

Health Information Technology Trends in Social Media: Using Twitter Data

Jisan Lee¹, Jeongeun Kim^{2,3}, Yeong Joo Hong², Meihua Piao², Ahjung Byun², Healim Song⁴, Hyeong Suk Lee²

¹Department of Nursing Science, College of Life & Health Sciences, Hoseo University, Asan, Korea

²College of Nursing, Seoul National University, Seoul, Korea

³Research Institute of Nursing Science, Seoul National University, Seoul, Korea

⁴Medical Intensive Care Unit, Samsung Medical Center, Seoul, Korea

Objectives: This study analyzed the health technology trends and sentiments of users using Twitter data in an attempt to examine the public's opinions and identify their needs. **Methods:** Twitter data related to health technology, from January 2010 to October 2016, were collected. An ontology related to health technology was developed. Frequently occurring keywords were analyzed and visualized with the word cloud technique. The keywords were then reclassified and analyzed using the developed ontology and sentiment dictionary. Python and the R program were used for crawling, natural language processing, and sentiment analysis. **Results:** In the developed ontology, the keywords are divided into 'health technology' and 'health information'. Under health technology, there are six subcategories, namely, health technology, wearable technology, biotechnology, mobile health, medical technology, and telemedicine. Under health information, there are four subcategories, namely, health information, privacy, clinical informatics, and consumer health informatics. The number of tweets about health technology has consistently increased since 2010; the number of posts in 2014 was double that in 2010, which was about 150 thousand posts. Posts about mHealth accounted for the majority, and the dominant words were 'care', 'new', 'mental', and 'fitness'. Sentiment analysis by subcategory showed that most of the posts in nearly all subcategories had a positive tone with a positive score. **Conclusions:** Interests in mHealth have risen recently, and consequently, posts about mHealth were the most frequent. Examining social media users' responses to new health technology can be a useful method to understand the trends in rapidly evolving fields.

Keywords: Consumer Health Informatics, Data Mining, Natural Language Processing, Social Media, Public Opinion

Submitted: February 14, 2019

Revised: March 25, 2019

Accepted: April 3, 2019

Corresponding Author

Hyeong Suk Lee

College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul 03080, Korea. Tel: +82-2-2287-1024, E-mail: leehyeongsuk@snu.ac.kr (<https://orcid.org/0000-0001-7597-3163>)

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.

© 2019 The Korean Society of Medical Informatics

1. Introduction

There has been a rapid transformation in health technology trends in the recent years. Among various ways of exploring these trends, social media analysis has surfaced as a useful methodology [1]. With the shift of empowerment of health information to consumers, health consumers are aggressively using social media to share their experiences and collect opinions from others [2]. Among the various types of social media, Twitter is characterized by its real-time features, strong delivery, publicness, causal ambience, and individuality [3]. Furthermore, unlike other social media platforms

that feature a bidirectional network among users, Twitter features a unidirectional structure, wherein users simply follow and read tweets by other users, companies, and media of their interest [4]. This structure enhances Twitter's ability to disseminate information.

In a previous study that analyzed Twitter data in 2009 during the spread of the H1N1 virus, only about 5% of the messages contained incorrect information, and Twitter users acquire personal experiences by retweeting information that they have gained indirectly [5]. In other words, Twitter users share personal experiences and concerns related to health and diseases by tweeting, and the retweeting of these messages by others indicates that their content is interesting and potentially influential. Twitter data contains a writer's opinions or emotions about a topic; therefore, it is possible to analyze the public's mood about a specific topic through opinion mining, through which extreme words are extracted from sentences containing keywords of interest [6]. Applying opinion mining and sentiment analysis techniques to big online data for the extraction of useful information on any event or topic is gaining more and more interest with the growing number of internet users and recent developments in information and communication technologies (ICT) [7]. Understanding user profiles, preferences, and barriers can help providers prioritize where to direct efforts when using evidence-based social media in their practice [8]. Moreover, social media, a communication boon for the public health community has the potential to promote and change many health-related behaviors and issues, particularly in times of crisis [9].

Therefore, analysis of social media data would be beneficial for understanding potential healthcare consumers because this data is voluntarily created by users, unlike the data produced in controlled environments, such as data obtained through surveys or interviews.

Considering the rapid advances in health technology, investigating traditional standardized data would be a limited approach to trend analysis. In the present study, we analyzed health technology trends and the sentiments of users who have posted content on relevant topics in the past seven years by using Twitter data to examine the public's opinions about health technology and identify their needs.

II. Methods

1. Collection of Twitter Data

Twitter data related to health technology, from January 2010 to October 2016, were collected using Python. About

440 thousand tweets were collected, and about 1.76 million words were retrieved through natural language processing using R (Table 1).

2. Development of Ontology

This study used open data. The characteristics of Twitter users were not included in the analysis; therefore, Institutional Review Board approval for this research was not necessary. To collect and classify big social data, two methods can be used. One is the top-down method, in which an ontology is developed by analyzing the theoretical background of the topic of interest, and then collecting keywords in the ontology; the other method is bottom-up, in which the topic of interest is collected using a web crawler and then classified [10]. In the present study, we used the top-down method to extract and semantically classify keywords related to health technology (Figure 1). These keywords were collected from Google Trends, Web of Science, MEDLINE, and Hashtagify [11]. Then they were classified with reference to the classification system shown in Medical Informatics [12]. Google Trends was used to find the most popular keywords in the search engine, and Web of Science and MEDLINE were included as a search pool for the extraction of the most frequently used keywords in academia. We could find the most frequently used keywords on Twitter by using Hashtagify. From the 103 keywords that were identified, duplicate words and 9 additional words were excluded based on review by experts, including professors and PhD students in nursing informatics, resulting in a total of 54 keywords for semantic classification.

3. Data Analysis

Python was used for crawling, and the R program was used for natural language processing. Frequently occurring keywords were analyzed by year using the R program and visualized using the word cloud technique. Then the keywords

Table 1. Number of Twitter postings related to health technology

Year of posting	Number of postings
2010	5,257
2011	7,587
2012	12,086
2013	13,637
2014	220,207
2015	27,813
2016	148,624

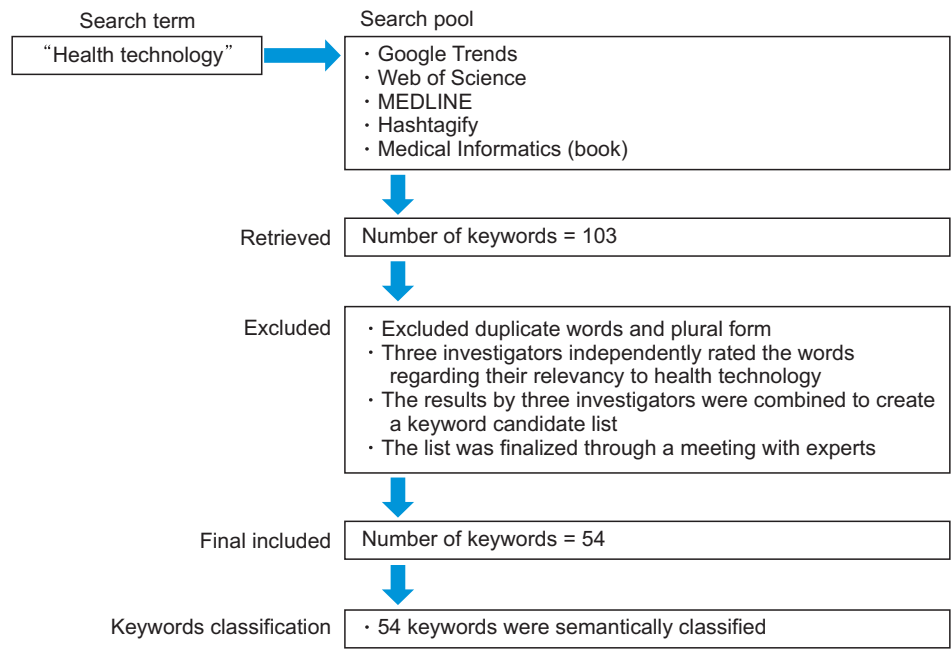


Figure 1. Keywords retrieval process.

Table 2. Developed ontology of health technology

Category	Keywords
Healthcare technology	
Healthcare technology	Health and technology, Healthcare information technology, Healthcare technology, Health information technology, Health science technology, Health technology, Healthcare technology, Healthcare technology solutions, Healthcare technology trends, HITsm
Wearable technology	Wearable technology
Biotechnology	Bioinformatics, Biomedical enhancement, Biomedical technology, Genetic enhancement, Nanomedicine
Mobile health	Digital health, Digital health technology, eHealth, Home telemedicine devices, mHealth, Mobile health, Mobile health technology, Mobile telemedicine
Medical technology	Medical technology, Advanced medical technology, Medical device innovation
Telemedicine	Telehealth, Telehomecare, Telemedicine, Telemonitoring
Health information	
Health information	Health informatics, Health information, Health information exchange, Health information management, Health information system, Healthcare data, Healthcare data analytics, Healthcare information, HealthMap
Privacy	Health information privacy, Health information security, Protected health information
Clinical informatics	Clinical data, Electronic health records, Medical informatics, Medical information
Consumer health informatics	Healthvet, HealthVault, Healthwise, Patient portals, Patient web portals, Personal health records, Public health informatics

were reclassified and analyzed using the developed ontology and sentiment dictionary. A sentiment dictionary is a dictionary in which positive words and negative words are defined. In this study, we used SentiWordNet [13]. Sentiment analysis, referred to as sentiment classification, separates text

items (such as tweets, product reviews, blog posts, etc.) into positive or negative opinions and expresses the degree of positivity or negativity as a score [14]. When a subcategory is mentioned with a positive word, it is given a positive score (+1), and when it is accompanied by a negative word, it is

given a negative score (-1). Sentiment scores were calculated by dividing the total number of text items by the numbers of positive and negative scores [14,15].

III. Results

1. Developed Ontology of Health Technology

The ontology developed after the classification of health technology keywords is shown in Table 2. The keywords are broadly divided into 'health technology' and 'health information'. Under health technology, there are six subcategories, namely, health technology, wearable technology, biotechnology, mobile health, medical technology, and telemedicine; under health information, there are four subcategories, namely, health information, privacy, clinical informatics, and consumer health informatics.

2. Frequency of Posts on Twitter Related to Health Technology

The number of tweets about health technology has consistently increased since 2010; the number of these posts in 2014 was double of that in 2010, which was about 150 thousand posts. Twitter posts about health technology were classified by year into 10 subcategories after the removal of duplicate tweets based on time and ID. It was observed that posts about mHealth accounted for the majority (Figure 2). Figure 3 shows a word cloud illustrating the frequency of words in Twitter posts related to health technology. The left part of Figure 3 shows the Nightingale word cloud of posts related to health technology from 2010 to 2016, and the right part shows the Alice word cloud of posts from 2016. The dominant words were 'care', 'new', 'mental', and 'fitness'.

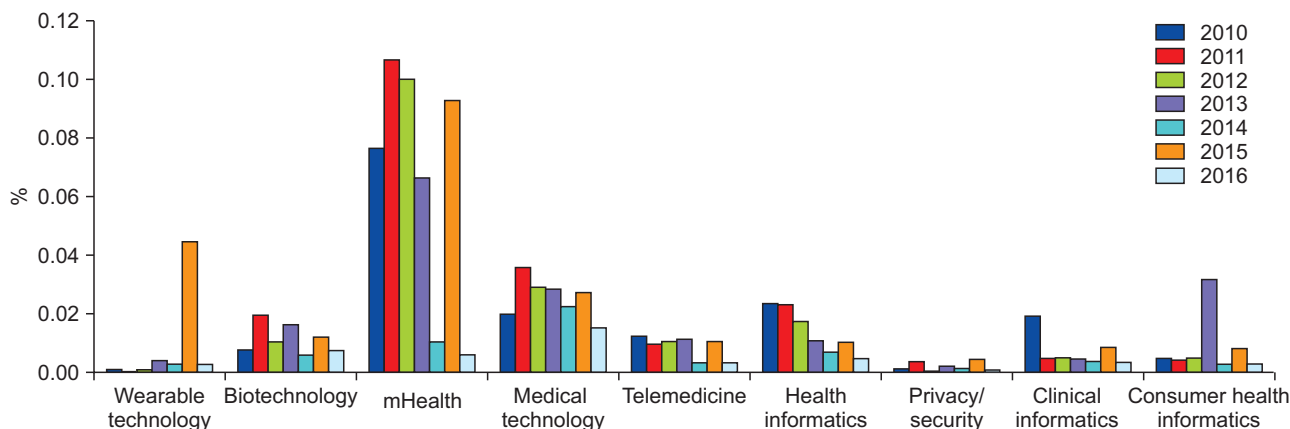


Figure 2. Frequency of Twitter posting of categories.

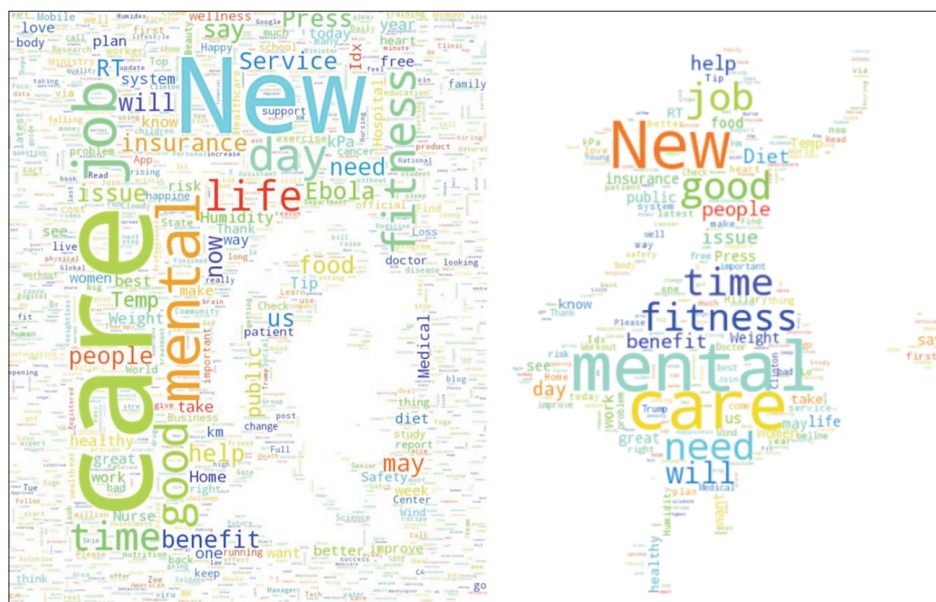


Figure 3. Word clouds of Twitter posts related to health technology.

Table 3. Results of sentiment analysis of subcategories related to health technology

Year	Sentiment score ^a								
	Wearable technology	Biotechnology	mHealth	Medical technology	Telemedicine	Health informatics	Privacy	Clinical informatics	Consumer health informatics
2010	0.17	0.18	0.20	0.33	0.17	0.18	0.00	0.64	0.42
2011	0.00	0.19	0.20	0.13	0.18	0.26	0.97	0.32	0.78
2012	0.38	0.11	0.19	0.21	0.50	0.07	0.17	0.26	0.12
2013	0.27	0.17	0.27	-0.34	0.35	-0.01	-0.13	0.44	0.13
2014	0.27	0.15	0.00	0.02	0.07	0.22	-0.03	0.31	0.05
2015	0.15	0.23	0.22	0.30	0.56	0.22	-0.28	0.14	0.16
2016	0.22	0.14	0.19	0.12	0.09	0.33	0.19	0.21	0.20

^aA positive score means that the subcategory was mentioned with positive words, and a negative score means that the subcategory was mentioned with negative words.

3. Sentiment Analysis by Subcategory

Sentiment analysis by subcategory showed that most of the posts in nearly all subcategories had a positive tone with a positive score. However, for some years, posts about mHealth, medical technology, health informatics, and privacy had negative scores. In 2013, the sentiment scores for each subcategory were -0.34, -0.01, and -0.13, respectively (Table 3).

IV. Discussion

Recent advances in mobile internet and ICT have enhanced connectivity, regardless of time and place, and have thus contributed significantly to various healthcare solutions. Various health problems, such as the increasing number of chronic diseases and the high cost of health services, highlight the need to empower patients and families to practice self-care; hence, the need to provide direct access to health services has emerged over the years [16]. Recently, mHealth solutions have been found to address these health problems regardless of time and place. For this reason, interests in mHealth have risen; thus, posts about mHealth were found to be the most frequent in our study. Myriad mobile applications have been introduced that promote health and disease self-management. As of 2012, there were about 13,000 health apps for consumers on the Apple AppStore, of which 5.8% were related to mental health, 4.13% to sleep, and 11.44% to stress and relaxation [17]. A 2013 study reported the existence of 14,000 health apps, of which 558 were for mental health and behavioral disorders, out of which two-thirds were for autism, anxiety, depression, and attention deficit hyperactivity disorder [18].

Clearly, mHealth has an enormous potential for enhancing

healthcare and quality of life. Multiple studies have reported that utilizing mHealth can cut costs and improve patients' clinical outcomes [19,20]. Furthermore, big data obtained from mHealth are expected to offer new insights for research that aims to enhance the quality of various healthcare services.

According to annual trends, posts about telemedicine, privacy, and consumer health informatics peaked in 2013 and 2015. Telemedicine reimbursement is now being provided by several individual states for Medicaid and/or commercial payers. These state-level policy changes are likely to have significant impact on the viability of telemedicine programs and the utilization of services from all payers, and not just on those services and payers who are affected directly by state policy. This is because telemedicine programs are required to serve patients from almost all payers [21]. In 2013 and 2015, the telemedicine parity legislation was passed or expanded in several states of the United States. This shows that advances in telemedicine were coupled with institutional measures, which led to increasing interest in the matter. The introduction and usage of telehealth in Korea is lagging behind its implementation in other countries, largely due to the lack of policy and adequate legislation [22]. In addition to continuous exploration of the needs for telemedicine and analysis of its effects through pilot studies, discussions regarding relevant institutional support are also needed.

Sentiment analysis of healthcare suggests the services that consumers prefer [23]. In this study, most sentimental scores for health technology were positive. These results, like those of previous studies, have proven that Twitter data helps us better understand consumers' positive feelings about health technology and that Twitter is a useful platform for sharing positive opinions on this topic [24]. However, posts

about privacy had negative sentimental scores in 2013 and 2015. This suggests that consumers have concerns regarding privacy as their health information is increasingly being handled digitally. Along with the expansion of the digital world, public attitudes toward privacy are also evolving [25]. According to one report, 80% of the digital data stored in the United States are pertinent to consumers, and the majority are data related to consumers' lives, such as metadata, medical records, and imaging, as opposed to data created by consumers, such as emails and photographs. In a study that investigated the state of privacy practices for health on social networking sites, the authors pointed out that discussions about specific security measures to protect consumers' personal information are lacking [27]. The exchange of healthcare-related data containing personal information will be facilitated in the coming years. Therefore, continuous discussions on privacy are needed to ensure that healthcare consumers can utilize various ICT-based services more safely. In 2013, there were also negative scores for health information and medical technology. Considering that the number of postings related to telemedicine, privacy, and consumer health informatics in 2013 was high, the fact that consumer interest was high does not mean that this topic should be evaluated positively. It is necessary to review the implications of negative emotions for health technology through future research.

Examining social media users' responses to new health technology can be useful to understand the trends in rapidly evolving fields. Recent research has mainly focused on the 'health' keyword itself, which is rather broad. However, this study went one step forward by classifying health technology into subcategories, such as health technology, wearable technology, biotechnology, mobile health, and medical technology. Moreover, health information was classified into the four subcategories of health information, privacy, clinical informatics, and consumer health informatics. By classifying the word above, the significance of this study is that it visualized the trends of health information technology in social media. The impact of technology in the healthcare field has been investigated in many studies by healthcare providers. Also, most of them have focused on the expert's point of view for providing healthcare services to the patients. However, we were able to examine the interests and sentiments of potential healthcare consumers regarding health information technology by analyzing social media data in this study. It would be better to understand the demands for healthcare technology from the consumer's perspective and apply it to new system development. Nevertheless, this study had a few

limitations. First, there were more male Twitter users than female, with the majority of them being young adults and belonging to a single country. Therefore, our findings should be interpreted with caution. Furthermore, we only used subcategory titles and included the contents of the posts in the analysis. In subsequent studies, subcategory keywords should be used to observe the number of retweets (copying and reposting of tweets or replying to tweets), identify important users who post a significant number of tweets on the topic, and conduct analysis based on countries and regions so as to provide more prompt and accurate insights into trends related to healthcare personnel.

Conflict of Interest

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

Acknowledgments

This research supported by the Ministry of Trade, Industry & Energy of Korea under Industrial Technology Innovation Program (No. 10063098, Telepresence robot system development for the support of point-of-care service associated with ICT technology).

ORCID

Jisan Lee (<http://orcid.org/0000-0002-9039-7448>)
 Jeongeun Kim (<http://orcid.org/0000-0002-1249-9908>)
 Yeong Joo Hong (<http://orcid.org/0000-0002-6956-9532>)
 Meihua Piao (<http://orcid.org/0000-0003-2317-2908>)
 Ahjung Byun (<http://orcid.org/0000-0001-5914-1527>)
 Healim Song (<http://orcid.org/0000-0001-5632-9735>)
 Hyeong Suk Lee (<http://orcid.org/0000-0001-7597-3163>)

References

1. Denecke K, Nejd W. How valuable is medical social media data? Content analysis of the medical web. *Inform Sci* 2009;179(12):1870-80.
2. Lober WB, Flowers JL. Consumer empowerment in health care amid the internet and social media. *Semin Oncol Nurs* 2011;27(3):169-82.
3. Lee KH. A study on the introduction of Twitter according to its application types. *Korean Corp Manag Rev* 2011;37(1):279-97.
4. Kwak HW, Lee CH, Park HS, Moon SB. Is Twitter so-

- cial network? From the perspective of the network structure and information propagation. *J Commun Res* 2011;48(1):87-113.
5. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One* 2010;5(11):e14118.
 6. Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: review of the methods used. *JMIR Public Health Surveill* 2018;4(2):e43.
 7. Nawaz MS, Bilal M, Lali MI, Ul Mustafa R, Aslam W, Jajja S. Effectiveness of social media data in health-care communication. *J Med Imaging Health Inform* 2007;7(6):1365-71.
 8. Fisher J, Clayton M. Who gives a tweet: assessing patients' interest in the use of social media for health care. *Worldviews Evid Based Nurs* 2012;9(2):100-8.
 9. Gupta A, Tyagi M, Sharma D. Use of social media marketing in healthcare. *J Health Manag* 2013;15(2):293-302.
 10. Klischewski R. Top-down or bottom-up? How to establish a common ground for semantic interoperability within e-government communities. *Proceedings of 1st International Workshop on E-Government at ICAIL*; 2003 Jun 24; Edinburgh, Scotland. p. 17-26.
 11. Hashtagify [Internet]. [place unknown]: Hashtagify; c2019 [cited at 2019 Feb 14]. Available from: <https://hashtagify.me/>.
 12. Hoyt RE, Sutton M, Yoshihashi A. *Medical informatics: practical guide for the healthcare professional*. 3rd ed. Morrisville (NC): Lulu Press Inc.; 2009.
 13. Esuli A, Sebastiani F. Sentiwordnet: a publicly available lexical resource for opinion mining. *Proceedings of the 5th Conference on Language Resources and Evaluation*; 2006 May 24-26; Genoa, Italy. p. 417-22.
 14. Nakov P, Ritter A, Rosenthal S, Sebastiani F, Stoyanov V. SemEval-2016 Task 4: sentiment analysis in Twitter. *Proceedings of the 10th International Workshop on Semantic Evaluation*; 2016 Jun 16-17; San Diego, CA. p. 1-18.
 15. Annau M. *Online sentiment analysis using R* [Internet]. Wien, Austria: Vienna University of Economics and Business; 2010 [cited at 2019 Mar 24]. Available from: <http://statmath.wu.ac.at/courses/SNLP/Presentations/DA-Sentiment.pdf>.
 16. Silva BM, Rodrigues JJ, de la Torre Diez I, Lopez-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. *J Biomed Inform* 2015;56:265-72.
 17. Dolan B. Report: 13K iPhone consumer health apps in 2012 [Internet]. [place unknown]: MobiHealthNews; 2012 [cited at 2019 Feb 14]. Available from <https://www.mobihealthnews.com/13368/report-13k-iphone-consumer-health-apps-in-2012>.
 18. Aitken M, Gauntlett C. *Patient apps for improved healthcare: from novelty to mainstream*. Parsippany (NJ): IMS Institut e for Healthcare Informatics; 2013.
 19. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. *JAMA Pediatr* 2017;171(5):461-9.
 20. Larsen-Cooper E, Bancroft E, Rajagopal S, O'Toole M, Levin A. Scale matters: a cost-outcome analysis of an m-Health intervention in Malawi. *Telemed J E Health* 2016;22(4):317-24.
 21. Neufeld JD, Doarn CR, Aly R. State policies influence medicare telemedicine utilization. *Telemed J E Health* 2016;22(1):70-4.
 22. Oh JY, Park YT, Jo EC, Kim SM. Current status and progress of telemedicine in Korea and other countries. *Healthc Inform Res* 2015;21(4):239-43.
 23. Khan MT, Khalid S. Sentiment analysis for health care. *Int J Priv Health Inf Manag* 2015;3(2):676-89.
 24. Palomino M, Taylor T, Goker A, Isaacs J, Warber S. The online dissemination of nature-health concepts: lessons from sentiment analysis of social media relating to "Nature-Deficit Disorder". *Int J Environ Res Public Health* 2016;13(1):E142.
 25. Cohen JE. What privacy is for. *Harv Law Rev* 2013;126:1904-33.
 26. Gantz J, Reinsel D. *The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east* [Internet]. Hopkinton (MA): EMC Corp.; 2012 [cited at 2019 Apr 1]. Available from: <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.
 27. Charbonneau DH. Privacy practices of health social networking sites: implications for privacy and data security in online cancer communities. *Comput Inform Nurs* 2016;34(8):355-9.