J Korean Soc Phys Med, 2019; 14(3): 55-62
http://dx.doi.org/10.13066/kspm.2019.14.3.55

Online ISSN: 2287-7215
Print ISSN: 1975-311X

Research Article

Open Access

# Text-Mining of Online Discourse to Characterize the Nature of Pain in Low Back Pain

Young Uk Ryu, PhD, PT†

Department of Physical Therapy, Daegu Catholic University

Received: July 6, 2019 / Revised: July 14, 2019 / Accepted: August 12, 2019


**| Abstract |**

**PURPOSE:** Text-mining has been shown to be useful for understanding the clinical characteristics and patients' concerns regarding a specific disease. Low back pain (LBP) is the most common disease in modern society and has a wide variety of causes and symptoms. On the other hand, it is difficult to understand the clinical characteristics and the needs as well as demands of patients with LBP because of the various clinical characteristics. This study examined online texts on LBP to determine of text-mining can help better understand general characteristics of LBP and its specific elements.

**METHODS:** Online data from www.spine-health.com were used for text-mining. Keyword frequency analysis was performed first on the complete text of postings (full-text analysis). Only the sentences containing the highest frequency word, pain, were selected. Next, texts including the sentences were used to re-analyze the keyword frequency (pain-text analysis).

**RESULTS:** Keyword frequency analysis showed that pain is of utmost concern. Full-text analysis was dominated by structural, pathological, and therapeutic words, whereas pain-text analysis was related mainly to the location and quality of the pain.

**CONCLUSION:** The present study indicated that text-mining for a specific element (keyword) of a particular disease could enhance the understanding of the specific aspect of the disease. This suggests that a consideration of the text source is required when interpreting the results. Clinically, the present results suggest that clinicians pay more attention to the pain a patient is experiencing, and provide information based on medical knowledge.

**Key Words:** Text-mining, Low Back Pain, Pain, Patient Web Portal

## Ⅰ. Introduction

An increasing numbers of people are posting and sharing information about health via the internet. Articles written by professionals and patients are also being produced continuously [1]. More material is being added by blogs, message boards, and social network services (SNS), and the amount of information produced through these and other media is increasingly vast [2,3].

'Text-mining' is a method to analyze the massive amount of information and is used to classify and organize large amounts of text to extract common themes as well as lexical,

†Corresponding Author : Young Uk Ryu
ryuyounguk@gmail.com, http://orcid.org/0000-0003-1601-4477

phrasal, or sentential tendencies [4]. Recently, the text-mining method has spread to health-related issues, where it is currently used widely [5,6]. Recent research has shown that an analysis of the keywords of related texts posted by people online can be used effectively to detect health-related issues [7,8]. Text-mining has also been shown to be useful in understanding the clinical characteristics and patients' concerns regarding a specific disease [9-11].

Park and Ryu [10] used text-mining to analyze 399 samples of memoirs about fibromyalgia that had been posted on the internet. The clinical features of fibromyalgia and the patients' interests were found in the keywords. If text-mining analysis of the full text of postings is useful, such analysis may also be effective for certain elements or characteristics belonging to a specific disease. For example, 'phrase net' analysis shows the link status of a keyword and can demonstrate a connection relation of words that appear with the specific keyword [10,12]. Through this type of analysis, it is possible to identify the words used in a sentence with a specific keyword. On the other hand, this analysis is limited in that it can ignore various words or the importance (frequency) of words that are used with a specific keyword.

This study examined online texts on low back pain (LBP). LBP is one of the most common symptoms in modern society [13], and a huge number of texts left by people with LBP are available online. The causes of LBP are very diverse, and multiple causes may be present [13]. Acute LBP is often caused by sprains of muscles and ligaments [14]. In the case of chronic LBP, herniation, spinal stenosis, zygapophyseal (facet) joint syndrome, and sacroiliac joint dysfunction are known to be major causes; many other causes are also known [15]. For example, nerve compression may be the cause if radicular leg pain is seen in the lower extremities [16].

LBP is the most common disease in modern society with a wide variety of causes and symptoms. The clinical characteristics of LBP and the needs and demands of patients with LBP are difficult to understand because of the various clinical characteristics. General studies have many practical problems in helping to understand LBP and patient with LBP. In particular, the time and expense of interviewing or surveying patients with LBP is enormous. In addition, because it is difficult to target many people, there are restrictions on generalization. As reviewed previously, the analysis of vast amounts of data online can be useful to generalize the characteristics of a disease and the concerns of patients. Therefore, this study examined whether text-mining is effective in helping to better understand the general characteristics of LBP and its specific element.

## Ⅱ. Methods

The online forum data from www.spine-health.com was used. According to alexa.com (http://www.alexa.com/siteinfo/spine-health.com), as of September 11, 2018, this forum was a popular spine health-related website with a global ranking of 15,753 (6,009 in the US). The site is largely composed of content that communicates information on the spine's health and forums where information is exchanged by affected patients. The key search words to find this site include lower back pain (1.55%), sciatica (1.31%), sciatic nerve (1.28%), chiropractor (1.03%), and piriformis syndrome (0.74%). Regarding the demographics of the participants of this site, most were from the United States (51.4%), India (6.9%), United Kingdom (5.4%), Canada (4.2%), and Australia (4.2%).

To obtain the natural language data from the forum, the unstructured data of the postings were first collected through a crawling process. The text data of this study were collected from the forum under the title "lower back pain" (https://www.spine-health.com/forum/categories/lower-back-pain?utm_medium=web&utm_source=sites&utm_campaign=nav). Discussions from the first discussion (June

30, 2008) until December 31, 2017, were collected. In a discussion, if one speaker makes a post, other people can reply with comments to that post. On the other hand, comments were not used in this study, and only the main text of the discussion was extracted. A JSOUP (an open-source Java library) web crawler, which was developed to obtain language data from webpages that use programming languages such as Hypertext Markup Language (HTML) and Java Server Pages (JSP), was used to retrieve discussions from the target forum automatically. The retrieved data were saved into a text file to produce a structured and analyzable dataset. A total of 5,060 discussions were collected.

Next, information, such as keyword lists and word frequency lists, were generated using R, a free software available on the internet that is often used by corpus linguists and social scientists for statistical computing and graphics. A dataset was also made by categorizing the words of the text file into nouns and adjectives using a Stanford Part-Of-Speech Tagger, which assigns part-of-speech tags to individual words. The unnecessary and functional words, such as articles (e.g., 'a', 'the'), conjunctions ('and', 'but', etc.), and pronouns ('I', 'you', etc.) were then eliminated. Number-related words (e.g., 'one', 'two', etc.), calendar-related words (e.g., 'day', 'week', 'month', 'January', 'February', etc.), and common people-related words (e.g., 'anyone', 'everyone', etc.) were also eliminated. After these processes, 550,533 remaining words had LBP-relevant data.

## Ⅲ. Results

Table 1 lists the top 100 words appearing in the online forum based on the keyword frequency (nouns and adjectives). According to the analysis, 'pain' was the most frequently posted word. Pain is the main complaint of LBP patients and is a defining symptom of the disease. This means that pain is the most distinctive issue, suggesting it is the most important word characterizing LBP. Therefore, understanding how the word 'pain' is used in the postings may provide relevant information for understanding the pain of LBP. For the purpose of this study, only sentences containing the word 'pain' were extracted from the whole text data; the frequency of the keywords from those sentences was analyzed (pain-text analysis). A total of 164,740 words were obtained from pain-text analysis. Table 2 lists the top 100 words, and 'pain' is clearly the most frequent word. The two analyses revealed differences in the characteristics of the keywords. These differences represent the most important component of the results of this study. The main results of the two analyses are discussed in the following sections.

### 1) Full-text analysis is dominated by structural, pathological, and therapeutic words

In both keyword frequency analyses (full- and pain-text), 'back' was the next most-frequent word after 'pain.' 'Back' is used to describe the pain and is part of the name of the disease. 'Lower' was also ranked highly in each analysis (Tables 1, 2), indicating that 'lower back pain' is used mostly as the disease name in posts. These results suggest that users who left posts considered their illness to be a major concern. 'Disc' ranked 3rd in full-text analysis (0.75%), but it was 22nd (0.26%) in pain-text analysis (Tables 1, 2). 'Disc' refers to the disc structure between the vertebral bodies and is also a keyword for disc herniation. Similarly, words for structures such as 'spine' and 'lumbar' appeared more often in full-text analysis than in pain-text analysis (Tables 1, 2). Words related to anatomical structures such as 'spinal', 'nerve', 'facet', 'muscle', 'discs', 'disc', 'bone', and 'epidural' ranked in the top 100 keywords only for full-text analysis (Tables 1, 2). Although the pathology-related words 'herniated' and 'sciatica' were found in both analyses, other pathology-related words, such as 'bulge', 'stenosis', 'bulging', 'herniation', and 'injury' were found in the top 100 only after full-text analysis (Tables 1, 2). Words related to

Table 1. Rank of Keyword Frequencies Among 550,533 Words in Full-text Analysis. Words with *Italic* font Represent Words found Only in the Top 100 of Full-text Analysis

| Rnk | Word | Freq | Rnk | Word | Freq | Rnk | Word | Freq | Rnk | Word | Freq |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Pain | 18457 | 26 | Walk | 1148 | 51 | Hip | 812 | 75 | Sleep | 608 |
| 2 | Back | 12609 | 27 | Lumbar | 1101 | 52 | *Problems* | 808 | 77 | *Bulging* | 605 |
| 3 | Disc | 4131 | 28 | Good | 1099 | 53 | Bed | 800 | 78 | *Herniation* | 600 |
| 4 | Lower | 4096 | 29 | Better | 1042 | 54 | *Doctors* | 789 | 79 | Doc | 588 |
| 5 | Now | 3729 | 30 | Normal | 1036 | 55 | *Bulge* | 788 | 80 | Bit | 573 |
| 6 | Surgery | 3207 | 31 | *Spinal* | 1025 | 56 | Meds | 779 | 81 | *Injury* | 570 |
| 7 | Mri | 3186 | 32 | Problem | 1022 | 57 | Low | 760 | 82 | *Body* | 568 |
| 8 | Right | 2970 | 33 | Severe | 1010 | 58 | Sure | 756 | 83 | *Disk* | 563 |
| 9 | Time | 2860 | 34 | Legs | 968 | 59 | *Stenosis* | 754 | 84 | Different | 552 |
| 10 | Help | 2782 | 35 | Herniated | 957 | 60 | Foot | 746 | 85 | Couple | 545 |
| 11 | Left | 2751 | 36 | *Advice* | 955 | 61 | Symptoms | 734 | 86 | Morning | 537 |
| 12 | Doctor | 2334 | 37 | Life | 935 | 62 | Walking | 712 | 87 | Pretty | 536 |
| 13 | Leg | 2298 | 38 | *Facet* | 923 | 63 | Joint | 711 | 88 | Hard | 530 |
| 14 | Work | 1985 | 39 | Next | 921 | 64 | *Post* | 688 | 89 | *Bone* | 529 |
| 15 | Nerve | 1787 | 39 | Therapy | 921 | 65 | *Due* | 671 | 90 | *Treatment* | 524 |
| 16 | Last | 1762 | 41 | Muscle | 907 | 66 | Numbness | 666 | 91 | Chronic | 523 |
| 17 | Spine | 1695 | 42 | Physical | 906 | 67 | Point | 664 | 91 | Painful | 523 |
| 18 | Side | 1545 | 43 | Injections | 905 | 68 | *Issue* | 663 | 93 | *Looking* | 514 |
| 19 | Worse | 1529 | 44 | Feeling | 903 | 69 | Past | 636 | 94 | *Central* | 513 |
| 20 | Long | 1525 | 45 | Way | 888 | 70 | Sciatica | 626 | 95 | *Home* | 509 |
| 21 | Bad | 1416 | 46 | Area | 884 | 71 | Night | 624 | 96 | Working | 507 |
| 22 | Old | 1328 | 47 | Surgeon | 863 | 72 | Injection | 614 | 97 | *Narrowing* | 503 |
| 23 | Fusion | 1311 | 48 | Little | 830 | 73 | Cause | 610 | 98 | *Epidural* | 501 |
| 24 | New | 1290 | 49 | Level | 828 | 74 | Many | 609 | 99 | *Job* | 498 |
| 25 | *Mild* | 1160 | 49 | Relief | 828 | 75 | *Discs* | 608 | 100 | Experience | 497 |

treatment and diagnosis were also found in both analyses. Words related to therapeutic methods such as ‘surgery’, ‘fusion’, ‘therapy’, and ‘injections’ were found at a relatively higher rate in full-text analysis compared to pain-text analysis (Tables 1, 2). ‘MRI,’ a diagnostic method, was also found more frequently in full-text analysis than pain-text analysis (Tables 1, 2).

**2) Pain-text analysis was predominantly related to the location and quality of the pain**

Both analyses contained words related to body parts as well as the words ‘leg,’ ‘legs,’ ‘hip,’ and ‘foot,’ all of which occurred at a relatively higher rate in pain-text analysis than in full-text analysis. (Tables 1, 2). In addition, ‘knee,’ ‘feet,’ ‘butt,’ thigh,’ ‘hips,’ buttocks,’ and ‘buttock’ were found only in the top 100 of pain-text analysis (Table 2).

Table 2. Rank of Keyword Frequencies Among 164,740 Words in Pain-text Analysis. Words with *Italic* font Represent Words found Only in the Top 100 of Pain-text Analysis

| Rnk | Word | Freq | Rnk | Word | Freq | Rnk | Word | Freq | Rnk | Word | Freq |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Pain | 16662 | 26 | Spine | 424 | 51 | Point | 235 | 76 | *Medication* | 177 |
| 2 | Back | 5595 | 27 | Low | 415 | 52 | Level | 228 | 76 | Pretty | 177 |
| 3 | Lower | 2076 | 28 | Area | 383 | 52 | Sleep | 228 | 78 | Doc | 172 |
| 4 | Leg | 1381 | 29 | Chronic | 352 | 54 | *Need* | 227 | 79 | Surgeon | 171 |
| 5 | Now | 1365 | 30 | Numbness | 346 | 55 | Walking | 225 | 80 | Injection | 169 |
| 6 | Left | 1166 | 31 | *Management* | 345 | 56 | Morning | 217 | 81 | Fusion | 168 |
| 7 | Right | 1147 | 32 | *Constant* | 335 | 57 | Good | 215 | 82 | *Extreme* | 167 |
| 8 | Time | 848 | 33 | *Sharp* | 328 | 58 | *Standing* | 213 | 83 | *Trying* | 165 |
| 9 | Worse | 695 | 34 | Feeling | 325 | 59 | Next | 211 | 84 | Hard | 163 |
| 10 | Surgery | 686 | 35 | Foot | 324 | 59 | *Sciatic* | 211 | 85 | *Suffering* | 162 |
| 11 | Side | 676 | 36 | Relief | 320 | 61 | Physical | 209 | 86 | Painful | 159 |
| 12 | Doctor | 609 | 37 | Old | 315 | 61 | Problem | 209 | 87 | *Butt* | 158 |
| 13 | Bad | 605 | 38 | Life | 312 | 63 | Couple | 201 | 88 | *Thigh* | 157 |
| 14 | Severe | 595 | 39 | *Pains* | 307 | 64 | Normal | 200 | 89 | Different | 155 |
| 15 | Last | 587 | 40 | Muscle | 305 | 65 | Therapy | 197 | 90 | *Excruciating* | 152 |
| 16 | Help | 574 | 41 | Better | 301 | 66 | *Burning* | 193 | 91 | *Hips* | 151 |
| 17 | Legs | 501 | 42 | Sciatica | 284 | 67 | Symptoms | 191 | 91 | *Live* | 151 |
| 18 | Nerve | 493 | 43 | Little | 282 | 68 | *Free* | 189 | 93 | Herniated | 150 |
| 19 | Work | 478 | 44 | Cause | 277 | 69 | *Shooting* | 188 | 94 | *Intense* | 147 |
| 20 | Meds | 464 | 45 | Bed | 271 | 70 | *Knee* | 186 | 95 | Sure | 146 |
| 21 | Walk | 457 | 46 | Way | 262 | 71 | Bit | 183 | 96 | *Buttocks* | 145 |
| 22 | Long | 437 | 47 | New | 258 | 71 | *Experiencing* | 183 | 97 | Many | 144 |
| 23 | Hip | 434 | 47 | Past | 258 | 73 | Joint | 181 | 98 | Experience | 141 |
| 24 | Disc | 433 | 49 | Night | 254 | 74 | *Feet* | 179 | 99 | *Issues* | 140 |
| 25 | MRI | 428 | 50 | Injections | 242 | 75 | Lumbar | 177 | 100 | *Buttock* | 139 |

The words for these body parts correspond to the areas of back pain or where back pain radiates. No words related to head or upper limbs were found in the top 100 of either analysis. Words related to the quality or appearance of pain, such as ‘severe’, ‘numbness’, ‘chronic’, and ‘painful’ were found in both analyses but were observed at a relatively higher rate in pain-text analysis than in full-text analysis (Tables 1, 2). Other words related to the quality and appearance of pain, such as ‘constant’, ‘sharp’, ‘burning’, ‘extreme’, ‘excruciating’, and ‘intense’ were also found among the top 100 of pain-text analysis (Table 2).

## Ⅳ. Discussion

This study used a text-mining method to analyze 5,060 samples of discussions of the specific disease LBP that were posted in the forum www.spine-health.com. First, a keyword frequency analysis of the full text of the posts

was performed, and the most frequent word was ‘pain’. For the main purpose of the study, the same keyword frequency analysis was performed by only extracting the sentences that included the word ‘pain’. This section discusses the main results of this study.

First, the specific words for the anatomical structures, pathologies, and pain sites associated with LBP were abundant in both analyses. This suggests that these keywords represent the general clinical features of LBP. Several studies analyzing online illness experiences have already shown this trend [10,11,17,18]. Because illness experiences are disease specific, the keywords extracted from such texts often represent a number of clinical features of the disease.

This study also found, however, that there were fewer keywords indicating the psychological status and social life of LBP patients than indicating the clinical features. Although ‘work’ and ‘life’ were observed as keywords, other analyses would be necessary to know how LBP is related to the patients’ work and life (analysis beyond the scope of this paper). Park and Ryu [10] analyzed the memoirs of patients with fibromyalgia and found that the most commonly used keywords were not only clinical words but also a number of desires as well as work-, occupation-, and human-related keywords. In contrast, the keywords in the present study were overwhelmingly related to the clinical disease characteristics. This might be because of the characteristics of the websites. The online texts used by Park and Ryu were extracted from sites that share stories or experiences related to a specific topic. In contrast, the postings used in this study were text from a discussion menu that shared questions and information regarding the diseases and symptoms on a site where medical information related to the spine was shared. In addition, as mentioned earlier, the top terms for visiting this site were also clinical words. Therefore, when analyzing text, the nature of an online site where texts are posted should be a factor to consider when interpreting the results [19].

The most important finding of this study was that there were differences in the nature of the keywords found in the full- and pain-text analyses. Keywords related to anatomical structures, pathology, diagnosis, and treatment were relatively more common in full-text analysis, whereas the most-common keywords in pain-text analysis were related to the quality, pattern, and location of pain. Because full-text analysis showed the general features of the disease, it is believed that various but general keywords related to LBP were observed. On the other hand, pain-text analysis provided information on ‘pain’ as a result of the ‘sub-analysis’ of LBP. These results suggest the possibility of sub-analysis regarding an element (keyword) that may provide information on the desired part of the entire text to be viewed selectively [20].

One thing to mention in pain-text analysis was the keywords regarding the intensity of pain, implying that the pain of those who posted was extreme, intense, and unbearable. In particular, ‘excruciating’ was the expression of the highest pain level according to the McGill Pain Questionnaire (MPQ) scale [21]. These observations suggest that patients with severe pain used the website more often than those with mild pain. When interpreting the results of online texts that are not controlled by the users, the main users can be estimated based on the results, which show their interests, indicating the possibility of using such results to provide certain services to them.

The results of this study suggest to clinicians what areas they should be concerned about when dealing with LBP. First, a more attentive approach to the pain a patient is experiencing is needed. The main concern for patients with LBP was pain. In addition, patients who left their own experiences online apparently had unusual pain. These results suggest that clinicians should be more interested in the pain their patients are experiencing and allow them to talk more about their own pain. This study also found that patients with LBP want to ask questions and know about their illness even online, meaning that they are

looking for more help. Finally, the results of this study also indicated the information that LBP patients should convey. Keyword analysis showed that the words left by patients were disease-oriented. This suggests that back pain patients want to know more about their clinical situation. That is to say that when we treat patients with are treated LBP in a clinic, it is better to inform them based on medical knowledge.

Although some important results were obtained from this study, there were limitations to this research method. First, the study could not control factors such as the characteristics (age, gender, educational level, social/ occupational status, etc.), disease state (diagnosis, duration of illness, pain level, etc.), and other environmental factors (culture, etc.) of the people who left postings. This inability was attributed to the processing of large amounts of unspecified data, which may be a limit of text-mining techniques that process significant amounts of data [10]. Second, although the data are uncontrolled, the study results can be biased by several factors, such as the nature of the website and the pathological status of the people posting the text. Therefore, it is important to understand the background of the source when analyzing and interpreting uncontrolled text data. Clinically, these results suggest that clinicians should pay more attention to the pain a patient is experiencing, and provide information based on their medical knowledge.

## Ⅴ. Conclusion

The results of this study confirmed that various clinical features related to LBP were well reflected in the keywords. More information on the characteristics of pain in LBP was found when sentences containing the most frequent keyword “pain” were analyzed for their keyword frequency than in full-text analysis. These results showed that text-mining for a specific element (keyword) of a particular disease could identify and enhance the understanding of that factor. In addition, when analyzing text online, the nature of the website where the text was posted can affect the results. This suggests that a consideration of the text source is required when interpreting the results.

## Acknowledgements

This study was supported by research grants (#20171131) from Daegu Catholic University in 2017.

## References

[1] Mazzoni D, Cicognani E. Sharing experiences and social support requests in an Internet forum for patients with systemic lupus erythematosus. J Health Psychol. 2014; 19(5):689-96.

[2] Allen C, Vassilev I, Kennedy A, et al. Long-term condition self-management support in online communities: a meta-synthesis of qualitative papers. J Med Internet Res. 2016;18(3):e61.

[3] Kingod N, Cleal B, Wahlberg A, et al. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res. 2017(1);27:89-99.

[4] Feldman R, Sanger J. The text mining handbook: advanced approaches in analyzing unstructured data. New York (NY): Cambridge University Press. 2007.

[5] Bellika J, Bravo-Salgado A, Brezovan M, et al. Text mining of web-based medical content (Vol. 1). Berlin: Walter de Gruyter GmbH & Co KG. 2014.

[6] Dreisbach C, Koleck TA, Bourne PE, et al. systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Inform. 2019;125:37-46.

[7] Lu Y, Zhang P, Liu J, et al. Health-related hot topic detection in online communities using text clustering. Plos one. 2013;8:e56221.

[8] Lazard AJ, Scheinfeld E, Bernhardt JM, et al. Detecting

themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. Am J Infect Control. 2015;43(3):1109-11.

[9] Vasconcellos-Silva PR, Carvalho D, Lucena C. Word frequency and content analysis approach to identify demand patterns in a virtual community of carriers of hepatitis C. Interact J Med Res. 2013;2(2):e12.

[10] Park J, Ryu YU. Online discourse on fibromyalgia: text-mining to identify clinical distinction and patient concerns. Med Sci Monitor. 2014;20:1858-64.

[11] Matsuda S, Aoki K, Tomizawa S, et al. Analysis of patient narratives in disease blogs on the internet: an exploratory study of social pharmacovigilance. JMIR Pub Health Sur. 2017;3(1):e10.

[12] Van Eck NJ, Waltman L. Text mining and visualization using VOSviewer. ISSI Newsletter. 2011;7:50-4.

[13] Maher C, Underwood M, Buchbinder R. Non-specific low back pain. The Lancet. 2017;389(10070):736-47.

[14] Borenstein DG. Epidemiology, etiology, diagnostic evaluation, and treatment of low back pain. Curr Opin Rheumatology. 2001;13(2):128-34.

[15] DePalma MJ, Ketchum JM, Saullo T. What is the source of chronic low back pain and does age play a role? Pain medicine. 2011;12(2):224-33.

[16] Koes BW, Van Tulder M, Thomas S. Diagnosis and treatment of low back pain. Bmj. 2006;332(7555):1430-4.

[17] Gupta S, MacLean DL, Heer J, et al. Induced lexico-syntactic patterns improve information extraction from online medical forums. J Am Med Inform Assn. 2014;21(5):902-9.

[18] Sunkureddi P, Gibson D, Doogan S, et al. Using self-reported patient experiences to understand patient burden: learnings from digital patient communities in ankylosing spondylitis. Adv Ther. 2018;35(3):424-37.

[19] Herring SC. Computer-mediated discourse analysis: An approach to researching online behavior. In: Designing for Virtual Communities in the Service of Learning. New York (NY): Cambridge University Press.

[20] Tighe PJ, Goldsmith RC, Gravenstein M, et al. The painful tweet: text, sentiment, and community structure analyses of tweets pertaining to pain. J Med Internet Res. 2015;17(4):e84.

[21] Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain. 1975;1(3):277-99.