

Review

Consumer-Generated Discourse on Cannabis as a Medicine: Scoping Review of Techniques

Sedigheh Khademi Habibabadi^{1*}, PhD; Christine Hallinan^{1,2*}, PhD; Yvonne Bonomo¹, MD, PhD; Mike Conway^{3*}, PhD

¹Department of General Practice, Melbourne Medical School, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Australia

²Health & Biomedical Research Information Technology Unit, The University of Melbourne, Melbourne, Australia

³School of Computing & Information Systems, The University of Melbourne, Melbourne, Australia

* these authors contributed equally

Corresponding Author:

Sedigheh Khademi Habibabadi, PhD

Department of General Practice

Melbourne Medical School

Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne

Grattan Street, Parkville, Victoria

Melbourne, 3010

Australia

Phone: 61 405761879

Email: sedigh.khademi@unimelb.edu.au

Abstract

Background: Medicinal cannabis is increasingly being used for a variety of physical and mental health conditions. Social media and web-based health platforms provide valuable, real-time, and cost-effective surveillance resources for gleaning insights regarding individuals who use cannabis for medicinal purposes. This is particularly important considering that the evidence for the optimal use of medicinal cannabis is still emerging. Despite the web-based marketing of medicinal cannabis to consumers, currently, there is no robust regulatory framework to measure clinical health benefits or individual experiences of adverse events. In a previous study, we conducted a systematic scoping review of studies that contained themes of the medicinal use of cannabis and used data from social media and search engine results. This study analyzed the methodological approaches and limitations of these studies.

Objective: We aimed to examine research approaches and study methodologies that use web-based user-generated text to study the use of cannabis as a medicine.

Methods: We searched MEDLINE, Scopus, Web of Science, and Embase databases for primary studies in the English language from January 1974 to April 2022. Studies were included if they aimed to understand web-based user-generated text related to health conditions where cannabis is used as a medicine or where health was mentioned in general cannabis-related conversations.

Results: We included 42 articles in this review. In these articles, Twitter was used 3 times more than other computer-generated sources, including Reddit, web-based forums, GoFundMe, YouTube, and Google Trends. Analytical methods included sentiment assessment, thematic analysis (manual and automatic), social network analysis, and geographic analysis.

Conclusions: This study is the first to review techniques used by research on consumer-generated text for understanding cannabis as a medicine. It is increasingly evident that consumer-generated data offer opportunities for a greater understanding of individual behavior and population health outcomes. However, research using these data has some limitations that include difficulties in establishing sample representativeness and a lack of methodological best practices. To address these limitations, deidentified annotated data sources should be made publicly available, researchers should determine the origins of posts (organizations, bots, power users, or ordinary individuals), and powerful analytical techniques should be used.

(*J Med Internet Res* 2022;24(11):e35974) doi: [10.2196/35974](https://doi.org/10.2196/35974)

KEYWORDS

social media; data mining; internet and the web technology; consumer-generated data; medicinal cannabis; medical marijuana

Introduction

Medicinal Cannabis Pharmacovigilance

Cannabis has been widely used for a variety of purposes, including medicinal applications, throughout human history. Over the last century, its use has been prohibited in Europe, Northern America, and Australasia [1]. Since 2016, these jurisdictions have incrementally authorized the use of medicinal cannabis for certain conditions [2]. Given the substantial public interest in cannabis as medicine, there is a pressing need to better understand its safety and efficacy.

However, aside from clinical trials, there are scant data regarding the efficacy and side effects of medicinal cannabis [3-6]. One of the main methods for postmarketing safety surveillance of medications is the use of established pharmacovigilance reporting systems, which rely on reporting of adverse events by individuals [7-9]. Cannabis users are often unaware of these systems or the importance of reporting. They may find them too difficult to use or may not want to divulge personal details if these are required [10]. Users may not even think of reporting their side effects because they consider them an inherent experience of cannabis consumption, especially if they are not using an approved medical cannabis product.

Increasing the understanding of the efficacy and safety of cannabis as medicine is warranted because cannabis is a nonstandardized product, given the wide variety in growing conditions and production specifications [11]. This includes variations in climate, soil (or other growth media), water, light, and other factors that affect plant growth. Even if cannabis medicines in a country or state must adhere to mandatory standards (good manufacturing practice), some cannabis users prefer to grow or import their own cannabis [12]. These factors make the systematic assessment of the effectiveness of medical cannabis and its side effects difficult.

Social Media as a Pharmacovigilance Data Source

To gain additional insights into cannabis use and its effects, researchers are now turning to social media and web-based health forums. These platforms are a place for both patients and the general population to freely express and exchange their experiences and thus provide a valuable additional data source for monitoring public health [13]. Unlike other forms of highly curated data collection methods, such as surveys or interviews, social media provides an organic view of everyday thoughts, behaviors, and activities of people. Therefore, social media has the potential to provide insights beyond the boundaries of

targeted investigations, including emergent events, observations of behavioral phenomena and subcultures, and insights for the social sciences [14].

The information contained in social media conversations is voluminous and not only potentially rich in content but also complex and varied. As an unstructured raw data source, credible information may be sparse and difficult to identify; there may be uncertainty about the origin of the data or the population they represent [15]. Furthermore, it is difficult to interpret the informal language and structure of social media posts, which are confounded by many competing sources, such as promotional posts, hashtags, and social media bots [16,17]. Social media bots automatically create content and interact with social media platform users [18]. A study found that between 9% and 15% of Twitter accounts are bots [19]. Notwithstanding these limitations, if these complexities can be successfully navigated, social media has the potential to be a great asset for increased understanding of cannabis as a medicine.

Our previous systematic scoping review [20] used PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [21] to understand the utility of web-based user-generated text in providing insight into the use of cannabis as a medicine. This paper examines the techniques, analyses, and limitations of these studies.

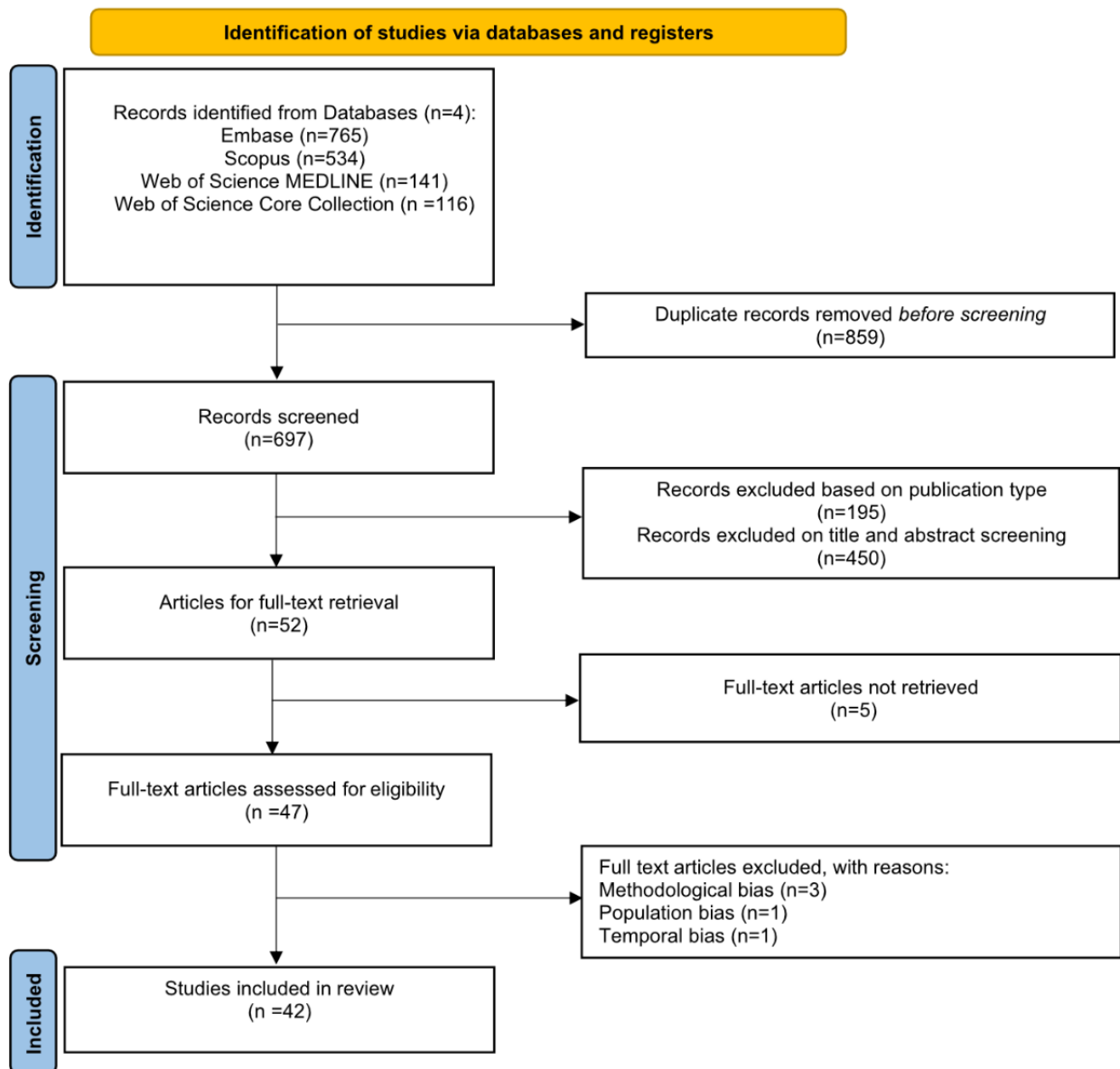
The objective of this research was to provide a review of studies that have used user-generated data in conjunction with computational methods to understand the medicinal use of cannabis in a population. We addressed the following research questions (RQs):

- RQ1: What consumer-generated data sources are used for studying cannabis?
- RQ2: What common techniques for collection and analysis of data are used?
- RQ3: What are the common limitations and challenges faced by the studies?

Methods

We searched for English-language studies that were indexed in MEDLINE, Embase, Web of Science, and Scopus databases and published between January 2010 and March 2022. Literature database queries were developed for these 4 databases. See Table S1 in [Multimedia Appendix 1](#) [22-63] for the details of search terms used and [Multimedia Appendix 1](#) Table S2 for the inclusion and exclusion criteria of the selected articles. A summary of the PRISMA flowchart is shown in [Figure 1](#) [20].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the study selection process [20].



Results

Overview

Table 1 provides a summary of each article that includes author names, publication year, data source, and duration of data collection, analysis, and number of items analyzed.

The year with the highest number of publications was 2020 (11/42, 26%), followed by 2017 and 2021 (6/40, 14%). Of the

42 studies, 6 (12%) were conducted in 2015 and 2019. The number of publications per year is shown in Table 2.

Regarding data sources, Twitter was used in 40% (17/42) of the reviewed studies, around 3 times the number of studies using either Reddit or web-based forums 14% (6/42). GoFundMe, YouTube, and Google Trends comprised 7% (3/42) of the total. Text was the focus of 83% (35/42) of the studies, whereas the others analyzed trends, videos, search logs, and images. Table 3 shows the distribution of the publications selected per data source.

Table 1. Articles included in the review.

Study	Source (duration)	Analysis	Number of items analyzed
McGregor et al [22], 2014	Web-based forums, Facebook, Twitter, and YouTube (not available)	<ul style="list-style-type: none"> Thematic and content analysis of glaucoma-related posts on the following: <ul style="list-style-type: none"> Analysis of the nature of the post (personal stories, information sharing or flagging, supportive comments, questions, answers, and general discussions) Sentiment analysis (positive or negative) 	3785 items
Cavazos-Rehg et al [23], 2015	Twitter (February to March 2014)	<ul style="list-style-type: none"> Cannabis-related chatter by influential users on the following: <ul style="list-style-type: none"> Sentiment analysis by using the Likert scale Thematic analysis of tweets Demographic analysis 	7000 tweets
Daniulaityte et al [24], 2015	Twitter (October to December 2014)	<ul style="list-style-type: none"> US dab-related tweets: <ul style="list-style-type: none"> Counting and normalizing based on cannabis legalization policy 	125,255 tweets (27,018 geolocated tweets)
Gonzalez-Estrada et al [25], 2015	YouTube (June 4-8, 2014)	<ul style="list-style-type: none"> Content analysis of asthma-related videos on the following: <ul style="list-style-type: none"> Source: professional society, media, asthma care provider, etc Content: personal experience, medical professional, advertisement, patient education, alternative treatment, or to increase awareness Quality scoring of misleading and useful info Video characteristics or video statistics 	200 most viewed videos
Krauss et al [26], 2015	YouTube (January 22, 2015)	<ul style="list-style-type: none"> Analysis of dabbing-related videos on the following: <ul style="list-style-type: none"> Characteristics of the people dabbing (age and skills) Characteristics of the session Messages included in the videos 	116 videos
Thompson et al [27], 2015	Twitter (March 2012 to July 2013)	<ul style="list-style-type: none"> Content analysis of cannabis-related tweets and retweets on the following: <ul style="list-style-type: none"> Adolescence users (age, inferred from the user profile) Sentiment (positive, negative, or unclear) Subject (self, other, general, or subject unclear) Use category (own use, use by others, or not mentioned) Related behaviors (habitual use, social aspect, etc) Positive aspects (better than other drugs and medical use) 	36,939 original tweets and 10,000 retweets
Cavazos-Rehg et al [28], 2016	Twitter (January 2015)	<ul style="list-style-type: none"> Dabbing-related tweets: <ul style="list-style-type: none"> Thematic analysis of tweets to 7 themes Subanalysis of 1 theme (extreme effects) into physiological or psychological effects Geotagged tweets analysis for number per state Demographic analysis 	5000 tweets
Lamy et al [29], 2016	Twitter (May to July 2015)	<ul style="list-style-type: none"> Content analysis of cannabis edible-related conversations: <ul style="list-style-type: none"> Tweet sources (media, retail, or users) Sentiment analysis (positive, negative, or neutral) Word frequency analysis Geotagging (policy impact on the volume of tweets) 	3000 tweets
Mitchell et al [30], 2016	Web-based forums (October 2014)	<ul style="list-style-type: none"> Thematic analysis of ADHD^a and cannabis web-based forum posts on the following: <ul style="list-style-type: none"> Impact of cannabis on ADHD symptoms (therapeutic, harmful, both, and none) Other domains (mood, psychiatric conditions, and other [sleep]) Comments about cannabis as medicinal (more effective than other ADHD medications, less effective, or not legal) 	268 threads

Study	Source (duration)	Analysis	Number of items analyzed
Andersson et al [31], 2017	Web-based forums (April 18-19, 2016)	<ul style="list-style-type: none"> Thematic analysis of conversations on headache-related posts 	32 topics
Dai and Hao [32], 2017	Twitter (August 2015 to April 2016)	<ul style="list-style-type: none"> Naive Bayes classifier on PTSD^b and cannabis-related tweets: <ul style="list-style-type: none"> Sentiment analysis Analysis of prevalence of support of cannabis use for PTSD in association with state level legislation and socioeconomic factors 	66,000 cannabis-related and 31,184 geolocated tweets
Greiner et al [33], 2017	Web-based forums (November 2014 to March 2015)	<ul style="list-style-type: none"> Content analysis of cannabis help forums on the following: <ul style="list-style-type: none"> Fields of interest (illness-related, social, financial, and legal issues) Self-help mechanisms (exchange of information, emotional support, group support) Analysis of sex and age when available Highly involved vs moderately involved users 	717 posts
Turner and Kantardzic [34], 2017	Twitter (August 2015 to April 2016)	<ul style="list-style-type: none"> Supervised and unsupervised machine learning techniques of cannabis-related tweets: <ul style="list-style-type: none"> Binary classification to identify marijuana-related tweet Topic modeling User social network analysis Spatiotemporal analysis of conversations 	40,509 geolocated tweets
Westmaas et al [35], 2017	Web-based forums (January 2000 to December 2013)	<ul style="list-style-type: none"> Topic modeling of Cancer Survivors Network: <ul style="list-style-type: none"> Analyze smoking or cessation-related content Analysis to determine the overall context in which these discussions occurred 	468,000 posts
Yom Tov and Lev Ran [36], 2017	Bing logs (November 2016 to April 2017)	<ul style="list-style-type: none"> Statistical analysis of cannabis-related query logs 	Not available
Cavazos-Rehg et al [37], 2018	YouTube (June 10-11, 2015)	<ul style="list-style-type: none"> Cannabis review web-based videos: <ul style="list-style-type: none"> Sentiment analysis Physical or mental effects; is it promotional, encourage follow-up; depiction of consumption; video details and engagement statistics Current users survey (demographics, reason for use, and use of reviews) 	83 videos
Glowacki et al [38], 2018	Twitter (August to October 2016)	<ul style="list-style-type: none"> Statistical analysis on opioid-related tweets: <ul style="list-style-type: none"> Clustering algorithm to find topics Analysis of trending hashtags, top influencers, and location of tweets 	73,235 tweets
Meacham et al [39], 2018	Reddit (January 2010 to December 2016)	<ul style="list-style-type: none"> Analysis of modes of cannabis use mentions on Twitter on the following: <ul style="list-style-type: none"> Most frequent words Mentions of adverse effects Subjective highness 	400,000 posts
Leas et al [40], 2019	Google Trends (January 2004 to April 2019)	<ul style="list-style-type: none"> Analysis on CBD^c and cannabidiol terms to evaluate public interest 	Not available
Meacham et al [41], 2019	Reddit (January 2017 to December 2019)	<ul style="list-style-type: none"> Content analysis of dabbing-related questions on the following: <ul style="list-style-type: none"> Topics of questions After engagement and the types and sentiment of information 	193 questions
Nasralah et al [42], 2019	Twitter (January 2015 to February 2019)	<ul style="list-style-type: none"> Analysis of opioid-dependent user's tweets: <ul style="list-style-type: none"> Thematic analysis of conversations Demographic analysis 	20,609 tweets

Study	Source (duration)	Analysis	Number of items analyzed
Pérez-Pérez et al [43], 2019	Twitter (February to August 2018)	<ul style="list-style-type: none"> • Lexicon- and rule-based analysis of bowel disease tweets on sentiments, network, gender, geolocation, symptoms, and food 	24,634 tweets
Shi et al [44], 2019	Google Trends and Buzzsumo (January 2011 to July 2018)	<ul style="list-style-type: none"> • Google Trends analysis on cancer therapies to evaluate interest in cannabis vs other therapies 	Not available
Allem et al [45], 2020	Twitter (May to December 2018)	<ul style="list-style-type: none"> • Topic analysis of cannabis-related tweets 	60,861 nonbot and 8874 bot tweets
Janmohamed et al [46], 2020	Blogs, news, forums, and <1% other (August 2019 to April 2021)	<ul style="list-style-type: none"> • Topic modeling on vaping-related conversations: <ul style="list-style-type: none"> • Analysis of word prevalence • Analysis of change of topics over time 	4,027,172 documents or blogs
Jia et al [47], 2020	Google, Facebook, and YouTube (September 2019)	<ul style="list-style-type: none"> • Content analysis of glaucoma and CBD posts on the following: <ul style="list-style-type: none"> • General discussion, information sharing, personal story, question, answer, and moderator comment • Quality of information • Source of information being professional or not and whether an opinion on glaucoma and medical cannabis use was expressed • Analysis of professional accounts 	51 Google websites, 126 Facebook posts, and 37 YouTube videos
Leas et al [48], 2020	Reddit (January 2014 to August 2019)	<ul style="list-style-type: none"> • Content analysis of reasons for CBD use: <ul style="list-style-type: none"> • Reasons for personal use (condition and wellness) • Analysis based on categorized diagnosable conditions 	104,917 posts
Merten et al [49], 2020	Pinterest (July 31, August 18, and September 1, 2018)	<ul style="list-style-type: none"> • Content analysis of CBD and cannabidiol posts on the following: <ul style="list-style-type: none"> • Mentions of mental and physical benefits • Emotional appeal analysis • Engagement statistics 	1280 pins
Mullins et al [50], 2020	Twitter (June to July 2017)	<ul style="list-style-type: none"> • Analysis of Ireland pain-related tweets on: <ul style="list-style-type: none"> • Topic analysis: sentiment analysis, analysis of most frequently occurring keywords, demographic analysis, and personal use analysis 	941 tweets
Saposnik and Huber [51], 2020	Google Trends (January 2004 to December 2019)	<ul style="list-style-type: none"> • Google Trends analysis on autism and cannabis to analyze trends in search volume about the causes and treatments of Autism spectrum disorder over time 	Not available
Song et al [52], 2020	GoFundMe (January 2012 to December 2019)	<ul style="list-style-type: none"> • Content analysis of alternative medicine and cancer campaigns on the following: <ul style="list-style-type: none"> • Themes of patient narratives • Types of alternative treatments used • Demographics (gender, cancer type, cancer stage, insurance status, past treatment, future treatment, and alternative treatment) 	1474 campaigns
Tran and Kavuluru [53], 2020	Reddit and or FDA comments (January to April 2019)	<ul style="list-style-type: none"> • Content analysis on CBD posts for therapeutic effects and popular modes of consumption compared with FDA^d comments 	64,099 Reddit and 3832 FDA comments
van Draanen et al [54], 2020	Twitter (January 2017 to June 2019)	<ul style="list-style-type: none"> • Cannabis-related US and Canada posts: <ul style="list-style-type: none"> • Topic modeling • Sentiment analysis based on cannabis legalization policies 	1,200,127 tweets
Zenone et al [55], 2020	GoFundMe (January 2017 to March 2019)	<ul style="list-style-type: none"> • Thematic analysis of cancer and cannabis campaigns: <ul style="list-style-type: none"> • Efficacy claims • Treatment regimen classification • CBD efficacy presentation • Content analysis for Other: cancer stage, raised money, and number of donors 	155 campaigns

Study	Source (duration)	Analysis	Number of items analyzed
Pang et al [56], 2021	Twitter (December 2019 to December 2020)	<ul style="list-style-type: none"> Thematic analysis of pregnancy- and cannabis-related tweets for safety during pregnancy, safety postpartum, and pregnancy-related symptoms 	17,238 tweets
Rhidenour et al [57], 2021	Reddit (January 2008 to December 2018)	<ul style="list-style-type: none"> Thematic analysis of veteran's cannabis posts on the following: <ul style="list-style-type: none"> Point of view, reasons for use, prescription drug use, or other substance use Test, legality, legal policy, and doctor-patient conversation 	974 posts
Smolev et al [58], 2021	Facebook (November 2018 to November 2019)	<ul style="list-style-type: none"> Thematic analysis of traumatic brachial plexus injury posts on: antio-pioid sentiment, preference for alternative options, and antigabapentin sentiment 	7694 posts
Soleymanpour et al [59], 2021	Twitter (July 2019)	<ul style="list-style-type: none"> Analysis of CBD marketing tweets and therapeutic claims 	2,200,000 tweets
Zenone et al [60], 2021	GoFundMe (June 2017 to May 2019)	<ul style="list-style-type: none"> Thematic analysis for informational pathways: self-directed research, recommendations from a trusted care provider, and insights shared by someone associated with or influencing the crowd funders personal network Content analysis for intended outcome, social media shares, number of donors, total requested, and total received 	164 campaigns
Turner et al [62] 2021	Twitter (October 2019 to January 2020)	<ul style="list-style-type: none"> Analysis of personal and commercial CBD-related tweets; term and sentiment analysis 	167,755 personal 143,322 commercial tweets
Allem et al [61], 2022	Twitter (January to September 2020)	<ul style="list-style-type: none"> Analysis of cannabis-related conversation for health-related motivations or perceived adverse health effects 	353,353 tweets
Meacham et al [63] 2022	Reddit (December 2015 to August 2019)	<ul style="list-style-type: none"> Analysis of cannabis-related posts from an opioid use and an opioid recovery subreddit 	908 posts from opioid re-covery subreddits and 4224 posts from opioid use subreddits

^aADHD: attention-deficit hyperactivity disorder.

^bPTSD: posttraumatic stress disorder.

^cCBD: cannabidiol.

^dFDA: Food and Drug Administration.

Table 2. Publications per year (n=42).

Year	Count, n (%)
2014	1 (2)
2015	5 (12)
2016	3 (7)
2017	6 (14)
2018	3 (7)
2019	5 (12)
2020	11 (26)
2021	6 (14)
2022	2 (5)

Table 3. Publications per data source (n=42).

Source	Count, n (%)
Twitter	17 (41)
Reddit	6 (14)
Web-based forums	6 (14)
GoFundMe	3 (7)
YouTube	3 (7)
Google Trends	3 (7)
Google, Facebook, and YouTube	1 (2)
Bing Search Engine	1 (2)
Facebook	1 (2)
Pinterest	1 (2)

Social Media Data Collection Strategies

Some studies obtained all their associated data from a specific subreddit [48,53,57] or a web-based forum [35] and subsequently sampled the data. Of 42 studies, 1 (2%) Twitter study collected tweets using a geolocation boundary box and then filtered the data for cannabis-related keywords [54].

Keyword-based filtering was used by many studies. Terms used for filtering were either common expressions for cannabis from dictionaries, such as Urban Dictionary, or were based on similar research in this domain. Of the 42 studies, 1 (2%) study [36] used Urban Dictionary and web forums to create a comprehensive list of 123 terms related to cannabis consumption. Another study [57] first found all the terms related to marijuana by searching on Thesaurus.com and then used the word embedding likeness perusal software [64] to generate synonyms.

In a nonmedical cannabis-related study, word embeddings created from Twitter and Reddit data sets discovered synonyms and slang terms that could not be identified using other means. The study recommends this method of synonym discovery in advance for any data collection based on keyword filtering [65].

Of the 42 studies, 3 (7%) studies were user focused, with data derived from specific highly influential users [23], opioid-dependent users [42], or a US veteran-specific subreddit [57].

The largest data set manually annotated by the researchers was collected using cannabis-related keywords and consisted of 36,939 original tweets and 10,000 retweets [27]. Apart from that study, the average size of annotated data sets was approximately 1450 records. Of the 42 studies, 2 (5%) studies [23,28] used crowdsourcing services to annotate tweets, whereas the rest conducted in-house annotation. The duration of data collection ranged from 1 month to 6 years. Of the 42 studies, 2 (5%) of these studies made their annotated data available to other researchers [30,60].

Types of Analysis

Overview

The studies included in this review used a variety of analytical methods, including qualitative analysis, quantitative content analysis, machine learning, rule-based, and statistical analysis. The types of analysis include sentiment assessment, thematic analysis, content analysis, named entity recognition, social networks, and geographic analysis. Table S3 in [Multimedia Appendix 1](#) summarizes the analyses.

Discovering Themes

Themes were identified in 62% (26/42) of the studies. Manual coding of the themes was performed by 69% (18/26) of the studies, either by using pre-existing categories or by observing a sample of the data and generating a codebook [22,23,25,26,28,30,31,37,41,47-49,52,55-58,60]. Of the 26 studies, 2 (7%) studies used the services of social media data analytics companies [42,50].

Of the 26 studies, 4 (15%) studies used topic modeling to infer themes or topics [34,35,46,54]. The algorithm of choice for this task is the latent Dirichlet allocation [66]. The choice of the number of topics was based on intrinsic evaluation metrics (eg, coherence and perplexity) and iterative qualitative analysis informed by prior experience with topic models. Of the 26 studies, 1 (4%) study used temporal topic modeling techniques to study changes in topics over time, with the goal of analyzing how web-based vaping narratives changed during the COVID-19 pandemic [46].

Of the 26 studies, 1 (4%) study identified themes by using rule-based methods. Frequency counts of the most common unigrams and bigrams were generated and formed the basis of the topics [45]. Another study used SAS Text Miner software, a text-topic node algorithm, to discover topics [38].

Demographic Analysis

Socioeconomic and demographic analyses of the study population were performed in 26% (11/42) of the studies. Of the 11 studies, 2 (27%) studies used the provided gender, age, and other user characteristics from user profiles or inferred from posts by users [33,52]. Of the 11 studies, 2 (27%) video-based

studies used the perceived age and gender of the subjects after observing the videos [25,26].

Of the 11 studies, 2 (18%) studies that used social media analytics providers obtained age and gender data by using the supplied analysis [42,50]. Of the 11 studies, 2 (18%) of the Twitter-based studies used a commercial tool called DemographicsPro, which uses proprietary algorithms to infer user demographic characteristics [23,28]. Other studies used existing census data [32], demographic information obtained from survey data [37], and a 2-step method based on a gender-name lexicon and a face recognition algorithm applied to users' profile information to identify the users' gender [43].

Geographic Analysis

Geolocation data analysis was performed in 40% (17/42) of the studies. User profiles or message metadata were used in 52% (9/17) of the studies [24,29,32,34,36,43,54,55,60]. Of the 17 studies, 2 (12%) studies used information provided by social media analytics companies [38,50]. The DemographicsPro tool was used in 5% (1/17) of studies [28]. Of the 17 studies, 3 (17%) studies used location information provided by Google Trends [40,44,51]. Another (1/17, 5%) study collected geographical information from survey data [37]. Of the 17 studies, 1 (5%) video-based study used the geographic location of video channels [26].

Sentiment Assessment

An individual's perception of a topic can be characterized as having a positive, negative, or neutral sentiment. The analysis of these sentiments is often performed using automated language tools and is named "sentiment analysis" [67].

Out of the 12 studies that performed sentiment analysis, 5 (42%) used automated methods. Of the 12 studies, 1 (8%) study trained a binary Naive Bayes classifier on a sample of 1000 "marijuana" related tweets to classify posts into 2 opinion polarities, positive and negative or neutral [32]. Another study used sentiment analysis provided by a social media analytics company [50]. Of the 12 studies, 3 (25%) studies used Valence Aware Dictionary and Sentiment Reasoner (VADER) [68], a lexicon and rule-based sentiment analysis tool [43,54,62]. The VADER performance was compared with in-house machine learning classifiers trained on 3000 manually coded cannabis-related tweets, which showed a 30% performance improvement over VADER. Although VADER is widely used for general tweet sentiment analysis, its performance suffers in substance-use-related domains where negative words are often used to carry positive sentiments. For example, "I took CBD oil, that stuff was bad" [69]—in this sentence, "bad" actually means good.

User Analysis

For conducting user analysis, 57% (24/42) of the studies examined either the subject of the posts, as from individuals or others (ie, from self, retail, media, or professionals), or who the post was about (self, others, or general) [22,23,25-29,33,37,41-43,45,47-50,52,55,57,58,60-62].

When manual data labeling was performed, the determination of both the poster and subject of the post was part of the labeling

process. Self-reporting and self-use were easily determined by observation of videos, as were most texts based on the structure of the language. For example, a study [27] first identified whether the subject of the tweet was about the self, other, or general and then identified whether the tweets were about actual cannabis use. This study included further categories of tone, related behavior, perceived impact, and social context. Automated labeling approaches look for phrases that indicate self-reporting. For example, a study on opioid addiction [42] looked for phrases such as "I am addicted" and "I have been addicted" in the context of opioid mentions. Classifiers were used in another study [59] to separate marketing tweets from nonmarketing tweets; however, their focus was on marketing tweets.

None of the studies used advanced natural language processing techniques to establish subjects and personal mentions. Social media bots are automated accounts that generate artificial activities on social media platforms [18]. Bot detection was used in only 4% (1/24) of studies, which used Twitter as a data source [45].

Other Analyses

Of the 42 studies, 2 (5%) studies examined the social networks of contributors to conversations. This allowed the identification of target communities and user interactions [34,43]. Of the 42 studies, 3 (7%) studies examined the impact of governmental cannabis legalization policies on the sentiments and opinions of people or on the volume of social media posts [24,28,54]. Term frequency and count analysis of words and phrases was performed in 12% (5/42) of studies [29,39,50,62,63].

Ethical Considerations

Institutional review boards (or their equivalents) ensured that research using human participants is conducted in an ethical manner [70]. Approval for and overseeing of a study by an institutional review board ensures that researchers adopt an ethically appropriate research protocol that respects the rights and interests of social media users; 62% (26/42) of the studies mentioned an ethics approval review being sought or the study being exempt from ethics requirements. There was no mention of ethics approval in 38% (16/42) of the studies.

External Validity

The use of standard reporting systems, such as the US Food and Drug Administration reports, helps to assess whether social media research findings can be generalized to real-world data. When a suitable ground-truth data set is not available, validating results against >1 social media platform improves the generalizability and validity of the results. Only a few studies used >1 social media data source or validated their findings against other data sources. Of the 42 studies, 2 (5%) studies used Food and Drug Administration data as an external ground-truth data source to validate their results [36,53]. Of the 42 studies, 1 (2%) study analyzed several web-based forums [31], and 2 (5%) other studies used several social media platforms as their data sources [22,47].

Discussion

In this study, we reviewed the technical aspects of peer-reviewed published works that used social media and other forms of user-generated data to understand the medicinal use of cannabis. All the studies concluded that these consumer-generated data sources are useful and provide a complementary resource for studying cannabis and medical conditions for which cannabis is used.

Principal Findings

The findings of this study are presented by answering the RQs.

RQ1: What Consumer-Generated Data Sources Are Used for Studying Cannabis?

Sources of consumer-generated data for cannabis research used by the reviewed studies include social media platforms, such as Twitter, Reddit, and YouTube; search queries, including Google Trend and Bing query logs; and web-based forums, crowdfunding platforms, blogs, and websites. Twitter was used in most of the studies. One of the studies concluded that, compared with unmoderated platforms, moderated sites focused more on evidence-based information and controlled misleading content [22].

RQ2: What Common Techniques for Collection and Analysis of Data Are Used?

Some studies have used social media analytics companies for some or all of their data collection and processing tasks. Other studies used application program interfaces to interact with Twitter and Reddit. Although Facebook allows researchers to access public posts from public pages through a dedicated platform [71], 2% (1/42) of studies [58] analyzed private Facebook posts—the method used to obtain data was not reported.

Approximately half of the studies used data sets of <8000 records and many of them used 1000 records. These studies either focused on understanding the characteristics and needs of users or the quality of information on the web, or they were directed by an RQ such as “Are individuals using CBD for diagnoseable conditions which have evidence-based therapies?” These analyses play a critical role in understanding the domain but are difficult to replicate and generalize.

More recent neural network–based natural language processing techniques have not been used in the studies in this review. These modern machine learning methods have the advantage that they require minimal data preparation and are characterized by the capacity to learn the nuance of language. However, to function effectively, they typically require high-quality annotated data—a scarce and expensive resource. Textual social media data are highly amenable to these techniques. Creating and sharing deidentified annotated data sets for this purpose should be encouraged within appropriate ethical, regulatory, and legal frameworks [72].

RQ3: What Are the Common Limitations and Challenges Faced by the Studies?

These limitations are mentioned in order of frequency.

Sample Representativeness

Most research on social media uses samples of available data. However, the extent to which the data samples are representative of the general population is often unclear. The limiting factors mentioned in these studies include sampling bias that is introduced as a result of the choice of keywords, data collection duration, and population biases.

Population biases often refer to the demographic composition of people using social media platforms being different from the general population and the difficulties in determining the demographic characteristics of users. Accessing accurate geographical locations has also been mentioned in previous studies. Obtaining these data is limited because even when users explicitly include demographic information (eg, with Facebook) or geographical information in their posts or profiles, these may be fabricated.

The choice of platform itself also imposes limitations. For instance, platform-specific features, such as sampling strategies, limit the amount of data that can be collected and the behavior and conversation of users depending on the platform or context. Of the 42 studies, 1 (2%) study mentioned that the forums they investigated could be very procannabis and are likely inhabited by more experienced cannabis users [41]. Another study stated that individuals posting on YouTube about cannabis are likely to seek social networking opportunities [37].

Complications also arise because platform-specific algorithms spot and further promote popular themes and users to deliberately manage behavior and attract more platform engagement. This needs to be ameliorated by detecting and accounting for the algorithms and potentially by sampling from >1 platform.

Methodology Constraints

The use of small data sets by some of the studies impacts the generalizability of the results, and some of the researchers acknowledged this and indicated a plan to replicate their studies with more data and the use of automated methods. Consequently, we observed that although such studies may be sampling social media data for hypothesis generation, they do not leverage one of the most important features of social media data, which is the ability to observe the continuous generation of big data to create long-term data-centric insights [73].

Biases that could have been introduced by the choice of theme were also mentioned in the studies. Most researchers have attempted to mitigate this by creating annotation guidelines, having >1 person labeling data, and resolving disagreements.

Actual Use Detection

A limitation mentioned by several studies is that web-based search activities and social media posts containing cannabis-related keywords do not necessarily represent the actual use of cannabis by the poster. Depending on the context and goals of the research (for instance, if the research seeks to study a cannabis-consuming population), advanced text processing techniques are required to establish when personal cannabis use can be inferred. For such studies, establishing its use should be a crucial initial step. However, the detection of

personal use is challenging, especially in the informal, diverse, and specialized language used by niche communities.

Source Identification

Identifying the source of posts (ie, whether they were generated organically by individual users or by organizations or bots) was a commonly mentioned limitation. Content generated by health and commercial organizations, power users, and nonindividual accounts was understood to comprise a considerable amount of social media post volume on the web.

Limitations

This review used 4 literature databases in the search process to allow the maximum coverage of existing publications. However, we cannot be certain that we have covered all relevant publications. The choice of keywords for the literature search could also have impacted capturing all the relevant studies in this domain, for instance, *infodemiology* and *infoveillance* were not in the keywords. Articles included in this study were selected following a systematic approach and underwent a bias assessment for quality; however, biases could not be completely avoided. This study was also limited to English-language articles.

Conclusions

The number of studies in this field has steadily increased over the last few years. Social media conversations are wide ranging

and offer opportunities for insights that cannot be obtained through formal information gathering. Researchers have realized the value of social media conversations as a place for users to freely express their experiences and concerns without risking judgment or penalty and that social media is the natural forum for many users of cannabis as medicine to share their insights into the benefits and issues they experience and perceive.

Manual qualitative analysis, statistical analysis, supervised and unsupervised machine learning, and rule-based methods are among methodologies used in these studies. Analyses of social media data that are limited to small data samples, although providing an effective means of hypothesis generation, are difficult to reliably reproduce and generalize. Where possible, the sharing of high-quality deidentified annotated data to allow the use of generalizable analytical techniques should be encouraged to advance this field.

To improve their validity and generalizability, studies could add additional social media data sources and check their results against established reporting systems. Studies could take advantage of emerging data analysis strategies that leverage big data, such as deep learning and transfer-learning-based approaches.

Acknowledgments

This review was supported by the Australian Centre for Cannabinoid Clinical and Research Excellence, funded by the National Health and Medical Research Council through the Centre of Research Excellence scheme (NHMRC CRE APP1135054).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supporting information (review keywords, inclusion and exclusion criteria, papers summary).

[\[OCX File, 45 KB-Multimedia Appendix 1\]](#)

References

1. Li HL. An archaeological and historical account of cannabis in China. *Econ Bot* 1973 Oct;28(4):437-448 [[FREE Full text](#)] [doi: [10.1007/BF02862859](https://doi.org/10.1007/BF02862859)]
2. Hallinan CM, Gunn JM, Bonomo YA. Implementation of medicinal cannabis in Australia: innovation or upheaval? Perspectives from physicians as key informants, a qualitative analysis. *BMJ Open* 2021 Oct 22;11(10):e054044 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2021-054044](https://doi.org/10.1136/bmjopen-2021-054044)] [Medline: [34686558](https://pubmed.ncbi.nlm.nih.gov/34686558/)]
3. Banerjee S, McCormack S. Medical Cannabis for the Treatment of Chronic Pain: A Review of Clinical Effectiveness and Guidelines. Ottawa, Canada: Canadian Agency for Drugs and Technologies in Health; 2019.
4. Kleeman-Forsthuber LT, Dennis DA, Jennings JM. Medicinal cannabis in orthopaedic practice. *J Am Acad Orthop Surg* 2020 May 01;28(7):268-277. [doi: [10.5435/JAAOS-D-19-00438](https://doi.org/10.5435/JAAOS-D-19-00438)] [Medline: [31800438](https://pubmed.ncbi.nlm.nih.gov/31800438/)]
5. Pawliuk C, Chau B, Rassekh SR, McKellar T, Siden HH. Efficacy and safety of paediatric medicinal cannabis use: a scoping review. *Paediatr Child Health* 2021 Jul;26(4):228-233 [[FREE Full text](#)] [doi: [10.1093/pch/pxaa031](https://doi.org/10.1093/pch/pxaa031)] [Medline: [34131459](https://pubmed.ncbi.nlm.nih.gov/34131459/)]
6. Pratt M, Stevens A, Thuku M, Butler C, Skidmore B, Wieland LS, et al. Benefits and harms of medical cannabis: a scoping review of systematic reviews. *Syst Rev* 2019 Dec 10;8(1):320 [[FREE Full text](#)] [doi: [10.1186/s13643-019-1243-x](https://doi.org/10.1186/s13643-019-1243-x)] [Medline: [31823819](https://pubmed.ncbi.nlm.nih.gov/31823819/)]
7. Martin JH, Lucas C. Reporting adverse drug events to the Therapeutic Goods Administration. *Aust Prescr* 2021 Mar;44(1):2-3 [[FREE Full text](#)] [doi: [10.18773/austprescr.2020.077](https://doi.org/10.18773/austprescr.2020.077)] [Medline: [33664539](https://pubmed.ncbi.nlm.nih.gov/33664539/)]
8. EudraVigilance. European Medicines Agency. URL: <https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance> [accessed 2022-05-24]

9. FDA Adverse Event Reporting System (FAERS) Public Dashboard. U.S. Food & Drug Administration. 2021. URL: <https://tinyurl.com/yh22mc2c> [accessed 2022-05-24]
10. Al Dweik R, Stacey D, Kohen D, Yaya S. Factors affecting patient reporting of adverse drug reactions: a systematic review. *Br J Clin Pharmacol* 2017 Apr;83(4):875-883 [FREE Full text] [doi: [10.1111/bcp.13159](https://doi.org/10.1111/bcp.13159)] [Medline: [27868226](https://pubmed.ncbi.nlm.nih.gov/27868226/)]
11. Chandra S, Lata H, ElSohly MA, Walker LA, Potter D. Cannabis cultivation: methodological issues for obtaining medical-grade product. *Epilepsy Behav* 2017 May;70(Pt B):302-312. [doi: [10.1016/j.yebeh.2016.11.029](https://doi.org/10.1016/j.yebeh.2016.11.029)] [Medline: [28202406](https://pubmed.ncbi.nlm.nih.gov/28202406/)]
12. Hakkarainen P, Frank VA, Barratt MJ, Dahl HV, Decorte T, Karjalainen K, et al. Growing medicine: small-scale cannabis cultivation for medical purposes in six different countries. *Int J Drug Policy* 2015 Mar;26(3):250-256. [doi: [10.1016/j.drugpo.2014.07.005](https://doi.org/10.1016/j.drugpo.2014.07.005)] [Medline: [25123694](https://pubmed.ncbi.nlm.nih.gov/25123694/)]
13. Paul MJ, Dredze M. Social monitoring for public health. In: Marchionini G, editor. *Synthesis Lectures on Information Concepts, Retrieval, and Services*. San Rafael, CA, USA: Morgan and Claypool Publishers; Aug 31, 2017:1-183.
14. Bode L, Davis-Kean P, Singh L, Berger-Wolf T, Budak C, Chi G, et al. Study designs for quantitative social science research using social media. *PsyArXiv* 2020:1-27. [doi: [10.31234/osf.io/zp8q2](https://doi.org/10.31234/osf.io/zp8q2)]
15. Mneimneh Z, Pasek J, Singh L, Best R, Bode L, Bruch E, et al. Data acquisition, sampling, and data preparation considerations for quantitative social science research using social media data. *PsyArXiv* 2021 Mar 15:1-45. [doi: [10.31234/osf.io/k6vyj](https://doi.org/10.31234/osf.io/k6vyj)]
16. Conway M, Hu M, Chapman WW. Recent advances in using natural language processing to address public health research questions using social media and consumer-generated data. *Yearb Med Inform* 2019 Aug;28(1):208-217 [FREE Full text] [doi: [10.1055/s-0039-1677918](https://doi.org/10.1055/s-0039-1677918)] [Medline: [31419834](https://pubmed.ncbi.nlm.nih.gov/31419834/)]
17. Allem JP, Ferrara E. Could social bots pose a threat to public health? *Am J Public Health* 2018 Aug;108(8):1005-1006. [doi: [10.2105/AJPH.2018.304512](https://doi.org/10.2105/AJPH.2018.304512)] [Medline: [29995482](https://pubmed.ncbi.nlm.nih.gov/29995482/)]
18. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. *Commun ACM* 2016 Jun 24;59(7):96-104. [doi: [10.1145/2818717](https://doi.org/10.1145/2818717)]
19. Varol O, Ferrara E, Davis CA, Menczer F, Flammini A. Online human-bot interactions: detection, estimation, and characterization. *arXiv* 2017.
20. Khademi Habibabadi S, Bonomo YA, Conway M, Hallinan CM. Social media discourse and internet search queries on cannabis as a medicine: A systematic scoping review. *medRxiv* 2022. [doi: [10.1101/2022.05.16.22275171](https://doi.org/10.1101/2022.05.16.22275171)]
21. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009 Jul 21;339:b2700 [FREE Full text] [doi: [10.1136/bmj.b2700](https://doi.org/10.1136/bmj.b2700)] [Medline: [19622552](https://pubmed.ncbi.nlm.nih.gov/19622552/)]
22. McGregor F, Somner JE, Bourne RR, Munn-Giddings C, Shah P, Cross V. Social media use by patients with glaucoma: what can we learn? *Ophthalmic Physiol Opt* 2014 Jan;34(1):46-52. [doi: [10.1111/opo.12093](https://doi.org/10.1111/opo.12093)] [Medline: [24325434](https://pubmed.ncbi.nlm.nih.gov/24325434/)]
23. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Gruzca RA, Bierut LJ. Twitter chatter about marijuana. *J Adolesc Health* 2015 Mar;56(2):139-145 [FREE Full text] [doi: [10.1016/j.jadohealth.2014.10.270](https://doi.org/10.1016/j.jadohealth.2014.10.270)] [Medline: [25620299](https://pubmed.ncbi.nlm.nih.gov/25620299/)]
24. Daniulaityte R, Nahhas RW, Wijeratne S, Carlson RG, Lamy FR, Martins SS, et al. "Time for dabs": analyzing Twitter data on marijuana concentrates across the U.S. *Drug Alcohol Depend* 2015 Oct 01;155:307-311 [FREE Full text] [doi: [10.1016/j.drugalcdep.2015.07.1199](https://doi.org/10.1016/j.drugalcdep.2015.07.1199)] [Medline: [26338481](https://pubmed.ncbi.nlm.nih.gov/26338481/)]
25. Gonzalez-Estrada A, Cuervo-Pardo L, Ghosh B, Smith M, Pazheri F, Zell K, et al. Popular on YouTube: a critical appraisal of the educational quality of information regarding asthma. *Allergy Asthma Proc* 2015;36(6):e121-e126. [doi: [10.2500/aap.2015.36.3890](https://doi.org/10.2500/aap.2015.36.3890)] [Medline: [26534743](https://pubmed.ncbi.nlm.nih.gov/26534743/)]
26. Krauss MJ, Sowles SJ, Mylvaganam S, Zewdie K, Bierut LJ, Cavazos-Rehg PA. Displays of dabbing marijuana extracts on YouTube. *Drug Alcohol Depend* 2015 Oct 01;155:45-51 [FREE Full text] [doi: [10.1016/j.drugalcdep.2015.08.020](https://doi.org/10.1016/j.drugalcdep.2015.08.020)] [Medline: [26347408](https://pubmed.ncbi.nlm.nih.gov/26347408/)]
27. Thompson L, Rivara FP, Whitehill JM. Prevalence of marijuana-related traffic on Twitter, 2012-2013: a content analysis. *Cyberpsychol Behav Soc Netw* 2015 Jul;18(6):311-319 [FREE Full text] [doi: [10.1089/cyber.2014.0620](https://doi.org/10.1089/cyber.2014.0620)] [Medline: [26075917](https://pubmed.ncbi.nlm.nih.gov/26075917/)]
28. Cavazos-Rehg PA, Sowles SJ, Krauss MJ, Agbonavbare V, Gruzca R, Bierut L. A content analysis of tweets about high-potency marijuana. *Drug Alcohol Depend* 2016 Oct 01;166:100-108 [FREE Full text] [doi: [10.1016/j.drugalcdep.2016.06.034](https://doi.org/10.1016/j.drugalcdep.2016.06.034)] [Medline: [27402550](https://pubmed.ncbi.nlm.nih.gov/27402550/)]
29. Lamy FR, Daniulaityte R, Sheth A, Nahhas RW, Martins SS, Boyer EW, et al. "Those edibles hit hard": exploration of Twitter data on cannabis edibles in the U.S. *Drug Alcohol Depend* 2016 Jul 01;164:64-70 [FREE Full text] [doi: [10.1016/j.drugalcdep.2016.04.029](https://doi.org/10.1016/j.drugalcdep.2016.04.029)] [Medline: [27185160](https://pubmed.ncbi.nlm.nih.gov/27185160/)]
30. Mitchell JT, Sweitzer MM, Tunno AM, Kollins SH, McClernon FJ. "I Use Weed for My ADHD": a qualitative analysis of online forum discussions on cannabis use and ADHD. *PLoS One* 2016 May 26;11(5):e0156614 [FREE Full text] [doi: [10.1371/journal.pone.0156614](https://doi.org/10.1371/journal.pone.0156614)] [Medline: [27227537](https://pubmed.ncbi.nlm.nih.gov/27227537/)]
31. Andersson M, Persson M, Kjellgren A. Psychoactive substances as a last resort—a qualitative study of self-treatment of migraine and cluster headaches. *Harm Reduct J* 2017 Sep 05;14(1):60 [FREE Full text] [doi: [10.1186/s12954-017-0186-6](https://doi.org/10.1186/s12954-017-0186-6)] [Medline: [28870224](https://pubmed.ncbi.nlm.nih.gov/28870224/)]

32. Dai H, Hao J. Mining social media data on marijuana use for Post Traumatic Stress Disorder. *Comput Human Behav* 2017 May;70(C):282-290. [doi: [10.1016/j.chb.2016.12.064](https://doi.org/10.1016/j.chb.2016.12.064)]
33. Greiner C, Chatton A, Khazaal Y. Online self-help forums on cannabis: a content assessment. *Patient Educ Couns* 2017 Oct;100(10):1943-1950. [doi: [10.1016/j.pec.2017.06.001](https://doi.org/10.1016/j.pec.2017.06.001)] [Medline: [28602568](https://pubmed.ncbi.nlm.nih.gov/28602568/)]
34. Turner J, Kantardzic M. Geo-social analytics based on spatio-temporal dynamics of marijuana-related tweets. In: *Proceedings of the 2017 International Conference on Information System and Data Mining*. 2017 Presented at: ICISDM '17; April 1-3, 2017; Charleston, SC, USA p. 28-38. [doi: [10.1145/3077584.3077588](https://doi.org/10.1145/3077584.3077588)]
35. Westmaas JL, McDonald BR, Portier KM. Topic modeling of smoking- and cessation-related posts to the American Cancer Society's Cancer Survivor Network (CSN): implications for cessation treatment for cancer survivors who smoke. *Nicotine Tob Res* 2017 Aug 01;19(8):952-959. [doi: [10.1093/ntr/ntx064](https://doi.org/10.1093/ntr/ntx064)] [Medline: [28340059](https://pubmed.ncbi.nlm.nih.gov/28340059/)]
36. Yom-Tov E, Lev-Ran S. Adverse reactions associated with cannabis consumption as evident from search engine queries. *JMIR Public Health Surveill* 2017 Oct 26;3(4):e77 [FREE Full text] [doi: [10.2196/publichealth.8391](https://doi.org/10.2196/publichealth.8391)] [Medline: [29074469](https://pubmed.ncbi.nlm.nih.gov/29074469/)]
37. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Murphy GM, Bierut LJ. Exposure to and content of marijuana product reviews. *Prev Sci* 2018 Feb;19(2):127-137 [FREE Full text] [doi: [10.1007/s11121-017-0818-9](https://doi.org/10.1007/s11121-017-0818-9)] [Medline: [28681195](https://pubmed.ncbi.nlm.nih.gov/28681195/)]
38. Glowacki EM, Glowacki JB, Wilcox GB. A text-mining analysis of the public's reactions to the opioid crisis. *Subst Abus* 2018;39(2):129-133. [doi: [10.1080/08897077.2017.1356795](https://doi.org/10.1080/08897077.2017.1356795)] [Medline: [28723265](https://pubmed.ncbi.nlm.nih.gov/28723265/)]
39. Meacham MC, Paul MJ, Ramo DE. Understanding emerging forms of cannabis use through an online cannabis community: an analysis of relative post volume and subjective highness ratings. *Drug Alcohol Depend* 2018 Jul 01;188:364-369 [FREE Full text] [doi: [10.1016/j.drugalcdep.2018.03.041](https://doi.org/10.1016/j.drugalcdep.2018.03.041)] [Medline: [29883950](https://pubmed.ncbi.nlm.nih.gov/29883950/)]
40. Leas EC, Nobles AL, Caputi TL, Dredze M, Smith DM, Ayers JW. Trends in Internet searches for cannabidiol (CBD) in the United States. *JAMA Netw Open* 2019 Oct 02;2(10):e1913853 [FREE Full text] [doi: [10.1001/jamanetworkopen.2019.13853](https://doi.org/10.1001/jamanetworkopen.2019.13853)] [Medline: [31642924](https://pubmed.ncbi.nlm.nih.gov/31642924/)]
41. Meacham MC, Roh S, Chang JS, Ramo DE. Frequently asked questions about dabbing concentrates in online cannabis community discussion forums. *Int J Drug Policy* 2019 Dec;74:11-17 [FREE Full text] [doi: [10.1016/j.drugpo.2019.07.036](https://doi.org/10.1016/j.drugpo.2019.07.036)] [Medline: [31400582](https://pubmed.ncbi.nlm.nih.gov/31400582/)]
42. Nasralah T, El-gayar O, Wang Y. What social media can tell us about opioid addicts: Twitter data case analysis. In: *Proceedings of the 25th Americas' Conference on Information Systems*. 2019 Presented at: AMCIS '19; August 15-17, 2019; Cancún, Mexico p. 15.
43. Pérez-Pérez M, Pérez-Rodríguez G, Fdez-Riverola F, Lourenço A. Using Twitter to understand the human bowel disease community: exploratory analysis of key topics. *J Med Internet Res* 2019 Aug 15;21(8):e12610 [FREE Full text] [doi: [10.2196/12610](https://doi.org/10.2196/12610)] [Medline: [31411142](https://pubmed.ncbi.nlm.nih.gov/31411142/)]
44. Shi S, Brant A, Sabolch A, Pollom E. False news of a cannabis cancer cure. *Cureus* 2019 Jan 19;11(1):e3918 [FREE Full text] [doi: [10.7759/cureus.3918](https://doi.org/10.7759/cureus.3918)] [Medline: [30931189](https://pubmed.ncbi.nlm.nih.gov/30931189/)]
45. Allem JP, Escobedo P, Dharmapuri L. Cannabis surveillance with Twitter data: emerging topics and social bots. *Am J Public Health* 2020 Mar;110(3):357-362. [doi: [10.2105/AJPH.2019.305461](https://doi.org/10.2105/AJPH.2019.305461)] [Medline: [31855475](https://pubmed.ncbi.nlm.nih.gov/31855475/)]
46. Janmohamed K, Soale AN, Forastiere L, Tang W, Sha Y, Demant J, et al. Intersection of the Web-based vaping narrative with COVID-19: topic modeling study. *J Med Internet Res* 2020 Oct 30;22(10):e21743 [FREE Full text] [doi: [10.2196/21743](https://doi.org/10.2196/21743)] [Medline: [33001829](https://pubmed.ncbi.nlm.nih.gov/33001829/)]
47. Jia JS, Mehran N, Purgert R, Zhang QE, Lee D, Myers JS, et al. Marijuana and glaucoma: a social media content analysis. *Ophthalmol Glaucoma* 2021;4(4):400-404. [doi: [10.1016/j.ogla.2020.11.004](https://doi.org/10.1016/j.ogla.2020.11.004)] [Medline: [33242684](https://pubmed.ncbi.nlm.nih.gov/33242684/)]
48. Leas EC, Hendrickson EM, Nobles AL, Todd R, Smith DM, Dredze M, et al. Self-reported cannabidiol (CBD) use for conditions with proven therapies. *JAMA Netw Open* 2020 Oct 01;3(10):e2020977 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.20977](https://doi.org/10.1001/jamanetworkopen.2020.20977)] [Medline: [33057645](https://pubmed.ncbi.nlm.nih.gov/33057645/)]
49. Merten JW, Gordon BT, King JL, Pappas C. Cannabidiol (CBD): perspectives from pinterest. *Subst Use Misuse* 2020;55(13):2213-2220. [doi: [10.1080/10826084.2020.1797808](https://doi.org/10.1080/10826084.2020.1797808)] [Medline: [32715862](https://pubmed.ncbi.nlm.nih.gov/32715862/)]
50. Mullins CF, Ffrench-O'Carroll R, Lane J, O'Connor T. Sharing the pain: an observational analysis of Twitter and pain in Ireland. *Reg Anesth Pain Med* 2020 Aug;45(8):597-602. [doi: [10.1136/rapm-2020-101547](https://doi.org/10.1136/rapm-2020-101547)] [Medline: [32503862](https://pubmed.ncbi.nlm.nih.gov/32503862/)]
51. Saposnik FE, Huber JF. Trends in Web searches about the causes and treatments of autism over the past 15 years: exploratory infodemiology study. *JMIR Pediatr Parent* 2020 Dec 07;3(2):e20913 [FREE Full text] [doi: [10.2196/20913](https://doi.org/10.2196/20913)] [Medline: [33284128](https://pubmed.ncbi.nlm.nih.gov/33284128/)]
52. Song S, Cohen AJ, Lui H, Mmonu NA, Brody H, Patino G, et al. Use of GoFundMe @ to crowdfund complementary and alternative medicine treatments for cancer. *J Cancer Res Clin Oncol* 2020 Jul;146(7):1857-1865. [doi: [10.1007/s00432-020-03191-0](https://doi.org/10.1007/s00432-020-03191-0)] [Medline: [32219517](https://pubmed.ncbi.nlm.nih.gov/32219517/)]
53. Tran T, Kavuluru R. Social media surveillance for perceived therapeutic effects of cannabidiol (CBD) products. *Int J Drug Policy* 2020 Mar;77:102688 [FREE Full text] [doi: [10.1016/j.drugpo.2020.102688](https://doi.org/10.1016/j.drugpo.2020.102688)] [Medline: [32092666](https://pubmed.ncbi.nlm.nih.gov/32092666/)]
54. van Draanen J, Tao H, Gupta S, Liu S. Geographic differences in cannabis conversations on Twitter: infodemiology study. *JMIR Public Health Surveill* 2020 Oct 05;6(4):e18540 [FREE Full text] [doi: [10.2196/18540](https://doi.org/10.2196/18540)] [Medline: [33016888](https://pubmed.ncbi.nlm.nih.gov/33016888/)]
55. Zenone M, Snyder J, Caulfield T. Crowdfunding cannabidiol (CBD) for cancer: hype and misinformation on GoFundMe. *Am J Public Health* 2020 Oct;110(S3):S294-S299. [doi: [10.2105/AJPH.2020.305768](https://doi.org/10.2105/AJPH.2020.305768)] [Medline: [33001729](https://pubmed.ncbi.nlm.nih.gov/33001729/)]

56. Pang RD, Dormanesh A, Hoang Y, Chu M, Allem JP. Twitter posts about cannabis use during pregnancy and postpartum: a content analysis. *Subst Use Misuse* 2021;56(7):1074-1077. [doi: [10.1080/10826084.2021.1906277](https://doi.org/10.1080/10826084.2021.1906277)] [Medline: [33821757](https://pubmed.ncbi.nlm.nih.gov/33821757/)]
57. Rhidenour KB, Blackburn K, Barrett AK, Taylor S. Mediating medical marijuana: exploring how veterans discuss their stigmatized substance use on Reddit. *Health Commun* 2022 Sep;37(10):1305-1315. [doi: [10.1080/10410236.2021.1886411](https://doi.org/10.1080/10410236.2021.1886411)] [Medline: [33602000](https://pubmed.ncbi.nlm.nih.gov/33602000/)]
58. Smolev ET, Rolf L, Zhu E, Buday SK, Brody M, Brogan DM, et al. "Pill pushers and CBD oil"-a thematic analysis of social media interactions about pain after traumatic brachial plexus injury. *J Hand Surg Glob Online* 2021 Jan;3(1):36-40 [FREE Full text] [doi: [10.1016/j.jhsg.2020.10.005](https://doi.org/10.1016/j.jhsg.2020.10.005)] [Medline: [33537664](https://pubmed.ncbi.nlm.nih.gov/33537664/)]
59. Soleymanpour M, Saderholm S, Kavuluru R. Therapeutic claims in cannabidiol (CBD) marketing messages on Twitter. *Proceedings (IEEE Int Conf Bioinformatics Biomed)* 2021 Dec;2021:3083-3088 [FREE Full text] [doi: [10.1109/bibm52615.2021.9669404](https://doi.org/10.1109/bibm52615.2021.9669404)] [Medline: [35096472](https://pubmed.ncbi.nlm.nih.gov/35096472/)]
60. Zenone MA, Snyder J, Crooks VA. What are the informational pathways that shape people's use of cannabidiol for medical purposes? *J Cannabis Res* 2021 May 06;3(1):13 [FREE Full text] [doi: [10.1186/s42238-021-00069-x](https://doi.org/10.1186/s42238-021-00069-x)] [Medline: [33957993](https://pubmed.ncbi.nlm.nih.gov/33957993/)]
61. Allem JP, Majmundar A, Dormanesh A, Donaldson SI. Identifying health-related discussions of cannabis use on Twitter by using a medical dictionary: content analysis of tweets. *JMIR Form Res* 2022 Mar 25;6(2):e35027 [FREE Full text] [doi: [10.2196/35027](https://doi.org/10.2196/35027)] [Medline: [35212637](https://pubmed.ncbi.nlm.nih.gov/35212637/)]
62. Turner J, Kantardzic M, Vickers-Smith R. Infodemiological examination of personal and commercial tweets about cannabidiol: term and sentiment analysis. *J Med Internet Res* 2021 Dec 20;23(12):e27307 [FREE Full text] [doi: [10.2196/27307](https://doi.org/10.2196/27307)] [Medline: [34932014](https://pubmed.ncbi.nlm.nih.gov/34932014/)]
63. Meacham MC, Nobles AL, Tompkins DA, Thrul J. "I got a bunch of weed to help me through the withdrawals": naturalistic cannabis use reported in online opioid and opioid recovery community discussion forums. *PLoS One* 2022 Feb 8;17(2):e0263583 [FREE Full text] [doi: [10.1371/journal.pone.0263583](https://doi.org/10.1371/journal.pone.0263583)] [Medline: [35134074](https://pubmed.ncbi.nlm.nih.gov/35134074/)]
64. Boyd RL. WELP: Word Embedding Likeness Perusal (v1.03). Ryan Boyd. 2018. URL: <https://github.com/ryanboyd/WELP> [accessed 2022-10-30]
65. Adams N, Artigiani EE, Wish ED. Choosing your platform for social media drug research and improving your keyword filter list. *J Drug Issues* 2019 Mar 13;49(3):477-492. [doi: [10.1177/0022042619833911](https://doi.org/10.1177/0022042619833911)]
66. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003;3:993-1022 [FREE Full text] [doi: [10.1162/jmlr.2003.3.4-5.993](https://doi.org/10.1162/jmlr.2003.3.4-5.993)]
67. Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In: Aggarwal CC, Zhai C, editors. *Mining Text Data*. New York, NY, USA: Springer; 2012:415-463.
68. Hutto CJ, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*. 2014 Presented at: AAAI '14; June 1-4, 2014; Ann Arbor, MI, USA.
69. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. "When 'bad' is 'good'": identifying personal communication and sentiment in drug-related tweets. *JMIR Public Health Surveill* 2016 Oct 24;2(2):e162 [FREE Full text] [doi: [10.2196/publichealth.6327](https://doi.org/10.2196/publichealth.6327)] [Medline: [27777215](https://pubmed.ncbi.nlm.nih.gov/27777215/)]
70. Grady C. Institutional review boards: purpose and challenges. *Chest* 2015 Dec;148(5):1148-1155 [FREE Full text] [doi: [10.1378/chest.15-0706](https://doi.org/10.1378/chest.15-0706)] [Medline: [26042632](https://pubmed.ncbi.nlm.nih.gov/26042632/)]
71. CrowdTangle | Content Discovery and Social Monitoring Made Easy. URL: <https://www.crowdtangle.com/> [accessed 2022-05-30]
72. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. *Yearb Med Inform* 2017 Aug;26(1):214-227 [FREE Full text] [doi: [10.15265/IY-2017-029](https://doi.org/10.15265/IY-2017-029)] [Medline: [29063568](https://pubmed.ncbi.nlm.nih.gov/29063568/)]
73. Sarker A, DeRoos A, Perrone J. Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework. *J Am Med Inform Assoc* 2020 Feb 01;27(2):315-329 [FREE Full text] [doi: [10.1093/jamia/ocz162](https://doi.org/10.1093/jamia/ocz162)] [Medline: [31584645](https://pubmed.ncbi.nlm.nih.gov/31584645/)]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RQ: research question

VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by G Eysenbach; submitted 24.12.21; peer-reviewed by A Dormanesh, K O'Connor, J Thrul; comments to author 26.03.22; revised version received 16.06.22; accepted 27.07.22; published 16.11.22

Please cite as:

Khademi Habibabadi S, Hallinan C, Bonomo Y, Conway M

Consumer-Generated Discourse on Cannabis as a Medicine: Scoping Review of Techniques

J Med Internet Res 2022;24(11):e35974

URL: <https://www.jmir.org/2022/11/e35974>

doi: [10.2196/35974](https://doi.org/10.2196/35974)

PMID:

©Sedigheh Khademi Habibabadi, Christine Hallinan, Yvonne Bonomo, Mike Conway. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 16.11.2022. 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.