancer,” “colorectal cancer awareness campaign,” “gold ribbon,” and/or “love handle,” from August 1 to September 30, 2014. Twitter users and contents were analyzed, and the credibility of information-sharing tweets throughout the study period was evaluated.

Results: In total, 10,387 tweets shared by 1,452 unique users were analyzed. As for users, 57.8% were individuals whereas 5.8% were organizations/communities; spambots accounted for a considerable percentage (36.4%). As for content, most tweets were spam (n = 8,736, 84.1%), repetitively advertising unverified commercial folk remedies, followed by tweets that shared information (n = 1,304, 12.6%) and non-information (n = 347, 3.3%). In the credibility assessment, only 80.6% of the information-sharing tweets were medically correct. After spam tweets had been excluded, a significant increase was seen in the percentage of information-sharing tweets (77.1% to 81.1%, P = 0.045) during the awareness campaign month.

Conclusion: Most Korean tweets regarding colorectal cancer during the study months were commercial spam tweets; informative public tweets accounted for an extremely small percentage. The transmissibility of the awareness campaign among Twitter users was questionable at best. To expand the reach of credible medical information on colorectal cancer, public health institutions and organizations must pay greater attention to social media.

Keywords: Social media; Twitter; Colorectal neoplasms; Awareness campaign; Medical information

Received: June 9, 2016 • Accepted: September 6, 2016

Correspondence to: Heung-Kwon Oh, M.D.

Department of Surgery, Seoul National University Bundang Hospital,
82, Gumi-ro 173beon-gil, Bundang-gu, Seongnam 13620, Korea
Tel: +82-31-787-7105, Fax: +82-31-787-4078, E-mail: crsohk@gmail.com

Co-correspondence to: Bongwon Suh, Ph.D.

Department of Transdisciplinary Studies, Graduate School of Convergence
Science and Technology, Seoul National University, 1, Gwanak-ro,
Gwanak-gu, Seoul 08826, Korea
Tel: +82-31-888-9124, Fax: +82-31-888-9148, E-mail: bongwon@snu.ac.kr

The preliminary results of this study were presented at the annual meeting
of the Korean Society of Coloproctology, held from April 10–12, 2015 in
Busan, Korea

© 2016 The Korean Society of Coloproctology

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Worldwide, colorectal cancer (CRC) is known to be the third most common form of cancer among males (746,000 cases, 10.0% of the total) and the second most common form among females (614,000 cases, 9.2% of the total) [1]. The incidence of age-standardized CRC in Korea is estimated to be as high as 45 per 100,000 according to GLOBOCAN 2012, and Korea has shown a sharp increase in that incidence recently [2, 3]. Based on National Cancer Registration Statistics, 27,618 new CRC cases occurred in 2013 in Korea [4].

CRC screening is important and is widely recommended as standard practice for early detection in many countries. However, one of the most promising avenues for reducing the number of

cancer deaths is through prevention. The majority of cancers, including CRC, are preventable through primary or secondary prevention (such as lifestyle modification or colonoscopy) [5]. In the United Kingdom, a cancer awareness campaign entitled 'Be Clear on Cancer' was held in January 2012 and was thought to have improved public awareness of CRC and to have encouraged the public to seek urgent medical attention when certain signs/symptoms are present [6]. Similarly, every March has been 'National Colorectal Cancer Awareness Month' in the US since 2000: "Getting screened could save your life," as early detection for CRC will enable earlier treatment; as a result of this effort, CRC mortality has decreased over the past decade in the United States [7]. Since 2007, the Korean Society of Coloproctology has dedicated the month of September as annual colorectal awareness month. This campaign, also known as the Gold Ribbon Campaign, is a major event to increase public awareness of CRC and to promote screening by colonoscopy.

The Internet is a common venue for disseminating and accessing health information, and traditional mass media coverage is now further enhanced by social networking services (SNS) [8]. Twitter is a popular microblogging service where users create status messages (called "tweets") among SNS. Tweets are text-based posts of up to 140 characters in length, and as many as 500 million are sent per day [9]. Twitter users can follow each other to receive a real-time feed of the users' respective tweets. Users can also pass others' messages on to their followers (called "retweets") as well as make explicit references to others by username (called "mentions"), which puts the tweet into an additional subscription feed. With the context of previous studies about CRC in tweets in English [8], this study adopted the same approach for tweets written in the Korean language. This study aimed to analyze the users and the contents of Korean tweets regarding CRC, and then to estimate the transmissibility of the awareness campaign among Twitter users.

METHODS

Collecting data

As a program to promote awareness about the prevention and the treatment of CRC, a campaign called the 'gold ribbon campaign' has been held annually in Korea since 2007 [10]. In 2014, the catchphrase 'Catch love handle, catch CRC' was especially used. For potential improvements in the campaign's effectiveness and acceptability, we collected and analyzed tweets related to CRC written in Korean from August 1, 2014, through September 30, 2014, to help us understand how the recently increased attention to social media might be utilized to increase people's awareness of a healthcare issue. Utilizing Twitter's public Application Programming Interface (API) and streaming API for real-time updates scripted in Python language, we were able to collect publicly available tweets containing colorectal-cancer-related keywords. The keywords we used for the search were generated by combining

"colorectal cancer," "colorectal cancer awareness campaign," "gold ribbon," and/or "love handle," in the Korean language. Each acquired tweet contained at least one of these keywords.

Classifying users

One of our interests was to be able to identify a tweet written by a medical professional or a medical institute. In order to determine the authorship, we collected the user's public profile for each tweet. We classified the user as either an individual or an organization based on the user's profile description. Users with a blank profile description were categorized as unverifiable. We further classified individuals into medical professionals or laypersons. For the organization category, we manually subcategorized them by using inductive reasoning. The subcategories included "news/media," "foundations/communities," "medical institutions," "government affiliates," and "other."

Categorizing contents

The contents were classified into categories of "spam," "informative tweet," or "communicative tweet." A spam tweet was defined as one that repetitively advertised unverified commercial folk remedies for ailments. Informative tweets educated and informed readers about cancer prevention, treatment, news, prognosis, and screening and promoted hospitals. Communicative tweets contained personal chats, jokes, curse phrases, celebrity news and/or simple questions.

Identifying external sources of information

Each tweet is limited to 140 characters or less. In order to conserve characters, a tweet normally contains a shortened uniform resource locator (URL) such as bit.ly or tinyurl.com to point to the external source of information. We performed unshortening to identify the original URL for each shortened URL. Because a shortened URL may refer to another shortened URL and so on, we expanded the shortened URLs into their original forms by following HTTP redirects.

Assessing the credibility of the information

We formed a committee of physicians to review the contents of informative tweets. The committee members included four board-certified physicians specializing in gastroenterology, family medicine, colorectal surgery, and medical oncology, respectively. The committee referred to previous literature and quality standard criteria to confirm or refute the correctness of the classified tweets. Due to the limited number of available models for information credibility on Twitter, all the committee members reviewed each tweet and analyzed the content to reach an agreement on its classification: "medically correct," "medical misinformation," and "unverifiable" [11]. An unverifiable tweet was defined as an informative tweet without any academic evidence from peer-reviewed journals.

Assessing the transmissibility of the awareness campaign

We compared the numbers of CRC-related tweets before and during the awareness campaign month to identify the transmissibility of the awareness campaign among Twitter users. We also counted the number of tweets directly related to the awareness campaign by using such phrases as “colorectal cancer awareness campaign,” “gold ribbon,” and/or “love handle” during the awareness campaign month.

Statistical analysis

Categorical variables were compared using the chi-square test or Fisher exact test, as appropriate. Continuous variables were compared using the Student unpaired t-test. All reported P-values are 2-tailed, with a P-value of 0.05 indicating a statistically significant difference. Analyses were conducted using STATA 14 (StataCorp LP, College Station, TX, USA).

RESULTS

Number of tweets collected, user analysis, and account analysis

Using the above listed terms, the total number of collected tweet was 10,387 from 1,452 distinct users, consisting of 9,665 original tweets and 722 retweets. The number of tweets from individual users was 6,005 (57.8%), followed by spambots (n = 3,780, 36.4%), and organizational users (n = 602, 5.8%). In terms of the tweet account analysis (n = 1,452), the number of individual user accounts was 794, among which 786 (98.9%) had user profiles belonging to laypersons, and only 8 (1.1%) had user profiles belonging to medical professionals (Table 1).

Analyses of the contents and the credibilities of the tweets

The total numbers of tweets for each content category were as follows: 8,736 (84.1%) for spam, 1,304 (12.6%) for informative tweets and 347 (3.3%) for communicative tweets. Excluding

spam, 1,651 relevant tweets, including informative and communicative tweets, remained. Of them, 979 tweets (59.2%) had one or more URLs embedded in them. Table 2 shows the top 10 domains. Portal sites and news resource websites had the greatest share on Twitter. The contents of the 1,304 informative tweets were classified into categories of individual tweets (n = 263, 20.2%) and organizational tweets (n = 1,041, 79.8%). The contents of these informative tweets included cancer prevention (37.4%), treatment (20.0%), news (17.5%), prognosis (7%), screening (6%), hospital promotion (2%), and others (10%) (Fig. 1). We assessed the credibilities of the 1,304 informative tweets by judging the quality of the information. We determined that 1,052 tweets (80.6%) were medically correct, 15 (1.2%) were medically incorrect and the remaining 237 (18.2%) were unverifiable. Representative tweets are presented in Table 3.

Transmissibility of the awareness campaign

Over the period of the study, out of the total 1,651 relevant tweets, 864 (52.3%) were identified in August and 787 (47.7%) in September (awareness campaign month) (Fig. 2). The numbers of tweets

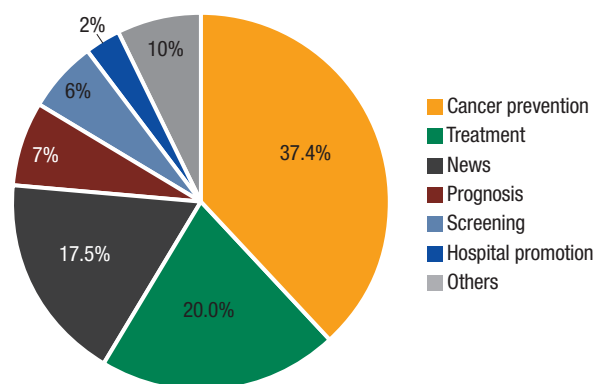


Fig. 1. Classification of Informative tweets (n = 1,304).

Table 2. Top 10 URLs shared most frequently by users in tweets with URL links (n = 979)

Rank	Domain	Domain type	No. (%)
1	media.daum.net	Portal site	84 (8.58)
2	blog.naver.com	Portal site	69 (7.05)
3	blog.daum.net	Portal site	61 (6.23)
4	cancer_info.blog.me	Government	55 (5.62)
5	news.naver.com	News	52 (5.31)
6	medi-talk.co.kr	Medical blog	44 (4.49)
7	www.pressian.com	News	38 (3.88)
8	www.mediup.co.kr	Portal	36 (3.68)
9	www.facebook.com	Social media	28 (2.86)
10	opencast.naver.com	Portal site	27 (2.76)

Table 1. User account analysis (n = 1,452)

Category	No. (%)
Individuals	794 (54.7)
Medical professionals	8 (1.1)
Laypersons	786 (98.9)
Organizations	81 (5.6)
Government	18 (22.2)
Medical institutes	21 (25.9)
Academic organizations	3 (3.7)
Nongovernmental organizations	1 (1.2)
Mass media	30 (37.0)
Commercial firms	8 (9.9)
Spambot	577 (39.7)

created per day were 27.8 ± 20.1 in August and 26.2 ± 14.2 in September. Thus, no significant increase in the number of CRC-related tweets was noted during the awareness campaign month ($P = 0.716$). However, a significant increase in the percentage of in-

formation-sharing tweets ($n = 666, 77.1\%$ vs. $n = 638, 81.1\%$, $P = 0.045$) was noted during the awareness campaign month (Table 4). This qualitative aspect shows that the number of tweets with medically-correct information increased in September ($n = 528, 67.1\%$), compared to August ($n = 524, 60.7\%$) (Fig. 3).

Table 3. Representative tweets according to medical correctness after a credibility analysis of the informative tweets ($n = 1,304$)

Classification	Representative tweets
Medically correct	Waist size, a stronger predictor than body mass index of colorectal cancer
Medically incorrect	Colorectal cancer is found late, more than a half of the cases
Unverifiable	Drinking mushroom reduces the risk of colorectal cancer?

Table 4. Content analysis of colorectal-cancer-related tweets during August and September 2014

	August 2014 ($n = 864$)	September 2014 ($n = 787$)	P-value
Informative	666 (77.08)	638 (81.07)	0.045
Communicative	198 (22.92)	149 (18.93)	

Values are presented as number (%).

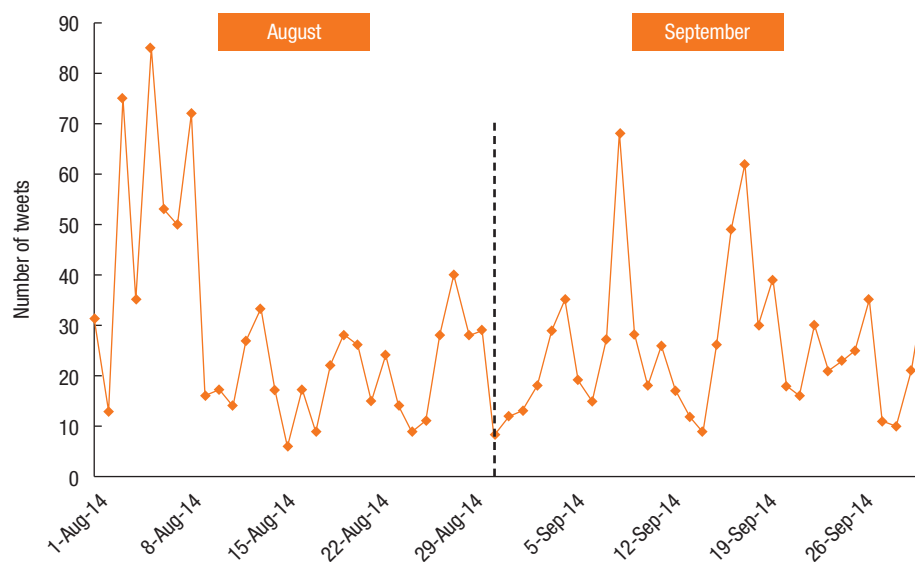


Fig. 2. Tweets related to colorectal cancer (CRC). Excluding spam, 864 tweets were identified in August 2014 and 787 tweets in September 2014. No increase of CRC-related tweets was noted during the month of the awareness campaign ($P = 0.716$).

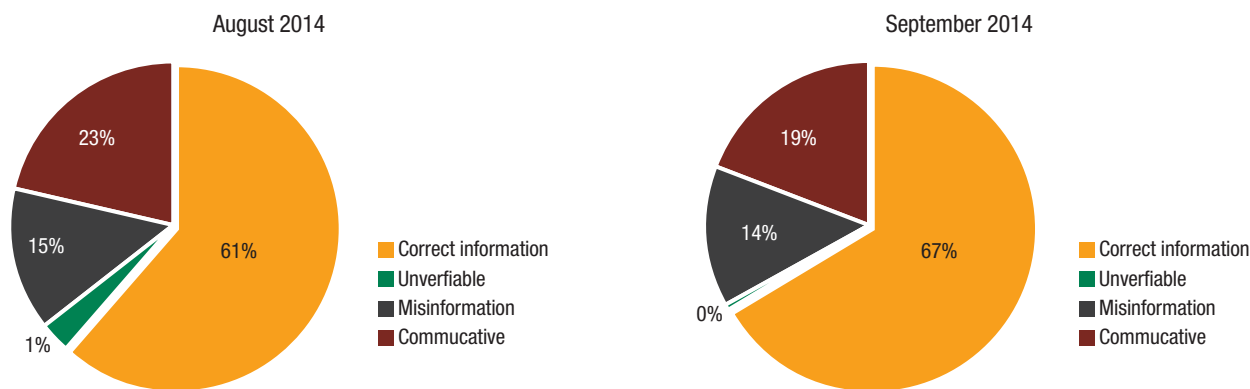


Fig. 3. Results for the credibility analysis of relevant tweets ($n = 1,651$) during August and September 2014. The number of tweets with medically-correct information increased in September ($n = 528, 67.1\%$) compared to August ($n = 524, 60.7\%$).

DISCUSSION

In this study, we adopted a data-driven approach to analyze the source and the credibility of CRC information on Twitter in Korea. We evaluated the source of CRC-related health information, the type of collected information, and the validity of that information. Our study found that the majority of the tweets written in Korean regarding CRC were mostly commercial spam or had inappropriate content. An extremely small percentage consisted of informative public tweets.

SNS represent an important communications medium that enables users to exchange information and relate socially [12]. A tweet is not a simple act of copying information, but enables amplification of the information for a specific audience [13]. Moreover, it makes users publicly agree with someone or validate others' thoughts. In the whole, it constitutes a conversational environment in which a shared conversational context is formed through a public interplay of voices. For this reason, our study used CRC-related tweets over the month of the awareness campaign and the previous month to examine how particular health information was dispersed throughout a network of interconnected users on Twitter.

This study showed that ordinary individuals produced almost 4 times more tweets about CRC information than organizational users. They frequently shared colorectal-cancer prevention and/or treatment related information. In places where content was solely determined by its own users, the dynamics of collective attention may cause a reinforcement effect, resulting in users having an incorrect understanding of a certain topic. Considering that the number of nonmedical professionals constitutes more than 98% of the individual users of the collected data, nonmedical sources most likely play a major role in spreading health information faster and increasing its exposure to large groups.

Our analysis indicated that some tweets shared misinformation and disseminated confusing information on colorectal health issues. About a quarter of the collected tweets written by nonmedical individual users were classified as noninformative messages. In August, a month before the awareness campaign, we collected almost three times as many tweets than we did in September, including spambots. However, after irrelevant spam had been excluded from the tweets collected over the 2 months, we obtained similar numbers of relevant tweets in both months. To some extent, the awareness campaign might have contributed to a reduction in the amount of spam during the awareness campaign (September) month. Furthermore, a significant increase in the percentage of information-sharing tweets was noted during the awareness campaign month.

More than half the collected tweets contained URLs, helping to identify the original sources of information. We found that a quarter of the links pointed to other blogging sites and contained posts about other patients' personal experiences. A handful of medical institutes or government authorities were mentioned in the tweets. This implies a propagation of information across dif-

ferent social networks [13], forming richer sets of publicly available health information beyond the initially isolated domain of analysis in Twitter. Moreover, with the several types of connected electronic devices these days, including mobile smartphones, many people may talk about a particular topic at once. A stream of messages would make individuals peripherally aware of a topic even without active participation in the conversation. With people's increased level of interest in health information on the Web [14], the creation and subsequent distribution of unverified health information will hinder people from obtaining a correct understanding of the related health issues.

Based on the above, medical professionals and institutes have to pay more attention to the effect of message diffusion through SNS in the present a fast-changing environment [15, 16]. Active and explicit promotion and monitoring on Twitter during the awareness campaign month could affect tweeting practice about CRC, thereby reducing the frequency of irrelevant or inappropriate messages. If the groups of people who do not actively tweet but have access to published tweets and monitor the transmissions are considered, one cannot ignore the potential impact of inappropriate information sharing. For trusted and high-quality knowledge sharing, medical professionals should not simply send messages to audiences, but should validate such messages and engage with people to promote and guide health information properly.

This study had some limitations. The main limitation is the small sample size, which may not be representative of social media for health communication. Another limitation is the 140-character format, which is not sufficient for delivering the intent of the tweet's author properly. This is particularly true for a retweet, for which the original intent of a tweet may be incrementally misunderstood. Tweets containing URLs to external social media services do not provide an answer at this moment, and a large fraction of those are abused for spam links. Future studies can go in various directions. If medical professionals' will participate actively in using Twitter, we will be able to assess the influence, the legitimacy, and the utility of social media as a new communication medium for health information. Also, the timeframe for collecting larger data sets may be extended to evaluate the accuracy of our model better.

In conclusion, most of the Korean tweets regarding CRC were commercial spam tweets; informative public tweets accounted for an extremely small percentage. The transmissibility of the awareness campaign among Twitter users was questionable at best. If the reach of credible medical information on CRC is to be extended, public health institutions and organizations need to pay greater attention to social media. The Korean Society of Coloproctology also needs to participate in social media and provide the public with reliable healthcare information about CRC.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

The authors are indebted to J. Patrick Barron, Professor Emeritus, Tokyo Medical University and Adjunct Professor, Seoul National University Bundang Hospital for his pro bono editing of this manuscript.

REFERENCES

1. Chiu HM, Chang LC, Hsu WF, Chou CK, Wu MS. Non-invasive screening for colorectal cancer in Asia. *Best Pract Res Clin Gastroenterol* 2015;29:953-65.
2. International Agency for Research on cancer. Globocan 2012 [Internet]. Lyon(FR): International Agency for Research on cancer; c2016 [cited 2016 Jun 1]. Available from: <http://globocan.iarc.fr/Default.aspx>.
3. Shin A, Jung KW, Woo H, Jeong SY. Colorectal cancer incidence in Korea is not the highest in the world. *Cancer Res Treat* 2016; 48:864-7.
4. National Cancer Registration and statistics. Cancer incidence [Internet]. Goyang (KR): National Cancer Registration and statistics; c2009 [cited 2016 Aug 16]. Available from: <https://ncrs.cancer.go.kr/index.do>.
5. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 2015;347:78-81.
6. Khong TL, Naik K, Sivakumar R, Shah S. Impact of the United Kingdom national bowel cancer awareness campaigns 2012 on colorectal cancer services and patient survival. *Colorectal Dis* 2015;17:1055-61.
7. American Cancer Society. Cancer reference information [Internet]. Atlanta (GA): American Cancer Society; c2009 [cited 2016 Aug 1]. Available from: <http://www.cancer.org/myacs/midwest/areahighlights/march-motm-2014>.
8. Park S, Oh HK, Park G, Suh B, Bae WK, Kim JW, et al. The source and credibility of colorectal cancer information on twitter. *Medicine (Baltimore)* 2016;95:e277.
9. DMR. 170 Amazing twitter statistics and facts [Internet]. Digital Marketing Ramblings; c2009 [cited 2016 Aug 1]. Available from: <http://expandedramblings.com/index.php/march-2013-by-the-numbers-a-few-amazing-twitter-stats/10/>.
10. The Korean Society of coloproctology [Internet]. Seoul (KR): The Korean Society of coloproctology; c2005 [cited 2016 Jun 1]. Available from: <http://www.colon.or.kr/colonlife/info/index.kin?gubun=2&subid=1>.
11. Oyeyemi SO, Gabarron E, Wynn R. Ebola, Twitter, and misinformation: a dangerous combination? *BMJ* 2014;349:g6178.
12. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *J Med Internet Res* 2008;10:e22.
13. Evers JL. Re-tweeting science. *Hum Reprod* 2014;29:639.
14. Van de Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. *J Med Internet Res* 2013;15:e220.
15. Lee JL, DeCamp M, Dredze M, Chisolm MS, Berger ZD. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. *J Med Internet Res* 2014;16:e237.
16. Chretien KC, Azar J, Kind T. Physicians on twitter. *JAMA* 2011; 305:566-8.