

Original Paper

Investigating Individuals' Perceptions Regarding the Context Around the Low Back Pain Experience: Topic Modeling Analysis of Twitter Data

Robert^{1*}; Pari Delir Haghighi^{1*}, PhD; Frada Burstein^{1*}, PhD; Donna Urquhart^{2*}, PhD; Flavia Cicuttini^{2*}, PhD, MD

¹Department of Human-Centred Computing, Faculty of Information Technology, Monash University, Caulfield East, Australia

²Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

* all authors contributed equally

Corresponding Author:

Pari Delir Haghighi, PhD

Department of Human-Centred Computing

Faculty of Information Technology

Monash University

Building H, Level 6

900 Dandenong Road

Caulfield East, 3145

Australia

Phone: 61 99032355

Email: pari.delirhaghighi@monash.edu

Abstract

Background: Low back pain (LBP) remains the leading cause of disability worldwide. A better understanding of the beliefs regarding LBP and impact of LBP on the individual is important in order to improve outcomes. Although personal experiences of LBP have traditionally been explored through qualitative studies, social media allows access to data from a large, heterogenous, and geographically distributed population, which is not possible using traditional qualitative or quantitative methods. As data on social media sites are collected in an unsolicited manner, individuals are more likely to express their views and emotions freely and in an unconstrained manner as compared to traditional data collection methods. Thus, content analysis of social media provides a novel approach to understanding how problems such as LBP are perceived by those who experience it and its impact.

Objective: The objective of this study was to identify contextual variables of the LBP experience from a first-person perspective to provide insights into individuals' beliefs and perceptions.

Methods: We analyzed 896,867 cleaned tweets about LBP between January 1, 2014, and December 31, 2018. We tested and compared latent Dirichlet allocation (LDA), Dirichlet multinomial mixture (DMM), GPU-DMM, biterm topic model, and nonnegative matrix factorization for identifying topics associated with tweets. A coherence score was determined to identify the best model. Two domain experts independently performed qualitative content analysis of the topics with the strongest coherence score and grouped them into contextual categories. The experts met and reconciled any differences and developed the final labels.

Results: LDA outperformed all other algorithms, resulting in the highest coherence score. The best model was LDA with 60 topics, with a coherence score of 0.562. The 60 topics were grouped into 19 contextual categories. "Emotion and beliefs" had the largest proportion of total tweets (157,563/896,867, 17.6%), followed by "physical activity" (124,251/896,867, 13.85%) and "daily life" (80,730/896,867, 9%), while "food and drink," "weather," and "not being understood" had the smallest proportions (11,551/896,867, 1.29%; 10,109/896,867, 1.13%; and 9180/896,867, 1.02%, respectively). Of the 11 topics within "emotion and beliefs," 113,562/157,563 (72%) had negative sentiment.

Conclusions: The content analysis of tweets in the area of LBP identified common themes that are consistent with findings from conventional qualitative studies but provide a more granular view of individuals' perspectives related to LBP. This understanding has the potential to assist with developing more effective and personalized models of care to improve outcomes in those with LBP.

(*J Med Internet Res* 2021;23(12):e26093) doi: [10.2196/26093](https://doi.org/10.2196/26093)

KEYWORDS

low back pain; Twitter; content analysis; social media; topic modeling; patient-centered approach; pain experience; context of pain

Introduction

Low back pain (LBP) is the leading cause of disability worldwide [1,2]. Approximately 50%-80% of adults experience LBP at least once in their lives [3] and it is a leading cause of work absence and limits physical activities, posing a large economic burden [1,4]. In the United States, the total cost associated with LBP exceeds US \$100 billion per year [5,6]. It is also a significant contributor to the current global epidemic of narcotic prescriptions [7].

Optimizing management of conditions such as LBP requires consumers to be engaged in their care. To enable this, health care providers need to have an understanding of the full context of the condition from the consumer perspective. “Contextual variables” here refer to any type of useful information about the context of an individual’s pain experience, such as physical, emotional, social, and/or occupational variables [8]. A better understanding of the contextual variables of individuals with LBP could provide clinicians and health providers with an alternative insight into patients’ concerns, beliefs, and expectations, and has the potential to improve outcomes in LBP [9]. Although there have been many studies examining individuals’ beliefs about LBP, patients’ perspectives remain inadequately understood [10]. Although qualitative studies—including systematic scoping reviews—investigating patients’ needs and expectations have been conducted, these have largely focused on a single topic, such as health care, with the findings extrapolated from heterogeneous studies that are of poor quality [11-13]. A further limitation of current approaches is that most traditional data collection methods use predefined frameworks that have the potential to constrain responses. For instance, validated questionnaires that provide statements about back pain and its consequences (such as “back pain must be rested”) and require the respondent to indicate their level of agreement on a scale are commonly used [12,13]. Moreover, for logistical and methodological reasons, many studies restrict the selection of populations to be studied.

With the current advances in online and web technologies, social media has emerged as a new and rich source of first-person health care data [14-16]. Social media platforms provide an opportunity to rapidly collect data from a larger and more diverse population in a cost-efficient manner. Health-related topics are commonly discussed on Twitter [17-19], a microblogging social media site [20]. A systematic review conducted by Sinnenberg et al [21] found six main uses of Twitter in health research: content analysis, surveillance, engagement, recruitment, intervention, and network analysis. Aggregation and analysis of large volumes of health-related data from social media sites could provide valuable information

from a first-person point of view [14,22]. In the area of LBP, this approach could be used to investigate individuals’ perspectives and the context around the LBP experience [15,23]. We hypothesize that the detected topics identify specific contexts around the LBP experience in individuals. Thus, the aim of this study was to identify contextual variables of the LBP experience from a first-person perspective using a topic modeling approach of Twitter data to provide useful insights into individuals’ beliefs and perceptions. This has the potential to inform more effective patient-centered approaches to the management of LBP.

Methods**Study Approach**

Our study approach was to undertake content analysis of Twitter data by applying topic modeling. Content analysis is a widely used technique for qualitative research [24] that enables studying patient experience in depth by deriving topics of interest from text documents [14,25].

Twitter Data

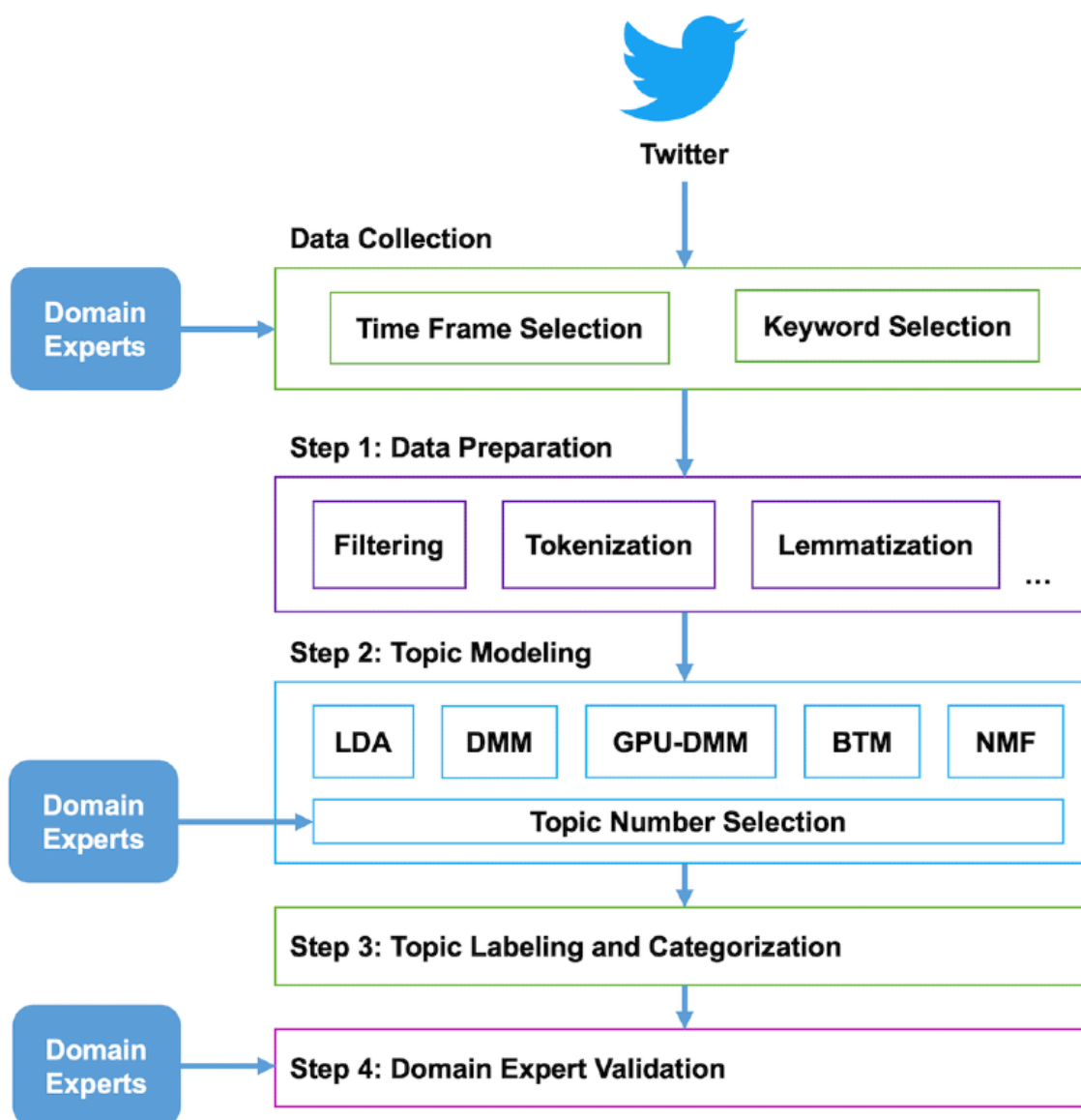
Twitter was used as the data source rather than other social media platforms, blog posts, or news articles because individuals use this platform for expressing and sharing their feelings and opinions on health-related topics by posting short messages that can be easily collected through application programming interfaces (APIs) or other open sources [14-17,26]. We used an open-source Twitter scraping tool called Twint [27] for collecting tweets related to LBP that were written in English. Twint enables the collecting of Twitter data without using Twitter’s API through its publicly available library in the Python programming language [27,28]. We collected tweets posted between January 1, 2014, and December 31, 2018 (inclusive). The time frame of 5 years was selected to provide us with sufficient data to examine the patterns in emerging topics and the number of tweets over time. Since the number of active users on the social media platform increased in recent years and we needed a large volume of data for topic modeling, we did not consider tweets posted before 2014. We selected the search keywords based on 3 studies on back pain [15,29,30]. These are detailed in Table 1. Search keywords were verified by our domain experts (FC, a rheumatologist; DU, a physiotherapist) who have extensive research and clinical expertise in the area of LBP. Selecting search keywords and an appropriate time frame are important considerations in the data collection process. The Monash University Human Research Ethics Committee approved this study (project ID 19738).

Our data processing and analysis consisted of 4 steps (see Figure 1).

Table 1. Keywords used to search tweets related to low back pain.

Source	Study purpose	Keywords	Total, n
Lee et al, 2016 [15]	To quantify the risks associated with a new tweet about back pain	“painful back,” “sore back,” “back started hurting,” “buggered my back,” “hurt my back,” “I’ve got backache,” “injured my back,” “my back hurts,” “I’ve got back pain,” “pain in my back,” “put my back out,” “my back is killing me”	12
Ahlwardt et al, 2014 [30]	To compare self-reported toothache experiences in tweets with those of backache, earache, and headache	“backache,” “back ache,” “back aches,” “back hurt,” “back hurting,” “back hurts,” “back killin’,” “back killing,” “back pain,” “back sore”	10
Campbell et al, 2013 [29]	A systematic review to study the influence of employment social support in nonspecific back pain	“lumbago,” “backache,” “back ache,” “back pain,” “low back ache,” “low back pain,” “lower back pains”	7

Figure 1. The overall data analysis workflow. The analysis consists of four steps: (1) data preprocessing, (2) thematic analysis using topic modelling, (3) topic labeling and categorization, and (4) domain expert validation. BTM: biterm topic model; DMM: Dirichlet multinomial mixture; GPU-DMM: General Pólya Urn Dirichlet Multinomial Mixture; LDA: latent Dirichlet allocation; NMF: nonnegative matrix factorization.



Step 1: Data Preprocessing

We removed duplicates, retweets, URLs, and tweets related to marketing and advertisements, which reduced the data set from 7,892,210 to 2,825,645. We filtered the data further by removing tweets that did not contain first person pronouns [15]. As a result, the remaining data set size was 2,010,295.

We replaced contractions with their expanded forms (eg, “didn’t” to “did not”). We converted the HTML characters to ASCII characters and removed hashtags, Unicode strings (eg, “\u2026”), numbers, and punctuation. We replaced abbreviations, elongated words (eg, “goood” to “good”), and emoticons and emojis with their equivalent English expressions. We then performed spelling correction, lowercasing, tokenization, and lemmatization, created n-grams, removed stop words (eg, common terms such as “the” and “is”). We again removed the duplicates and the remaining data set was 1,249,576 tweets.

After completing the abovementioned steps, we excluded tweets with less than three words because in topic modeling, the document size is important to achieve high accuracy [31]. This reduced the data set to 896,867 tweets.

Step 2: Topic Modeling

Topic modeling is a technique used to provide a summary of a large collection of documents by extracting “topics” that represent the dominant themes [32]. It allows the uncovering of common, hidden themes from a corpus of text documents such as tweets. We tested 5 well-established topic modeling algorithms for detecting topics in a text-based corpus, namely latent Dirichlet allocation (LDA) [33], Dirichlet multinomial mixture (DMM) [34], GPU-DMM [35], biterm topic model (BTM) [36], and nonnegative matrix factorization (NMF) [37].

LDA is a generative probabilistic model that assumes each document can be represented by distribution over topics and each topic by distribution over words [33,38]. DMM is also a generative model but it assumes that each document is associated with one single topic [34,39]. GPU-DMM is an extended method of DMM that considers semantic similarity between words to provide semantic understanding of text documents and improve topic inference [35,40]. BTM uncovers topics by modeling the word co-occurrence patterns (ie, biterns) rather than using the document-level word co-occurrences [36,41]. NMF is able to learn the latent features in data using a nonnegative representation and improve latent semantic topic identification [37,42,43].

To use these models (except for NMF), we used a Java-based open-source library for short text topic modeling algorithms called STTM (version 1.8) [44], whereas for NMF we used the sklearn [45] library. For each approach, we performed a series of experiments ranging from 5 topics to 200 topics. We applied the 5 algorithms to the 896,867 tweets to determine the best model and the optimal number of topics.

Choosing the right number of topics is a crucial step in topic modeling because it can affect the accuracy of results. The quantitative approach computes the coherence score and perplexity, which helps in determining the optimal number of

topics [46]. The coherence score measures the sum of the pairwise word-similarity scores of the words in the topic, using the pointwise mutual information (PMI) score [47]. Best collocation pairs usually have a high PMI. On the other hand, the qualitative approach requires humans and domain experts to examine the topics. Human judgment is extremely important because topic modeling uses a form of unsupervised learning.

As a quantitative approach, we calculated the coherence score of each model on different numbers of topics ranging from 5 to 200, based on the PMI score [47,48]. The coherence score was used to evaluate the quality of the topic-word distribution. LDA outperformed other approaches (ie, DMM, GPU-DMM, BTM, and NMF).

Additionally, we used a qualitative approach to select the most representative topics. We manually examined the topics, their top 20 terms, and a random sample of tweets in each topic. We also created a word cloud for each topic and evaluated word clouds and their sample tweets. We identified the number of topics that provided us with distinct and meaningful topics; if we exceeded this number of topics, we started to notice an increase in duplicates and overlapping topics. We used both quantitative and qualitative approaches to select the optimal number of topics.

Step 3: Topic Labeling and Categorization

Topic labeling is a process of representing the meaning of a topic by assigning each topic a descriptive word or phrase [49]. Although automatic labeling approaches can reduce costs and time required, they are not able to achieve high semantic validity and accuracy [50,51]. In our study, we used the “eyeballing” method, which refers to reading and inspecting the top words in a topic and manually assigning a label [50]. We made sure that the results met the requirements of a “good” label: (1) semantically relevant, (2) meaningful, (3) representative, (4) adequate, and (5) understandable [34,49].

LDA assumes that each document (tweet) is a mix of topics with different proportions [33]. We were interested to examine tweets based on their dominant topic to gain a better understanding of the frequency of topics across all tweets. Therefore, we performed further analysis, and used the label of the dominant topic to represent each tweet, and then calculated the total number of tweets per topic.

To improve the results of thematic analysis, low-order topics can be grouped under broad, higher-order categories [52]. Higher-level categories can provide a better overview of the key topics discussed by individuals. To this end, after manual topic labeling, we performed topic categorization and assigned a category label to the topics that represented common themes. To identify the important and widely discussed categories, we then calculated the percentage of all tweets that corresponded to each individual category.

Step 4: Domain Expert Validation

Two domain experts (FC, a rheumatologist; DU, a physiotherapist), actively working clinically and researchers in the area of LBP, independently examined the selected topics from the previous step where each topic included the top 20

words to determine face validity. As previously described, in topic modeling, the top words of each topic provide the description of that topic, thereby assisting the domain experts with inferring its meaning [49]. The experts then met to reconcile any differences and develop the final labels.

Results

Overview

The total number of collected tweets about LBP was 7,892,210 from 2,420,258 unique users from 2014 to 2018. The average number of words in each tweet increased from 2017 onward ([Multimedia Appendix 1](#)), in line with Twitter doubling the character limit of tweets from 140 to 280 characters as of November 2017 [53].

Step 1: Data Preprocessing

After performing comprehensive data preprocessing, the final number of retained tweets was 896,867, which represents 11% (896,867/7,892,210) of the original raw data we collected, with a vocabulary size of 29,539. The minimum length of tweets was 4 words and the maximum length was 20 words.

Step 2: Topic Modeling

After testing 5 topic modeling algorithms and the number of topics based on the coherence score and our manual examination, we selected the best model that included 60 topics, detected from 896,867 self-reported tweets about LBP. [Multimedia Appendix 2](#) shows the coherence score of different models with a different number of topics ranging from 5 to 200. The best model was the LDA model with 60 topics, which had

a coherence score of 0.562. [Multimedia Appendix 3](#) shows the best model selected with 60 topics and their top 20 terms.

Step 3: Topic Labeling and Categorization

The 60 topics were examined and manually given a topic label. The common and duplicate labels were then grouped into higher-order categories. Word clouds for the two categories of “pain regions” and “sleep” after combining the related topics are provided in [Multimedia Appendix 4](#). The prevalence of the 60 manually labeled topics is presented in [Multimedia Appendix 5](#).

Step 4: Domain Expert Validation

Independent examination of selected topics by two domain experts and reconciliation of any differences resulted in 19 contextual categories, with details presented in [Multimedia Appendix 6](#). The total number of tweets within each of 19 contextual categories is presented in [Figure 2](#), with more details in [Multimedia Appendix 7](#). The “emotion and beliefs” category had the largest proportion of the total tweets, followed by “physical activity” and “daily life.” The lowest proportion of tweets belonged to the categories of “food and drink,” “weather,” and “not being understood.”

The proportion of tweets for each higher-level category over the years showed that all 19 categories had been discussed by individuals with relatively similar frequency every year (see [Figure 3](#)). However, the proportion of “emotion and beliefs” decreased from 2014 to 2018. The number of tweets about other categories, such as “aggravating factors” and “symptoms,” increased over that time period. An example of a tweet for each category is presented in [Table 2](#) to illustrate the type of personal point of view related to each category.

Figure 2. The 19 categories and their proportions based on all tweets posted from 2014 to 2018.

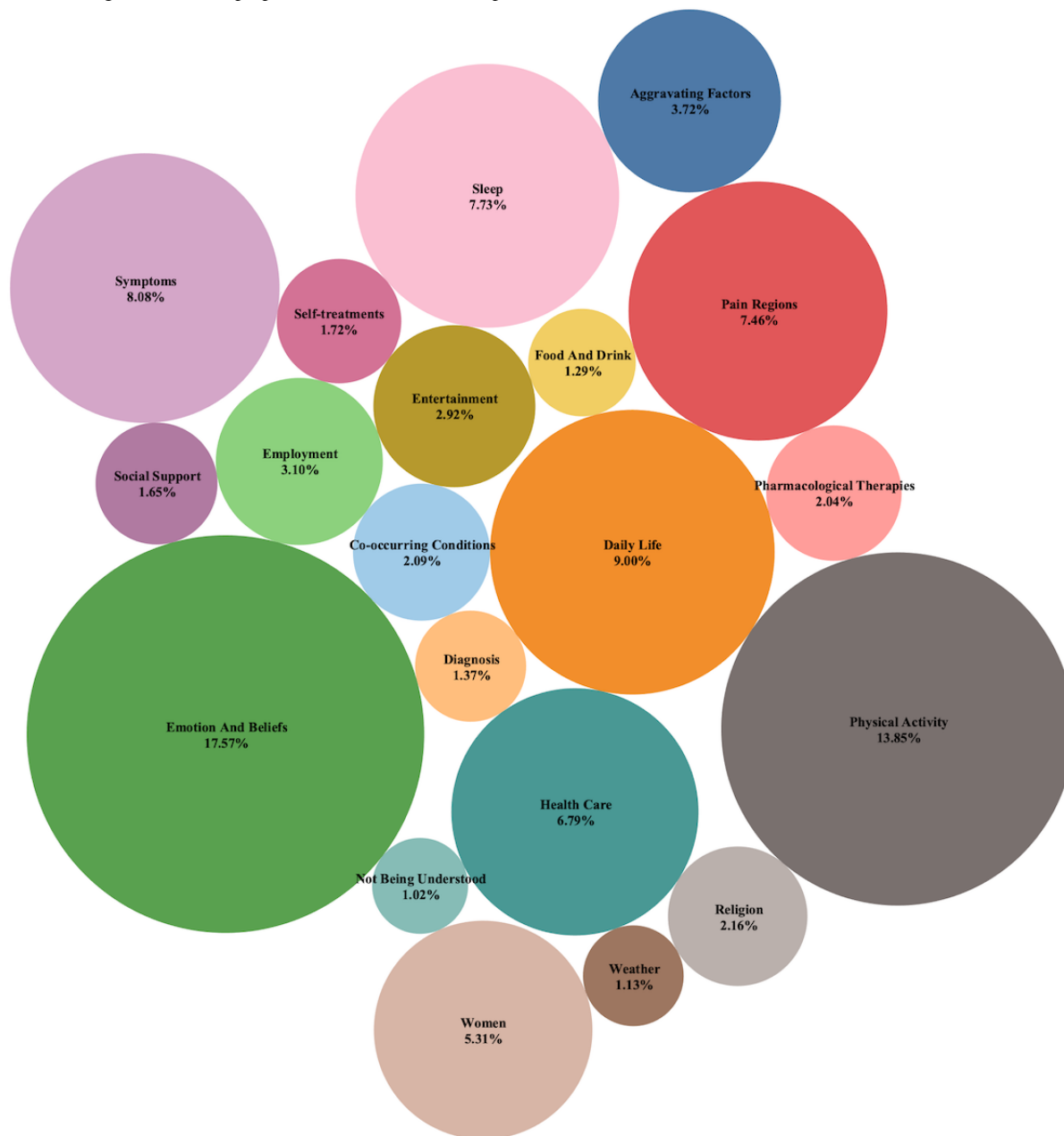


Figure 3. The proportions of 19 categories based on the dominant topic per year.



Table 2. An example of tweets for each contextual category.

Categories	Examples of tweets
Emotion and beliefs	<ul style="list-style-type: none"> My back hurts, feeling sad because I wanna get up and do something ! I hate staying in bed :(
Physical activity	<ul style="list-style-type: none"> I did 6 miles on my exercise bike yesterday, felt really pleased with myself, and ate healthy. My back hurts today
Daily life symptoms	<ul style="list-style-type: none"> So my back hurts like hell and I can hardly sit here and do my hair. I hate it when my lower back hurts and sends shooting pains down my legs, making them ache and throb. Ugh.
Sleep	<ul style="list-style-type: none"> Every time I sleep in my sis guest bedroom my back hurts, that bed is not comfortable. I'd prolly be better off sleeping on the floor
Pain regions	<ul style="list-style-type: none"> today is not a good day. my back hurts, my shoulder hurts, my elbow is tingly, a little numb down to my hand and to top it off now my left knee hurts a little.
Health care	<ul style="list-style-type: none"> So I have found one good physio and one good chiropracter, both same price, who would you see if you had lower back pain?
Women	<ul style="list-style-type: none"> Being pregnant is literally taking everything out of me. I'm exhausted, my back is killing me and I stay moody...
Aggravating factors	<ul style="list-style-type: none"> Yesterday I tried doing a back flip on my trampoline. Now, every time I walk my back hurts. When I did the back flip I landed on my head.
Employment	<ul style="list-style-type: none"> Hurt my back at work yesterday and I'm working a full 12 hours tomorrow without getting paid. Lovin life right now.
Entertainment	<ul style="list-style-type: none"> Watching Cirque Du Soleil: Michael Jackson my back hurts just from watching it
Religion	<ul style="list-style-type: none"> Testimony Time! i want to give God the glory for healing me from a severe back pain
Co-occurring conditions	<ul style="list-style-type: none"> I don't know if my back pain is causing depression or my depression is causing back pain...
Pharmacological therapies	<ul style="list-style-type: none"> I just took my very first Oxycodone for lower back pain. I think I'm in love. It didn't just kill the pain. It assassinated it.
Self-treatments	<ul style="list-style-type: none"> Coconut oil epsom salt & vapor bath oil just soothed my back pain away
Social support	<ul style="list-style-type: none"> Told mom my back hurts she offered to rub my feet an back I have the best mom ever
Food and drink	<ul style="list-style-type: none"> my back is killing me cant get out ov bed but need coffee
Weather	<ul style="list-style-type: none"> I love cold weather but it's really not helping with my back pain. Where is that warm summer weather atttttt.
Not being understood	<ul style="list-style-type: none"> OMG no one understands the pain I'm in right now. My back is killing me.

Discussion

Principal Results

In this study, we identified 60 specific topics from 896,867 tweets about LBP and grouped them into 19 categories that relate to contextual variables of LBP. The top category was “emotion and beliefs,” with 157,563/896,867 tweets (17.6%), followed by “physical activity” (124,251/896,867, 13.85%) and “daily life” (80,730/896,867, 9%), while “food and drink,” “weather,” and “not being understood” had the lowest proportions of tweets (11,551/896,867, 1.29%; 10,109/896,867, 1.13%; and 9180/896,867, 1.02%, respectively). There were 11 topics within the category of “emotion and beliefs”; of 157,563 tweets in this category, 113,562 (72%) expressed negative sentiment. Our results were consistent with the general findings from traditional study methods in the area of LBP but provided

more in-depth detail on the context of LBP from the individual perspective.

Comparison With Prior Work

Our study examined contextual variables to provide a novel insight into first-person perspectives of the LBP experience and confirmed the broad areas that have previously been identified using more traditional data collection methods from qualitative and quantitative studies. For example, psychosocial factors have an important role in LBP [54] and, from our analysis of tweets, “emotion and beliefs” was the most common topic we identified, with 157,563 of 896,867 tweets. This is consistent with LBP being widely recognized as a biopsychosocial condition, and growing evidence to show that psychological factors, such as beliefs and emotions, play an important role [55]. For instance, systematic reviews have highlighted that beliefs about back pain and negative consequences resulting from these beliefs are

common across different countries and populations [56], and affect both treatment efficacy and prognosis [57]. Moreover, mass media campaigns that target negative beliefs have been implemented in an effort to influence how people manage their back pain on a population level [58]. Our study has also provided novel findings with respect to emotions. Although we found a range of emotions, from positive emotions (such as happy, love, or fun) to negative emotions (including hell, bad, or disgusting), the majority were found to be of negative affect. Although several studies have examined the role of specific emotions, such as anger [59,60], in LBP research, our understanding of the array of emotions experienced by individuals with back pain, specifically negative emotions, is limited.

Our study also highlighted areas related to the pain experience in individuals that have not been adequately explored in the literature but that play an important role in the effectiveness of LBP interventions and self-management behaviors, such as the “not being understood,” “religion,” and “food and drink” categories. We found that although the category of “not being understood” had the smallest proportion of tweets with a total of 9180 tweets, it had the top five words: “make,” “people,” “stop,” “thing,” and “complain.” This is consistent with a previous systematic scoping review that examined what patients want from their medical care, which reported that patients felt misunderstood and wanted legitimization of their LBP [11]. Patients with LBP report negative social stereotyping from health care professionals, family and friends, and the community [61] and that they are dissatisfied with the inadequate advice they receive from medical practitioners and have identified an unmet need for care providers that show more understanding and empathy [11].

The category of “food and drink” is novel and interesting. The tweets included words relating to the type of food (eg, pizza, chocolate, cookies and cream), mealtimes (such as breakfast and lunch), and the process of bringing or making food. Although they reflect important daily habits of eating and drinking, they may also highlight issues around pain affecting an individual’s capacity to eat and drink and/or problems associated with weight and in particular obesity [62], which is a major public health issue [63].

There are well-described sex differences in the prevalence of back pain [64]. Analysis of tweets identified 3 topics under the “women” category including “motherhood,” “large breasts problem,” and “female health complaints.” LBP has been reported in more than two-thirds of pregnant women [65]. Improving psychological well-being, physical fitness, and general well-being may reduce LBP in women [65-67]. The topics identified in tweets may provide more direction in relation to the personal topics that warrant further examination (eg, the potential effect of “large breasts problem” and whether this is a cause of LBP or a potential confounding variable). Identifying possible mechanisms for the association with topics such as “motherhood” or “female health complaints” could also help with understanding whether these associations are due to psychosocial factors or biomechanical factors such as the lifting

and carrying of children. Understanding the context of LBP could offer valuable insights into how people with LBP view and experience their condition; this could lead to the identification of new areas of research in exploring the causes of LBP, as well as the opportunity to identify areas of potential misinformation that need to be addressed.

Limitations

There are some limitations to our study. Although the keywords were taken from existing studies about LBP and approved by domain experts, some keywords, such as “back hurt” and “back pain,” were very broad. Therefore, the data collected might not have been specific to LBP. Selection of the right keywords in Twitter data analysis is very important to avoid unrelated data that could reduce the accuracy of results. Filtering and cleaning of Twitter data is also crucial for achieving high accuracy of results. In our study, we performed vigorous data cleaning, but our manual examination showed that there was a group of tweets that contained a few lines from the lyrics of a famous hip-hop song (Bad and Boujee) by Migos. These lines included “...So my money makin' my back ache.” One of our search keywords was “back ache.” Although there are many tools and methods available to automatically perform data cleaning, it is always necessary to manually inspect the results.

Twitter users tend to be younger and might not represent the general population; therefore, the results must be carefully interpreted [68]. Similar to other social media studies in health care, we cannot verify that individuals who tweeted about LBP were actually real patients [15]. However, the filtering based on first-person pronouns (eg, I, my, or mine) that we performed is likely to have reduced this.

To determine the optimal number of topics, we used the coherence score, a widely used method, and then manually examined and compared the models. This process can be further improved by using other measures such as heuristic approaches [69] or perplexity measures [70].

We also recognize that manual labeling of topics can be subjective. Two domain experts with extensive knowledge were involved in the labeling and examination of selected topics but future work in this area could involve a greater number of and more diverse domain experts to further reduce this subjectivity.

Conclusions

Our findings provided useful insights into individuals’ beliefs and perspectives regarding their needs and concerns related to LBP that complement the information available in the literature. Considering the contextual factors identified in this study rather than simply focusing on a biomedical model of LBP could address the needs of patients more holistically, help with improving LBP outcomes, and increase patient satisfaction. These findings have the potential to assist health care providers and clinicians with developing more effective, personalized therapies for LBP. There is also the potential to use social media to identify any major changes in community beliefs and needs regarding LBP that can be addressed in a timelier manner.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. DU was supported by a National Health and Medical Research Council Career Development Fellowship (Level 2; 1142809).

Authors' Contributions

PDH, FB, DU, and FC contributed to study concept and design. R contributed to data collection and topic modeling. PDH, DU, and FC contributed to topic labeling and clustering. PDH, FB, DU, and FC contributed to interpretation of data. R and PDH contributed to drafting of the initial manuscript. PDH, FB, DU, and FC contributed to critical revision of the manuscript for important intellectual content. R and PDH provided administrative, technical, or material support. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The average number of words in tweets per year.

[\[DOCX File , 83 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Coherence score for latent Dirichlet allocation, Dirichlet multinomial mixture (DMM), General Pólya Urn Dirichlet Multinomial Mixture (GPU-DMM), biterm topic model, and nonnegative matrix factorization with number of topics 5-200 .

[\[DOCX File , 197 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

The best model selected with 60 topics and their top 20 terms.

[\[DOCX File , 35 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Word clouds for the pain region and sleep categories.

[\[DOCX File , 244 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

Total number of tweets per each topic manually labelled.

[\[DOCX File , 23 KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

The 19 contextual categories related to low back pain.

[\[DOCX File , 23 KB-Multimedia Appendix 6\]](#)

Multimedia Appendix 7

The total and percentage of tweets for each contextual category.

[\[DOCX File , 21 KB-Multimedia Appendix 7\]](#)

References

1. Hoy D, March L, Brooks P, Blyth F, Woolf A, Bain C, et al. The global burden of low back pain: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014 Jun 24;73(6):968-974. [doi: [10.1136/annrheumdis-2013-204428](https://doi.org/10.1136/annrheumdis-2013-204428)] [Medline: [24665116](https://pubmed.ncbi.nlm.nih.gov/24665116/)]
2. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: [10.1016/S0140-6736\(15\)60692-4](https://doi.org/10.1016/S0140-6736(15)60692-4)] [Medline: [26063472](https://pubmed.ncbi.nlm.nih.gov/26063472/)]
3. Rubin DI. Epidemiology and risk factors for spine pain. *Neurol Clin* 2007 May;25(2):353-371. [doi: [10.1016/j.ncl.2007.01.004](https://doi.org/10.1016/j.ncl.2007.01.004)] [Medline: [17445733](https://pubmed.ncbi.nlm.nih.gov/17445733/)]

4. Hoy D, Brooks P, Blyth F, Buchbinder R. The Epidemiology of low back pain. *Best Pract Res Clin Rheumatol* 2010 Dec;24(6):769-781. [doi: [10.1016/j.berh.2010.10.002](https://doi.org/10.1016/j.berh.2010.10.002)] [Medline: [21665125](https://pubmed.ncbi.nlm.nih.gov/21665125/)]
5. Guo HR, Tanaka S, Halperin WE, Cameron LL. Back pain prevalence in US industry and estimates of lost workdays. *Am J Public Health* 1999 Jul;89(7):1029-1035. [doi: [10.2105/ajph.89.7.1029](https://doi.org/10.2105/ajph.89.7.1029)] [Medline: [10394311](https://pubmed.ncbi.nlm.nih.gov/10394311/)]
6. Katz JN. Lumbar disc disorders and low-back pain: socioeconomic factors and consequences. *J Bone Joint Surg Am* 2006 Apr;88 Suppl 2:21-24. [doi: [10.2106/JBJS.E.01273](https://doi.org/10.2106/JBJS.E.01273)] [Medline: [16595438](https://pubmed.ncbi.nlm.nih.gov/16595438/)]
7. Mafi JN, McCarthy EP, Davis RB, Landon BE. Worsening trends in the management and treatment of back pain. *JAMA Intern Med* 2013 Sep 23;173(17):1573-1581 [FREE Full text] [doi: [10.1001/jamainternmed.2013.8992](https://doi.org/10.1001/jamainternmed.2013.8992)] [Medline: [23896698](https://pubmed.ncbi.nlm.nih.gov/23896698/)]
8. Chou L, Cicuttini FM, Urquhart DM, Anthony SN, Sullivan K, Seneviwickrama M, et al. People with low back pain perceive needs for non-biomedical services in workplace, financial, social and household domains: a systematic review. *J Physiother* 2018 Apr;64(2):74-83 [FREE Full text] [doi: [10.1016/j.jphys.2018.02.011](https://doi.org/10.1016/j.jphys.2018.02.011)] [Medline: [29574167](https://pubmed.ncbi.nlm.nih.gov/29574167/)]
9. Lim YZ, Chou L, Au RT, Seneviwickrama KMD, Cicuttini FM, Briggs AM, et al. People with low back pain want clear, consistent and personalised information on prognosis, treatment options and self-management strategies: a systematic review. *J Physiother* 2019 Jul;65(3):124-135 [FREE Full text] [doi: [10.1016/j.jphys.2019.05.010](https://doi.org/10.1016/j.jphys.2019.05.010)] [Medline: [31227280](https://pubmed.ncbi.nlm.nih.gov/31227280/)]
10. De Souza LH, Frank AO. Subjective pain experience of people with chronic back pain. *Physiother Res Int* 2000 Nov;5(4):207-219. [doi: [10.1002/pri.201](https://doi.org/10.1002/pri.201)] [Medline: [11129663](https://pubmed.ncbi.nlm.nih.gov/11129663/)]
11. Chou L, Ranger TA, Peiris W, Cicuttini FM, Urquhart DM, Sullivan K, et al. Patients' perceived needs of health care providers for low back pain management: a systematic scoping review. *Spine J* 2018 Apr;18(4):691-711. [doi: [10.1016/j.spinee.2018.01.006](https://doi.org/10.1016/j.spinee.2018.01.006)] [Medline: [29373836](https://pubmed.ncbi.nlm.nih.gov/29373836/)]
12. Symonds TL, Burton AK, Tillotson KM, Main CJ. Absence resulting from low back trouble can be reduced by psychosocial intervention at the work place. *Spine (Phila Pa 1976)* 1995 Dec 15;20(24):2738-2745. [doi: [10.1097/00007632-199512150-00016](https://doi.org/10.1097/00007632-199512150-00016)] [Medline: [8747253](https://pubmed.ncbi.nlm.nih.gov/8747253/)]
13. Waddell G, Newton M, Henderson I, Somerville D, Main C. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain* 1993 Feb;52(2):157-168. [doi: [10.1016/0304-3959\(93\)90127-B](https://doi.org/10.1016/0304-3959(93)90127-B)] [Medline: [8455963](https://pubmed.ncbi.nlm.nih.gov/8455963/)]
14. Delir Haghighi P, Kang Y, Buchbinder R, Burstein F, Whittle S. Investigating Subjective Experience and the Influence of Weather Among Individuals With Fibromyalgia: A Content Analysis of Twitter. *JMIR Public Health Surveill* 2017 Jan 19;3(1):e4 [FREE Full text] [doi: [10.2196/publichealth.6344](https://doi.org/10.2196/publichealth.6344)] [Medline: [28104577](https://pubmed.ncbi.nlm.nih.gov/28104577/)]
15. Lee H, McAuley JH, Hübscher M, Allen HG, Kamper SJ, Moseley GL. Tweeting back: predicting new cases of back pain with mass social media data. *J Am Med Inform Assoc* 2016 May;23(3):644-648. [doi: [10.1093/jamia/ocv168](https://doi.org/10.1093/jamia/ocv168)] [Medline: [26661720](https://pubmed.ncbi.nlm.nih.gov/26661720/)]
16. Asghar MZ, Ahmad S, Qasim M, Zahra SR, Kundi FM. SentiHealth: creating health-related sentiment lexicon using hybrid approach. *Springerplus* 2016;5(1):1139 [FREE Full text] [doi: [10.1186/s40064-016-2809-x](https://doi.org/10.1186/s40064-016-2809-x)] [Medline: [27504237](https://pubmed.ncbi.nlm.nih.gov/27504237/)]
17. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:3 [FREE Full text] [doi: [10.1186/2047-2501-2-3](https://doi.org/10.1186/2047-2501-2-3)] [Medline: [25825667](https://pubmed.ncbi.nlm.nih.gov/25825667/)]
18. Bian J, Topaloglu U, Yu F. Towards Large-scale Twitter Mining for Drug-related Adverse Events. *SHB12 (2012) 2012* Oct 29;2012:25-32 [FREE Full text] [doi: [10.1145/2389707.2389713](https://doi.org/10.1145/2389707.2389713)] [Medline: [28967001](https://pubmed.ncbi.nlm.nih.gov/28967001/)]
19. Pershad Y, Hangege P, Albadawi H, Oklu R. Social Medicine: Twitter in Healthcare. *J Clin Med* 2018 May 28;7(6):121 [FREE Full text] [doi: [10.3390/jcm7060121](https://doi.org/10.3390/jcm7060121)] [Medline: [29843360](https://pubmed.ncbi.nlm.nih.gov/29843360/)]
20. Aichner T, Jacob F. Measuring the Degree of Corporate Social Media Use. *International Journal of Market Research* 2015 Mar 01;57(2):257-276. [doi: [10.2501/ijmr-2015-018](https://doi.org/10.2501/ijmr-2015-018)]
21. Sinnenberg L, Bittenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. *Am J Public Health* 2017 Jan;107(1):e1-e8. [doi: [10.2105/AJPH.2016.303512](https://doi.org/10.2105/AJPH.2016.303512)] [Medline: [27854532](https://pubmed.ncbi.nlm.nih.gov/27854532/)]
22. Jayaraman PP, Forkan ARM, Morshed A, Haghighi PD, Kang Y. Healthcare 4.0: A review of frontiers in digital health. *WIREs Data Mining Knowl Discov* 2019 Dec 25;10(2):1-23. [doi: [10.1002/widm.1350](https://doi.org/10.1002/widm.1350)]
23. Goh T, Delir Haghighi P, Burstein F, Buchbinder R. Developing a Contextual Model towards Understanding Low Back Pain. In: *Proceedings of the 19th Pacific Asia Conference on Information Systems*. 2015 Presented at: 19th Pacific Asia Conference on Information Systems; July 5, 2015; Singapore URL: <http://aisel.aisnet.org/pacis2015/64>
24. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005 Nov;15(9):1277-1288. [doi: [10.1177/1049732305276687](https://doi.org/10.1177/1049732305276687)] [Medline: [16204405](https://pubmed.ncbi.nlm.nih.gov/16204405/)]
25. Hewis J. Do MRI Patients Tweet? Thematic Analysis of Patient Tweets About Their MRI Experience. *J Med Imaging Radiat Sci* 2015 Dec;46(4):396-402. [doi: [10.1016/j.jmir.2015.08.003](https://doi.org/10.1016/j.jmir.2015.08.003)] [Medline: [31052120](https://pubmed.ncbi.nlm.nih.gov/31052120/)]
26. Prier K, Smith M, Giraud-Carrier C, Hanson C. Identifying Health-Related Topics on Twitter: An Exploration of Tobacco-Related Tweets as a Test Topic. In: Salerno J, Yang SJ, Nau D, Chai SK, editors. *Social Computing, Behavioral-Cultural Modeling and Prediction*. Berlin, Heidelberg: Springer; Mar 19, 2011:18-25.
27. TWINT - Twitter Intelligence Tool. GitHub. URL: <https://github.com/twintproject/twint> [accessed 2021-03-29]
28. Xavier C, Souza M. A Basic Approach for Extracting and Analyzing Data from Twitter. In: Roesler V, Barré E, Willrich R, editors. *Special Topics in Multimedia, IoT and Web Technologies*. Cham: Springer; Mar 3, 2020:185-211.

29. Campbell P, Wynne-Jones G, Muller S, Dunn KM. The influence of employment social support for risk and prognosis in nonspecific back pain: a systematic review and critical synthesis. *Int Arch Occup Environ Health* 2013 Feb 9;86(2):119-137 [FREE Full text] [doi: [10.1007/s00420-012-0804-2](https://doi.org/10.1007/s00420-012-0804-2)] [Medline: [22875173](https://pubmed.ncbi.nlm.nih.gov/22875173/)]
30. Ahlwardt K, Heavilin N, Gibbs J, Page J, Gerbert B, Tsoh JY. Tweeting about pain: comparing self-reported toothache experiences with those of backaches, earaches and headaches. *J Am Dent Assoc* 2014 Jul;145(7):737-743 [FREE Full text] [doi: [10.14219/jada.2014.30](https://doi.org/10.14219/jada.2014.30)] [Medline: [24982280](https://pubmed.ncbi.nlm.nih.gov/24982280/)]
31. Jian T, Zhaoshi M, Xuanlong N, Qiaozhu M, Ming Z. Understanding the limiting factors of topic modeling via posterior contraction analysis. 2014 Jun 21 Presented at: 31st International Conference on International Conference on Machine Learning; 2014; Beijing, China p. 190-198.
32. Blei D, Carin L, Dunson D. Probabilistic Topic Models: A focus on graphical model design and applications to document and image analysis. *IEEE Signal Process Mag* 2010 Nov 01;27(6):55-65 [FREE Full text] [doi: [10.1109/MSP.2010.938079](https://doi.org/10.1109/MSP.2010.938079)] [Medline: [25104898](https://pubmed.ncbi.nlm.nih.gov/25104898/)]
33. Blei D, Ng A, Jordan M. Latent dirichllocation. *Journal of Machine Learning Research* 2003;3:993-1022 [FREE Full text]
34. Nigam K, McCallum AK, Thrun S, Mitchell T. Text Classification from Labeled and Unlabeled Documents using EM. *Machine Learning* 2000;39:103-134. [doi: [10.21236/ada350490](https://doi.org/10.21236/ada350490)]
35. Li C, Wang H, Zhang Z, Sun A, Ma Z. Topic Modeling for Short Texts with Auxiliary Word Embeddings. In: SIGIR '16: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. New York, NY: ACM Press; 2016 Jul 17 Presented at: SIGIR '16; July 17 - 21; Pisa, Italy p. 165-174. [doi: [10.1145/2911451.2911499](https://doi.org/10.1145/2911451.2911499)]
36. Cheng X, Yan X, Lan Y, Guo J. BTM: Topic Modeling over Short Texts. *IEEE Trans Knowl Data Eng* 2014 Dec 1;26(12):2928-2941. [doi: [10.1109/tkde.2014.2313872](https://doi.org/10.1109/tkde.2014.2313872)]
37. Lee D, Seung S. Algorithms for Non-negative Matrix Factorization. : MIT Press; 2000 Jan 1 Presented at: 13th International Conference on Neural Information Processing; 2000; Denver, CO p. 535-541.
38. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study. *J Med Internet Res* 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: [10.2196/22624](https://doi.org/10.2196/22624)] [Medline: [33006937](https://pubmed.ncbi.nlm.nih.gov/33006937/)]
39. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *J Med Internet Res* 2016;18(8):e232 [FREE Full text] [doi: [10.2196/jmir.6045](https://doi.org/10.2196/jmir.6045)] [Medline: [27573910](https://pubmed.ncbi.nlm.nih.gov/27573910/)]
40. Liang W, Feng R, Liu X, Li Y, Zhang X. GLTM: A Global and Local Word Embedding-Based Topic Model for Short Texts. *IEEE Access* 2018;6:43612-43621. [doi: [10.1109/access.2018.2863260](https://doi.org/10.1109/access.2018.2863260)]
41. Mackey T, Kalyanam J, Klugman J, Kuzmenko E, Gupta R. Solution to Detect, Classify, and Report Illicit Online Marketing and Sales of Controlled Substances via Twitter: Using Machine Learning and Web Forensics to Combat Digital Opioid Access. *J Med Internet Res* 2018 Dec 27;20(4):e10029 [FREE Full text] [doi: [10.2196/10029](https://doi.org/10.2196/10029)] [Medline: [29613851](https://pubmed.ncbi.nlm.nih.gov/29613851/)]
42. Odlum M, Yoon S, Broadwell P, Brewer R, Kuang D. How Twitter Can Support the HIV/AIDS Response to Achieve the 2030 Eradication Goal: In-Depth Thematic Analysis of World AIDS Day Tweets. *JMIR Public Health Surveill* 2018 Nov 22;4(4):e10262 [FREE Full text] [doi: [10.2196/10262](https://doi.org/10.2196/10262)] [Medline: [30467102](https://pubmed.ncbi.nlm.nih.gov/30467102/)]
43. Wang Y, Zhang Y. Nonnegative Matrix Factorization: A Comprehensive Review. *IEEE Trans Knowl Data Eng* 2013 Jun;25(6):1336-1353. [doi: [10.1109/tkde.2012.51](https://doi.org/10.1109/tkde.2012.51)]
44. STTM: A Library of Short Text Topic Modeling. GitHub. URL: <https://github.com/qiang2100/STTM> [accessed 2021-02-17]
45. scikit-learn: Machine Learning in Python. URL: <https://scikit-learn.org/stable/> [accessed 2021-02-17]
46. Chang J, Gerrish S, Wang C, Boyd-graber J, Blei D. Reading tea leaves: how humans interpret topic models. 2009 Presented at: 22nd International Conference on Neural Information Processing Systems; December 7, 2009; Vancouver, British Columbia, Canada p. 288-296 URL: <https://papers.nips.cc/paper/2009/hash/f92586a25bb3145facd64ab20fd554ff-Abstract.html>
47. Church K, Hanks P. Word association norms, mutual information, and lexicography. *Computational Linguistics* 1990 Mar 1;16(1):76-83. [doi: [10.3115/981623.981633](https://doi.org/10.3115/981623.981633)]
48. Newman D, Bonilla E, Buntine W. Improving topic coherence with regularized topic models. 2011 Presented at: 24th International Conference on Neural Information Processing Systems; December 12, 2011; Granada, Spain p. 496-504 URL: <https://papers.nips.cc/paper/2011/hash/5ef698cd9fe650923ea331c15af3b160-Abstract.html>
49. Allahyari M, Pouriyeh S, Kochut K, Reza H. A Knowledge-based Topic Modeling Approach for Automatic Topic Labeling. *ijacsa* 2017;8(9):335-349. [doi: [10.14569/ijacsa.2017.080947](https://doi.org/10.14569/ijacsa.2017.080947)]
50. Morstatter F, Liu H. In Search of Coherence and Consensus: Measuring the Interpretability of Statistical Topics. *Journal of Machine Learning Research* 2017;18(1):6177-6208.
51. Grimmer J, Stewart BM. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 2017 Jan 04;21(3):267-297. [doi: [10.1093/pan/mps028](https://doi.org/10.1093/pan/mps028)]
52. Nowell LS, Norris JM, White DE, Moules NJ. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods* 2017 Oct 02;16(1):160940691773384. [doi: [10.1177/1609406917733847](https://doi.org/10.1177/1609406917733847)]

53. Boot AB, Tjong Kim Sang E, Dijkstra K, Zwaan RA. How character limit affects language usage in tweets. *Palgrave Commun* 2019 Jul 9;5(1):1-13. [doi: [10.1057/s41599-019-0280-3](https://doi.org/10.1057/s41599-019-0280-3)]
54. Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine (Phila Pa 1976)* 2002 Mar 01;27(5):E109-E120. [doi: [10.1097/00007632-200203010-00017](https://doi.org/10.1097/00007632-200203010-00017)] [Medline: [11880847](https://pubmed.ncbi.nlm.nih.gov/11880847/)]
55. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. *Lancet* 2017 Feb 18;389(10070):736-747. [doi: [10.1016/S0140-6736\(16\)30970-9](https://doi.org/10.1016/S0140-6736(16)30970-9)] [Medline: [27745712](https://pubmed.ncbi.nlm.nih.gov/27745712/)]
56. Morton L, de Bruin M, Krajewska M, Whibley D, Macfarlane G. Beliefs about back pain and pain management behaviours, and their associations in the general population: A systematic review. *Eur J Pain* 2019 Jan 07;23(1):15-30 [FREE Full text] [doi: [10.1002/ejp.1285](https://doi.org/10.1002/ejp.1285)] [Medline: [29984553](https://pubmed.ncbi.nlm.nih.gov/29984553/)]
57. Wertli MM, Rasmussen-Barr E, Held U, Weiser S, Bachmann LM, Brunner F. Fear-avoidance beliefs-a moderator of treatment efficacy in patients with low back pain: a systematic review. *Spine J* 2014 Nov 01;14(11):2658-2678. [doi: [10.1016/j.spinee.2014.02.033](https://doi.org/10.1016/j.spinee.2014.02.033)] [Medline: [24614254](https://pubmed.ncbi.nlm.nih.gov/24614254/)]
58. Urquhart DM, Bell RJ, Cicuttini FM, Cui J, Forbes A, Davis SR. Negative beliefs about low back pain are associated with high pain intensity and high level disability in community-based women. *BMC Musculoskelet Disord* 2008 Nov 04;9(1):148-148 [FREE Full text] [doi: [10.1186/1471-2474-9-148](https://doi.org/10.1186/1471-2474-9-148)] [Medline: [18983681](https://pubmed.ncbi.nlm.nih.gov/18983681/)]
59. Bruehl S, Liu X, Burns J, Chont M, Jamison R. Associations between daily chronic pain intensity, daily anger expression, and trait anger expressiveness: an ecological momentary assessment study. *Pain* 2012 Dec;153(12):2352-2358 [FREE Full text] [doi: [10.1016/j.pain.2012.08.001](https://doi.org/10.1016/j.pain.2012.08.001)] [Medline: [22940462](https://pubmed.ncbi.nlm.nih.gov/22940462/)]
60. Burns JW, Quartana P, Bruehl S. Anger suppression and subsequent pain behaviors among chronic low back pain patients: moderating effects of anger regulation style. *Ann Behav Med* 2011 Aug 5;42(1):42-54 [FREE Full text] [doi: [10.1007/s12160-011-9270-4](https://doi.org/10.1007/s12160-011-9270-4)] [Medline: [21544702](https://pubmed.ncbi.nlm.nih.gov/21544702/)]
61. Slade SC, Molloy E, Keating JL. Stigma experienced by people with nonspecific chronic low back pain: a qualitative study. *Pain Med* 2009 Jan 01;10(1):143-154. [doi: [10.1111/j.1526-4637.2008.00540.x](https://doi.org/10.1111/j.1526-4637.2008.00540.x)] [Medline: [19222775](https://pubmed.ncbi.nlm.nih.gov/19222775/)]
62. Chou L, Brady S, Urquhart D, Teichtahl AJ, Cicuttini FM, Pasco JA, et al. The Association Between Obesity and Low Back Pain and Disability Is Affected by Mood Disorders: A Population-Based, Cross-Sectional Study of Men. *Medicine (Baltimore)* 2016 Apr;95(15):e3367 [FREE Full text] [doi: [10.1097/MD.0000000000003367](https://doi.org/10.1097/MD.0000000000003367)] [Medline: [27082599](https://pubmed.ncbi.nlm.nih.gov/27082599/)]
63. Agha M, Agha R. The rising prevalence of obesity: part A: impact on public health. *Int J Surg Oncol (N Y)* 2017 Aug;2(7):e17 [FREE Full text] [doi: [10.1097/IJ9.0000000000000017](https://doi.org/10.1097/IJ9.0000000000000017)] [Medline: [29177227](https://pubmed.ncbi.nlm.nih.gov/29177227/)]
64. Wu A, March L, Zheng X, Huang J, Wang X, Zhao J, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. *Ann Transl Med* 2020 Mar;8(6):299-299 [FREE Full text] [doi: [10.21037/atm.2020.02.175](https://doi.org/10.21037/atm.2020.02.175)] [Medline: [32355743](https://pubmed.ncbi.nlm.nih.gov/32355743/)]
65. Liddle S, Pennick V. Interventions for preventing and treating low-back and pelvic pain during pregnancy. *Cochrane Database Syst Rev* 2015 Sep 30(9):CD001139 [FREE Full text] [doi: [10.1002/14651858.CD001139.pub4](https://doi.org/10.1002/14651858.CD001139.pub4)] [Medline: [26422811](https://pubmed.ncbi.nlm.nih.gov/26422811/)]
66. Brady SR, Hussain SM, Brown WJ, Heritier S, Wang Y, Teede H, et al. Course and Contributors to Back Pain in Middle-aged Women Over 9 Years: Data From the Australian Longitudinal Study on Women's Health. *Spine (Phila Pa 1976)* 2018 Dec 01;43(23):1648-1656. [doi: [10.1097/BRS.0000000000002702](https://doi.org/10.1097/BRS.0000000000002702)] [Medline: [29794589](https://pubmed.ncbi.nlm.nih.gov/29794589/)]
67. Ng SK, Cicuttini FM, Davis SR, Bell R, Botlero R, Fitzgibbon BM, et al. Poor general health and lower levels of vitality are associated with persistent, high-intensity low back pain and disability in community-based women: A prospective cohort study. *Maturitas* 2018 Jul;113:7-12. [doi: [10.1016/j.maturitas.2018.04.007](https://doi.org/10.1016/j.maturitas.2018.04.007)] [Medline: [29903650](https://pubmed.ncbi.nlm.nih.gov/29903650/)]
68. Zhang H, Wheldon C, Dunn A, Tao C, Huo J, Zhang R, et al. Mining Twitter to assess the determinants of health behavior toward GPU-DMM: General Pólya Urn Dirichlet Multinomial Mixturehuman papillomavirus vaccination in the United States. *J Am Med Inform Assoc* 2020 Feb 01;27(2):225-235 [FREE Full text] [doi: [10.1093/jamia/ocz191](https://doi.org/10.1093/jamia/ocz191)] [Medline: [31711186](https://pubmed.ncbi.nlm.nih.gov/31711186/)]
69. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics* 2015;16 Suppl 13:S8 [FREE Full text] [doi: [10.1186/1471-2105-16-S13-S8](https://doi.org/10.1186/1471-2105-16-S13-S8)] [Medline: [26424364](https://pubmed.ncbi.nlm.nih.gov/26424364/)]
70. Wallach H, Murray I, Salakhutdinov R, Mimno D. Evaluation methods for topic models. In: *Proceedings of the 26th Annual International Conference on Machine Learning.*: ACM Press; 2009 Presented at: 26th Annual International Conference on Machine Learning; June 14-18, 2009; Montreal, Quebec, Canada p. 1-8.

Abbreviations

- API:** application programming interface
- BTM:** biterm topic model
- DMM:** Dirichlet multinomial mixture
- GPU-DMM:** General Pólya Urn Dirichlet Multinomial Mixture
- LBP:** low back pain

LDA: latent Dirichlet allocation
NMF: nonnegative matrix factorization
PMI: pointwise mutual information
STTM: short text topic modeling algorithm

Edited by JMIRPE Office; submitted 27.11.20; peer-reviewed by G Chirambo, E O'Hagan; comments to author 22.01.21; revised version received 06.03.21; accepted 21.11.21; published 23.12.21

Please cite as:

Robert, Delir Haghighi P, Burstein F, Urquhart D, Cicuttini F

Investigating Individuals' Perceptions Regarding the Context Around the Low Back Pain Experience: Topic Modeling Analysis of Twitter Data

J Med Internet Res 2021;23(12):e26093

URL: <https://www.jmir.org/2021/12/e26093>

doi: [10.2196/26093](https://doi.org/10.2196/26093)

PMID:

© Robert, Pari Delir Haghighi, Frada Burstein, Donna Urquhart, Flavia Cicuttini. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 23.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.