Full title: Social media discourse and internet search queries on cannabis as a medicine: A systematic scoping review

Short title: Social media discourse and search queries on cannabis as medicine

Christine Mary Hallinan[1,2,* #a,*], Sedigheh Khademi Habibabadi[1], , Mike Conway[3, #b] , and Yvonne Ann Bonomo[4,5, #c]

[1]Department of General Practice, Faculty of Medicine, Dentistry & Health Sciences, The University of Melbourne, Victoria, Australia

[2]Health & Biomedical Research Information Technology Unit (HaBIC R2), Department of General Practice, Melbourne Medical School, Faculty of Medicine, Dentistry & Health Sciences, The University of Melbourne, Victoria, Australia

[3]*Centre for Digital Transformation of Health, Victorian Comprehensive Cancer Centre, The University of Melbourne, Victoria, Australia*

[4]St Vincent's Health - Department of Addiction Medicine, Melbourne, Victoria, Australia

[5]St Vincent's Clinical School, Melbourne Medical School, Faculty of Medicine, Dentistry & Health Sciences, The University of Melbourne, Victoria, Australia

[#a] Current address: Department of General Practice, Melbourne Medical School, Faculty of Medicine, Dentistry & Health Sciences of the University of Melbourne, Level 3, 780 Elizabeth Street, Carlton, Victoria, 3053

[#c] Current address: St Vincent's Health - Department of Addiction Medicine 38 Fitzroy Street, Fitzroy, Victoria, 3065

[#b] Current address: Centre for Digital Transformation of Health, Victorian Comprehensive Cancer Centre, The University of Melbourne, Victoria, Australia

[*] Corresponding authors

Email: hallinan@unimelb.edu.au

Abstract

The use of cannabis for medicinal purposes has increased globally over the past decade since patient access to medicinal cannabis has been legislated. Yet, evidence of cannabis efficacy for a suite of conditions is only just emerging. Although there is considerable engagement from many stakeholders to add to the evidence base through randomized control trials, many gaps in the literature remain. Data from real-world and patient reported sources can provide opportunities to address this evidence deficit. This real-world data can be captured from a variety of sources such as found in routinely collected health care and health services records that include but are not limited to patient generated data from medical, administrative and claims data, patient reported data from surveys, wearable trackers, patient registries, and social media. In this systematic scoping review, we seek to understand the utility of online user generated text into the use of cannabis as a medicine. The objective of this scoping review is to synthesize primary research that uses social media discourse and internet search engine queries to answer the following questions: (i) Does online user-generated text provide a useful data source for studying cannabis as a medicine? (ii) What are the aims, data sources, and research themes of studies using online user-generated text to discuss the medicinal use of cannabis? For this scoping review we used a framework for systematic reviews and meta-analyses, the PRISMA guidelines to inform our methods. We conducted a manual search of primary research studies which used online user-generated text as a data source using the MEDLINE, Embase, Web of Science, and Scopus databases in October 2022. Editorials, letters, commentaries, surveys, protocols, and book chapters were excluded from the review. Forty-two studies were included in this review, 22 studies used manually labelled data, four studies used existing meta-data (Google trends/geo-location data), two studies used data that was manually coded using crowdsourcing services, and

two used automated coding supplied by a social media analytics company, 15 used computational methods for annotating data. Our review reflects a growing interest in the use of user-generated content for public health surveillance. It also demonstrates the need for the development of a systematic approach for evaluating the quality of social media studies and highlights the utility of automatic processing and computational methods (machine learning technologies) for large social media datasets. This systematic scoping review has shown that user-generated content as a data source for studying cannabis as a medicine provides another means to understand how cannabis is perceived and used in the community. As such, it provides another potential 'tool' with which to engage in pharmacovigilance of, not only cannabis as a medicine, but also other novel therapeutics as they enter the market.

Introduction

Genetic analysis of ancients cannabis indicates the plant cannabis sativa was first cultivated for use as a as a medicinal agent up to 2400 years ago (1). From the 1800's, people in the United States (US), widely used cannabis as a medicine by either prescription or as an over the counter therapeutic (2). Yet by the mid-20th century, cannabis use was prohibited in many parts of the developed world with the passing of legislation in the US, the United Kingdom (UK) and various European countries that proscribed its use (3-6). Since the 2000s, the use of cannabis for medicinal purposes has been decriminalized in many countries including Israel, Canada, Netherlands, Uniter States, United Kingdom, and Australia (7-9). More recently, new evidence regarding the clinical efficacy of cannabis for some conditions (10) has triggered public interest in cannabis and cannabis-derived products (11, 12), resulting in a global trend towards public acceptance, and subsequent legalization of cannabis for both medicinal and recreational use.

There is emerging evidence of cannabis efficacy for childhood epilepsy, spasticity, and neuropathic pain in multiple sclerosis, acquired immunodeficiency syndrome (AIDS) wasting syndrome, and cancer chemotherapy-induced nausea and vomiting (13-15). Although researchers are investigating cannabis for treating cancer, psychiatric disorders (16), sleep disorders (17), chronic pain (18) and inflammatory conditions such as rheumatoid arthritis (19), there is currently insufficient evidence to support its clinical use. Scientific studies on emerging therapeutics typically exclude vulnerable populations such as pregnant women, young people, the elderly, patients with multimorbidity and polypharmacy, and this limits the availability of evidence for cannabis effectiveness across these population groups (20).

Cannabis as medicine is associated with a rapidly expanding industry (21). Patient demand is increasing, as is reflected in an increasing number of approvals for prescriptions over time (22) , with one study showing that 61% of Australian GPs surveyed reported one or more patient enquiries regarding medical cannabis (23). With this increasing demand, is sophisticated marketing by medicinal cannabis companies that leverages evidence from a small number of studies to promote their products (24, 25). In light of this, concerns regarding patient safety is warranted especially when marketing for some cannabinoid products is associated with inadequate labelling and/or inappropriate dosage recommendations (26). These concerns are compounded by the downscheduling of over-the-counter cannabis products which do not require a prescription (27) and the illicit drug market (28). Given this dynamic interplay between marketing, product innovation, regulation, and consumer demand, innovative methods are required to augment existing established approaches to the surveillance and monitoring of emerging and unapproved drugs.

Although there is considerable engagement from many stakeholders to improve the scientific evidence regarding the efficacy and safety of cannabis through randomized control trials, many gaps remain in the literature (29). Yet data from real-world and patient reported data sources could provide opportunities to address this evidence deficit (30). This real-world data can be captured from a variety of sources such as found in routinely collected health care and health services records that include but are not limited to patient generated data from medical, administrative and claims data, as well as patient reported data from surveys, wearable trackers, patient registries, and social media (31-33).

People readily consult the internet when looking for and sharing health information (34, 35). According to 2017 survey of Health Information National Trends, almost 78% of US adults used online searches first to inquire about health or medical information (34). Data resulting from these online activities is labelled 'user generated' and is increasingly becoming a component of surveillance systems in the health data domain (36). Monitoring user-generated data on the web can be a timely and inexpensive way to generated population-level insights (37). The collective experiences and opinions shared online are an easily accessible wide-ranging data source for tracking emerging trends – which might be unavailable or less noticeable by other surveillance systems.

The objective of this systematic scoping review is to understand the utility of online user generated text in providing insight into the use of cannabis as a medicine. In this review, we aim to systematically review existing work that utilizes user-generated content to explore cannabis as a medicine. The objective of this review is to synthesise primary research that uses social media discourse and internet search engine queries to answer the following questions:

- Does online user-generated text provide a useful data source for studying cannabis as a medicine?
- What are the aims, data sources, and research themes of studies using online user-generated text to discuss the medicinal use of cannabis?

Materials and methods

Search strategy

For this scoping review we used the PRISMA statement for systematic reviews and meta-analyses, to inform our systematic search methods (38).

Literature database queries were developed for four categories of studies. The first three categories used social media text as a data source, the fourth relied on internet search engine query data. For the first category, the database queries combined words used to describe social media forums, and cannabis-related keywords and general medical-related keywords (Table 1 Category 1). The second category also included the social media and cannabis-related keywords, but used keywords specific to psychiatric disorders, for which the use of medical cannabis has been described. Our search terms for this second category were informed by a systematic review of medicinal cannabis for psychiatric disorders (16) (Table 1 Category 2). The third category included social media and cannabis related keywords but focused on non-psychiatric medical conditions for which cannabis is sometimes used (Table 1 Category 3). The fourth category included studies using Internet search engine queries as a data source, there were no medical conditions included in these searches (Table 1 Category 4). A manual search of MEDLINE (Web of Science (1900-2022), Embase (OVID: 1974-2022), Web of Science Core Collection (1900-

112 2022), and Scopus (1996-2022) databases was conducted by SKH and CMH in May 2021 and
113 again in October 2022. The search was limited to English-language studies that were published
114 between January 1974 and April 2022 (Appendix I).

115 **Table 1. Search Strategy**

CATEGORY 1 - SOCIAL MEDIA, CANNABIS, AND MEDICAL TERMS AS KEYWORDS		
Social media related Keywords	*Cannabis keywords*	*Medical keywords*
"Social media" OR twitter OR reddit OR instagram OR youtube OR pinterest OR facebook OR "social network forum" OR "Online health community" OR "message board"	cannabis OR cannabis OR cannabinoids OR delta-9-tetrahydrocannabinol OR cannabidiol OR cbd OR cbg OR cbn OR thc OR weed	medical OR medicinal OR patient OR patients OR medicine OR doctor OR position OR care OR therapy OR therapeutic
CATEGORY 2 - SOCIAL MEDIA, CANNABIS, AND PSYCHIATRIC DISORDERS KEYWORDS		
Social media related Keywords	*Cannabis keywords*	*Psychiatric disorders*
"Social media" OR twitter OR reddit OR instagram OR youtube OR pinterest OR facebook OR "social network forum" OR "Online health community" OR "message board"	cannabis OR cannabis OR cannabinoids OR delta-9-tetrahydrocannabinol OR cannabidiol OR cbd OR cbg OR cbn OR thc OR weed	depression OR depressive OR "mental illness*" OR "mental disorder*" OR "mental health" OR "mood disorder*" OR "affective disorder*" OR anxi* OR "panic disorder" OR "obsessive compulsive" OR adhd OR "attention deficit" OR phobi* OR bipolar OR psychiat* OR psychological OR psychosis OR psychotic OR schizophr* OR "severe mental*" OR "serious mental*" OR antidepress* OR antipsychotic* OR "post traumatic*" OR "personality disorder*" OR stress
CATEGORY 3 - SOCIAL MEDIA, CANNABIS AND VARIOUS MEDICAL (NON-PSYCHIATRIC) CONDITIONS/ ILLNESSES KEYWORDS		
Social media related Keywords	*Cannabis keywords*	*Medical conditions*
"Social media" OR twitter OR reddit OR instagram OR youtube OR pinterest OR facebook OR "social network forum" OR "Online health community" OR "message board"	cannabis OR cannabis OR cannabinoids OR delta-9-tetrahydrocannabinol OR cannabidiol OR cbd OR cbg OR cbn OR thc OR weed	Pain, Opioid, Alzheimer, sleep OR insomnia, inflammatory, arthritis, Multiple Sclerosis, Endometriosis
CATEGORY 4 - SEARCH ENGINE QUERIES AND CANNABIS KEYWORDS		
Search Engine keywords	*Cannabis keywords*	
"Search engine" OR "search log" OR "search queries" OR "online search" OR "internet Search" OR "web search"	cannabis OR Cannabis OR Cannabinoids OR Delta-9-Tetrahydrocannabinol OR Cannabidiol OR CBD OR CBG OR CBN OR THC OR weed	

116

117 The inclusion criteria for this review were: (i) peer reviewed research studies, (ii) peer reviewed

118 conference papers (iii) studies which used online user-generated text as a data source, and (iv)

social media research that was either directly focused on cannabis and cannabis products that have an impact on health or were health-related studies that found medicinal use of cannabis.

Exclusion criteria comprised: (i) editorials, letters, commentaries, surveys, protocols and book chapters; (ii) studies that used social media for recruiting participants; (iii) studies where the full text of the publication was not available; (iv) conference abstracts (iv) studies primarily focused on electronic nicotine delivery systems adapted to deliver cannabinoids; (v) studies that used bots or autonomous systems as the main data source and (vi) studies that focused exclusively on synthetic cannabis.

All studies captured by the search queries listed in Table 1 were uploaded into excel to enable all duplicates to be removed. Following this all titles and abstracts were reviewed independently and in duplicate by CMH and SKH. Records were excluded based on title and abstract screening as well as publication type.

The full text articles that were identified for inclusion following screening process were then independently critiqued by pairs of reviewers using a checklist developed for this study. The purpose of the checklist that we developed for this systematic scoping review was to provide an overall assessment of quality rather than generate a specific score, (Appendix 1). Assessments of quality in each study were based on evidence of relative quality in the aims or objectives, main findings, data collection method, analytic methods, data source, and evaluation and interpretations of the study. CMH and SKH critiqued all articles, and YB and MC each critiqued a selection of studies to ensure each article had been independently reviewed by two researchers. Where initial disagreement existed between reviewers regarding the inclusion of a study, team members met to discuss the disputed article's status until consensus was achieved.

Study inclusion

Assessments of quality in each study were based on evaluating each study's aims and objectives, main findings, data collection and analytic methods, data sources, and evaluation and interpretations of the results. Social media studies were included if there were no major biases affecting the internal, external or construct validity of the study (39). In doing so, the internal validity of each study was determined by the quality of the data and analytic processes used, the external validity determined by the extent to which the findings can be generalized to other contexts, and the construct validity was ascertained by the extent to which the chosen measurement tool correctly measured what the study aimed to measure (Appendix II).

Of the 1,556 titles identified in the electronic database searches, 858 duplicate articles were removed, 449 were excluded following the screening of title and abstracts and 197 were excluded based on publication type (i.e., survey, letter, comment, abstract). This screening process provided 52 potentially relevant full text primary research studies to be included in the review. Of these, five articles were not able to be retrieved, two out of 47 articles had initial disagreement. Upon consensus, five were excluded with reasons (Appendix III) using the quality assessment checklist as described above. This provided 42 papers for inclusion in this systematic scoping review.

The PRISMA flow diagram is presented in Fig 1.

Results

Of the 42 papers published between 2014 and 2022 included in this systematic scoping review, most were journal articles 40/42 (95.2%), and two were conference-based publications 2/42 (4.8%). Although the first study was published eight years ago, nearly two-thirds 24/42 (57%),

have been published over the last four years. Table 2 provides a summary of each paper that includes author names, publication year, data source, duration of the study, number of collected posts, number of analyzed posts, and the coding/labelling approach used.

Data collection and annotation

The largest manually annotated dataset that contained 47,000 labelled tweets was published by (40) in 2015. This paper was one of 22 studies included in this review (54.8%) that either collected a limited number of data points, or sampled their collected data, and manually coded the data to gain an in-depth understanding of the domain (41-62). Four of the 42 studies (9.5%) used existing meta-data including Google trends summary data (63-65), and geo-location data (66). Two (4.8%) studies used data that was manually coded using crowdsourcing services (42, 46), and two (4.8%) used automated coding supplied by a social media analytics company (67, 68),. Fifteen of the 42 studies (35.7%) used automated methods for labelling data, that included the use of machine learning, lexicon, and rule-based algorithms (57, 69-82). Automated coding was increasingly used as an analytic tool for social media data on this topic from 2017 onward (Table 2).

183 **Table 2. Articles included in the review**

#	STUDY	SOURCE	DURATION	COLLECTED DATA	ANNOTATED/ ANALYSED	CODING/ LABELLING APPROACH
1	McGregor et al., 2014	Online forums, Facebook, Twitter, YouTube, Patient Opinion	Not Available	3,785 items	All data	Manual coding
2	Cavazos-Rehg, Sowles & Fisher et al., 2015	Twitter	February 2014—Mar 2014	7,653,738 tweets	7,000 tweets	Manual coding using crowdsourcing services
3	Daniulaityte et al., 2015	Twitter	October 2014—December 2014	125,255 tweets 27,018 geolocated	All data	No coding—Use existing geographical fields
4	Gonzalez-Estrada et al., 2015	YouTube	4-8 June 2014	200 most viewed videos	All data	Manual coding
5	Krauss et al., 2015	YouTube	22 January 2015	116 Videos	All data	Manual coding
6	Thompson, Rivara & Whitehill, 2015	Twitter	March 2012—July 2013	36,939 original tweets 10,000 retweets	~47,000 tweets	Manual coding
7	Cavazos-Rehg, Sowles & Krauss et al., 2016	Twitter	January 2015	206,854 tweets	5,000 tweets	Manual coding using crowdsourcing services
8	Lamy et al., 2016	Twitter	May 2015—July 2015	100,182 tweets 26,975 geolocated	3,000 tweets	Manual coding
9	Mitchell et al., 2016	Online Forums	October 2014	268 forum threads	46 threads 880 posts	Manual coding
10	Andersson, Persson & Kjellgren, 2017	Online Forums	18-19 April 2016	32 topics	All data	Manual coding
11	Dai & Hao, 2017	Twitter	August 2015—April 2016	1,253,872 post traumatic stress disorder (PTSD) tweets 66,000 cannabis PTSD tweets	2,000 labelled tweets, remaining tweets by machine learning	Automated coding using machine learning
12	Greiner, Chatton & Khazaal, 2017	Online Forums	November 2014—March 2015	717 posts	All data	Manual coding
13	Turner & Kantardzic, 2017	Twitter	August 2015—April 2016	40,509 geolocated tweets	2,000 labelled, remaining tweets by machine learning	Automated coding using machine learning
14	Westmaas, McDonald & Portier, 2017	Online Forums	January 2000—December 2013	468,000 posts	All data	Automated coding using topic modelling

15	Yom-Tov & Lev-Ran, 2017	Bing search engine	November 2016—April 2017	Not available	All data	Automated coding using lexicons
16	Cavazos-Rehg, Krauss & Sowles et al., 2018	YouTube	10-11 June 2015	83 videos	All data	Manual coding
17	Glowacki, Glowacki, & Wilcox., 2018	Twitter	August 2016—October 2016	73,235 tweets	All data	Automated coding using topic modelling
18	Meacham, Paul & Ramo, 2018	Reddit	January 2010–December 2016	~400,000 posts	All data	Automated coding using lexicons and pattern matching
19	Leas, Nobels & Caputi et al., 2019	Google Trends	January 2004—April 2019	Not available	Summary data	No coding – used Google trends data
20	Meacham, Roh & Chang et al., 2019	Reddit	January 2017—December 2017	193 dabbing questions and/or posts	All data	Manual coding
21	Nasralah, El-Gayar & Wang, 2019	Twitter	January 2015—February 2019	20,609 tweets	All data	Automated coding supplied by an analytics company
22	Pérez-Pérez et al., 2019	Twitter	February 2018—August 2018	24,634 tweets	All data	Automated coding using lexicons
23	Shi et al., 2019	Google Trends; Buzzsumo	January 2011—July 2018	Not available	Summary data	No coding – used Google trends data
24	Allem, Escobedo & Dharmapuri, 2020	Twitter	May 2018—December 2018	19,081,081 tweets	60,861 non-bots 8,874 bots	Automated coding using rule-based methods
25	Janmohamed et al., 2020	Blogs, news, forums , comments, professional reviews, Facebook	August 2019—April 2021	4,027,172 documents and/or blog entries	All data	Automated coding using topic modelling
26	Jia et al., 2020	Google, Facebook, YouTube	September 2019	51 Google websites 126 Facebook posts 37 YouTube videos	All data	Manual coding
27	Leas, Hendrickson & Nobels et al., 2020	Reddit	January 2014—August 2019	104,917 reddit posts	3,000 initial data 376 as testimonials	Manual coding
28	Merten, Gordon & King et al., 2020	Pinterest	January 2018—December 2018	1,280 pins	226 pins	Manual coding
29	Mullins et al., 2020	Twitter	June 2017—July 2017	941 tweets	All data	Automated coding supplied by an analytics company
30	Saposnik & Huber, 2020	Google Trends	January 2004—December 2019	Not available	Summary data	No coding – used Google trends data

31	Song et al., 2020	GoFundMe	January 2012—December 2019	1,474 campaigns	500 campaigns	Manual coding
32	Tran & Kavuluru, 2020	Reddit and FDA comments	January 2019—April 2019	64,099 reddit comments 3,832 FDA (U.S. Food & Drug Administration) comments	All data	Automated coding using rule-based methods
33	van Draanen et al., 2020	Twitter	January 2017—June 2019	1,200,127 tweets	All data	Automated coding using Topic modelling
34	Zenone, Snyder & Caul, 2020	GoFundMe	January 2017—May 2019	155 campaigns	All data	Manual coding
35	Pang et al., 2021	Twitter	December 2019—December 2020	17,238 tweets	1,000 tweets	Manual coding
36	Rhidenour et al., 2021	Reddit	January 2008—December 2018	974 posts	All data	Manual coding
37	Smolev et al., 2021	Facebook	November 2018—November 2019	7,694 posts	All data	Manual coding
38	Zenone, Snyder & Crooks, 2021	GoFundMe	June 2017—May 2019	164 campaigns	All data	Manual coding
39	Soleymanpour, Saderholm & Kavuluru, 2021	Twitter	July 2019	2,200,000 tweets	2,000 labelled, remaining data by machine learning	Automated coding using machine learning
40	Allem et al. 2022	Twitter	January 2020—September 2020	353,353 tweets	1,092 labelled, remaining data by classifiers	Automated coding using rule-based methods
41	Meacham, Nobles, Tompkins & Thrul, 2022	Reddit (2 subreddits)	December 2015—August 2019	16,791 opioid recovery subreddits 159,994 opioid use subreddits	200 posts labelled manually and analyzed 908 opioid recovery subreddits 4,224 opioid use subreddits	Manual coding and automated coding using rule-based methods
42	Turner, Kantardzic, & Vickers-Smith, 2022	Twitter		567,850 tweets collected –	5,496 tweets manually labelled; 167,755 personal use tweets 143,322 commercial tweets labelled by classifier	Automated coding using machine learning

Data analysis

The manually labelled studies calculated proportions and trends and developed themes (40, 41, 43, 44, 46, 47, 51-53, 57, 58, 61, 62, 79, 83-91). (63-65) reported on Google trends data that delivered an index of Google search trends over time. (66) processed Twitter data that contained existing geographical fields to identify geolocation at source. The studies that utilised a large volume of data used advanced computational methods, which included sentiment analysis, topic modeling, and rule-based text mining (57, 69, 74-79, 82). The use of sentiment analysis in the (57, 77, 92) enabled the analysis of people's sentiments, opinions, and attitudes. Topic modelling in the (71, 73, 76, 77, 82) studies enabled the development of themes via automatic machine learning methods. The use of rule-based text mining such as found in the (57, 69, 80, 93, 94) studies enabled the classification of posts into pre-existing health-related categories.

Research Themes

In this review, we categorized the 42 research articles into six broad themes. Themes were based on the research questions motivating the studies, where each paper was classified as belonging a primary theme, based on alignment with the research aims.

General cannabis-related

Nine studies were included in this theme (Table 3). The main keywords used in these studies included general terms such as 'cannabis', 'marijuana', 'pot' and 'weed'. The major aim of these studies was to either identify topics of conversations regarding cannabis, or to examine their sentiments. These studies are included because they reported on conversations around cannabis use for medical purposes, sentiment associated with perceptions of health benefits of cannabis,

209 and reports of adverse effects. For example, a study on veterans use of cannabis found that
210 cannabis is used to self-medicate a number of health issues, including Post-Traumatic Stress
211 Disorder (PTSD), anxiety and sleep disorder (91). Seven of the studies used Twitter as a data
212 source (40, 70, 75, 84, 93, 95, 96) , one examined the content of YouTube videos about cannabis
213 (97), one investigated online self-help forums (85) and another used Reddit data (91).

214 **Table 3. Papers studying general conversations**

STUDY	AIM	HEALTH-RELATED EFFECTS/CLAIMS	DATA IDENTIFICATION
Cavazos-Rehg, Krauss et al., 2015	To examine the sentiment and themes of cannabis-related tweets from influential users and to describe the users' demographics.	A common theme of pro tweets was that cannabis has health benefits. Anti-cannabis posts spoke of the harm experienced in using cannabis. 77% of posts had positive sentiments, with 12 times higher reach than other posts.	Cannabis-related keyword
Thompson, Rivara et al., 2015	To examine cannabis- related content in Twitter, especially content tweeted by adolescent users, and to examine any differences in message content before and after the legalization of recreational cannabis in two US states.	More tweets described perceived positive benefits of cannabis use, including relaxation and escaping life problems. Tweets described cannabis as less harmful than other drugs or as not harmful at all and suggested its medical role for conditions such as depression and cancer. Less than 1% of tweets expressed a concern about cannabis use.	Cannabis-related keyword
Greiner, Chatton et al., 2017	To investigate online content of cannabis use/addiction self-help forums.	Self-help forums on cannabis share a theme around cannabis users seeking help for addiction and withdrawal issues.	keywords "cannabis" and "forum" and "help"
Turner et al., 2017	To examine if cannabis legalization policies impact Twitter conversations and the social networks of users contributing to cannabis conversations.	Medical cannabis was a major topic in the conversations.	Cannabis-related keywords
Cavazos-Rehg, Krauss et al., 2018	To investigate cannabis product reviews and the relationship between exposure to product reviews and cannabis users' demographics and characteristics.	Product reviews promoted cannabis for helping with relaxation, pain relief, sleep, improving emotional well-being. Medical cannabis users are more likely to be exposed to cannabis product reviews.	"weed review", "marijuana review", "cannabis review"

Allem, Escobedo et al., 2020	To identify and describe cannabis-related topics of conversation on Twitter, and the public health implications of these.	Health and medical was the third most prevalent topic of the 12 topics identified in the data. Posts suggested that cannabis could help with cancer, sleep, pain, anxiety, depression, trauma, and posttraumatic stress disorder. Health-related posts from social bots were almost double of genuine posts.	Cannabis-related keywords
Van Draanen et al., 2020	To examine differences in the sentiment and content of cannabis-related tweets in the US (by state cannabis laws) and Canada.	Medical cannabis use was one of the main topics of conversations in cannabis-related tweets from both countries.	Tweets filtered on US and Canada geolocation and then further filtered on cannabis-related keywords
Rhidenour et al., 2021	To explore Veterans' Reddit discussions regarding their cannabis use.	Over a third of the Reddit posts described the use of medical cannabis as an aid for psychological and physical ailments. Overall, veterans discussed how the use of medical cannabis reduced PTSD symptoms, anxiety, and helped with their sleep.	The veteran subreddit
Allem, Majmundar et al., 2022	To determine the extent to which a medical dictionary could identify cannabis-related motivations for use and health consequences of cannabis use.	There were posts related to both health motivations and consequences of cannabis use. The health-related posts included issues with the respiratory system, stress to the immune system, and gastrointestinal issues, among others.	Cannabis-related keywords

Cannabis mode of use

Seven studies reported on the mode of use of cannabis as a medicine (Table 4). These studies collected data using keywords such as 'vape', 'vaping', 'dabbing', and 'edibles'. Conversations around modes of use revealed a theme about lacking, seeking, or sharing knowledge about possible health consequences of the modes of use. Another theme was around the perceived health benefits of cannabis and the various modes of use of cannabinoids that included sleep improvement and relaxation resulting from dabbing oils (46) or consuming 'edibles' (47). The

224 findings suggest that for emerging modes of use such as dabbing, where the availability of

225 evidence-based information is limited, people seek information from others' experiences.

226 **Table 4. Papers studying Cannabis mode of use**

STUDY	AIM	HEALTH-RELATED EFFECTS/CLAIMS	DATA IDENTIFICATION
Daniulaityte et al., 2015	To explore Twitter data on concentrate ('dabs') use and examine the impact of cannabis legalization policies on concentrate use conversations.	Twitter data suggest popularity of dabs in the US states with legalized recreational/medical use of cannabis. Dabbing as an emerging mode of use could carry significant health risks.	dab-related keywords for U.S. location
Krauss, Sowles et al., 2015	To explore the content of cannabis dabbing-related videos on YouTube.	Only 21% of videos contained some warning about dabbing, such as preventing explosions, injury, or negative side effects. 22% of videos specifically mentioned medical cannabis or getting "medicated", either in the video itself or in the accompanying text description.	Dabbing related keywords
Cavazos-Rehg, Sowles et al., 2016	To study themes of dabbing conversations and to investigate the consequences of high-potency cannabis consumption.	4th theme (of 7 themes) was about cannabis helping with relaxation, sleep or solving problems. Extreme effects were both physiological and psychological. The most common physiologic effects were passing out and respiratory, with coughing the most common respiratory effect.	Dabbing related keywords
Lamy, Daniulaite et al., 2016	To study themes of edibles conversations and examine legalization policies' impact on cannabis-related tweeting activity.	Twitter data suggest mostly positive attitudes toward cannabis edibles. Positive tweets describe the quality of the "high" experienced and how cannabis edibles facilitate falling asleep. Negative tweets discuss the unreliability of edibles' THC dosage and delayed effects that were linked to over-consumption, which could lead to potential harmful consequences.	Cannabis edible-related keywords
Meacham, Paul et al., 2018	To analyse discussions of emerging and traditional forms of cannabis use.	Less than 2% of conversations described adverse effects. The most mentioned adverse effects were anxiety-related in the context of smoking, edibles, and butane hash oil, and cough for vaping and dabbing.	A cannabis specific subreddit on various modes of use
Meacham, Roh et al., 2019	To study themes of dabbing- related questions and responses.	Health concerns are the 5th category of dabbing questions - including respiratory effects, anxiety, and vomiting.	Search for "Dab" and "question"

		Respondents in these conversations usually spoke from personal experience.	on cannabis subreddits
Janmohamed et al.,2020	To map temporal trends in the web-based vaping narrative, to indicate how the narrative changed from before to during the COVID-19 pandemic.	The emergence of a vape-administered CBD treatment narrative around the COVID-19 pandemic.	Vape-related keywords

227

228

229 ## Cannabis as a medicine for a specific health issue

230 Six studies were included in this theme (Table 5). These studies investigated conversations
231 around the use of cannabis or cannabidiol for a specific health issue. The health conditions
232 included glaucoma (53), PTSD (69), cancer (62, 98), Attention Deficit Hyperactivity Disorder
233 (ADHD) (88) and pregnancy (89). These studies mostly discovered that conversations claimed
234 benefits of cannabis as an alternative treatment for these health conditions, although mentions of
235 harm, and both harm and therapeutic effects, were also present (88).

236 **Table 5. Papers studying Cannabis as a medicine for a specific health issue**

STUDY	AIM	HEALTH-RELATED EFFECTS/CLAIMS	DATA IDENTIFICATION
Mitchell et al., 2016	To examine the content of online forum threads on ADHD and cannabis use to identify trends about their relation, particularly regarding therapeutic and adverse effects of cannabis on ADHD.	25% of individual posts indicated that cannabis is therapeutic for ADHD, as opposed to 8% that claim it is harmful, 5% that it is both therapeutic and harmful, and 2% that it has no effect on ADHD.	Cannabis and ADHD keywords
Dai & Hao, 2017	To evaluate factors that could impact public attitudes to PTSD related cannabis use.	5.3% of all PTSD tweets were related to cannabis use and these tweets predominantly supported cannabis use for PTSD.	Cannabis and PTSD keywords

Shi et al., 2019	To characterize trends in use of cannabis for cancer and analysis of content and impact of popular news about cannabis for cancer.	Between 2011-2018, the relative google search volume of 'cannabis cancer' queries increased at a rate 10 times faster than 'standard cancer therapies' queries. Popular 'false news' stories had a much higher engagement than contrary 'accurate' news stories.	Cannabis vs standard therapies for cancer
Jia, Mehran et al., 2020	To analyze the content quality and risk of readily available online information regarding cannabis and glaucoma.	Although the American Glaucoma Society recommends against cannabis use for glaucoma treatment, 21% of Facebook, 24% of Google, and 59% of YouTube search results were pro cannabis use for glaucoma treatment.	Cannabis and Glaucoma keywords
Zenone et al., 2020	To use crowdfunding campaigns to understand how cannabidiol is represented/misrepresented as a cancer-related care.	CBD use was reported to reduce the side effects of conventional treatments or can be used with other complementary cancer treatments. Reported uses included stimulating appetite, general pain relief, assisting with sleep, countering nausea, or general recovery purposes. Most campaigners presented definite efficacy of CBD for pain or symptom management.	CBD term variants and "cancer"
Pang et al., 2021	To examine cannabis and pregnancy-related tweets over a 12-month period.	36% mentioned safety during pregnancy, 2.3% of posts asked about safety during postpartum, and 2.7% of posts expressed use of cannabis during pregnancy to help with pregnancy symptom i.e. to help with morning sickness, nausea, vomiting, headaches, pain, stress, and fatigue. The authors conclude that health providers discuss risks and provide official information about cannabis use in pregnancy.	Cannabis and pregnancy related keywords

Cannabis as a medicine as part of discourse on illness and disease

Eleven studies were included in this theme. In this category, the research focus was on social media topics relating to management and treatment options for a range of health conditions rather than on medicinal cannabis per se (Table 6). Health conditions included inflammatory and irritable bowel disease(99), opioid use disorder (67, 100, 101), pain (68), ophthalmic disease

245 (41), cluster headache and migraine(83), asthma (102), cancer (90, 103), autism disorder (104)

246 and brachial plexus injury (61).

247 **Table 6. Papers studying Cannabis as a medicine as part of discourse on illness and disease**

STUDY	AIM	HEALTH-RELATED EFFECTS/CLAIMS	DATA IDENTIFICATION
McGregor et al., 2014	To analyse the ophthalmic content of social media platforms.	Treatment was one of the main themes, with complementary therapy featuring most prominently on Twitter, where 87% of posts on complementary therapy described the use of medical cannabis for glaucoma.	Glaucoma patient forums and glaucoma keywords
Gonzalez-Estrada et al., 2015	To determine the educational quality of YouTube videos for asthma.	The most common video content was regarding alternative medicine (38%) and included cannabis as well as live fish ingestion; salt inhalers; raw food, etc.	Asthma related videos
Andersson et al., 2017	To understand the use of non-established or alternative pharmacological treatments used to alleviate cluster headaches and migraines.	Cannabis was discussed for its potential to alleviate symptoms or reduce the frequency of migraine attacks. Some had used cannabis for other purposes but experienced additional benefits for headaches. The effects of self-treatment with cannabis appeared more contradictory and complex than treatment with other substances.	Search for "treatment migraines" on three alternative treatment online forums
Westmaas et al., 2017	To investigate contexts in which smoking, or quitting is discussed in a cancer survivor network.	Use of cannabis (primarily for nausea), was the 4th topic.	Smoking/cessation-related keywords from the Cancer Survivors Network
Glowacki et al.,2018	To identify public reactions to the opioid epidemic by identifying the most popular topics.	Mentions of cannabis as an effective alternate to opioids for managing pain.	Opioid-related keywords
Nasralah et al., 2019	To understand the concerns of opioid-addicted users.	Cannabis was found in two of five main themes: "In recovery" and "taking illicit drugs" for pain management.	Users who self-identified as addicted to, or previously addicted to opioids

Pérez-Pérez et al., 2019	To characterize the bowel disease community on Twitter.	Medical cannabis was the 4th most mentioned term in the bowel disease (BD) community. Medical cannabis and its components were the most discussed drug, with mentions of its benefits in mitigating common BD symptoms.	Inflammatory bowel disease, Irritable bowel disease keywords
Mullins et al., 2020	To examine pain-related tweets in Ireland over a 2-week period.	The fourth most occurring keyword was cannabis. Ninety percent of cannabis related tweets were non-personal, with highly positive sentiment and highest number of impressions per tweet. Cannabis had by the largest number of tweets aimed at generating awareness.	Pain-related keywords
Saposnik & Huber, 2020	To analyse of trends in web searches for the cause and treatments of autism spectrum disorder (ASD).	ASD and cannabis web searches have continued to rise since 2009. Apart from searches on Applied Behavioral Analysis and Autism, cannabis and ASD have been searched more than other ASD interventions since 2013.	"Autism" and key search terms for causes and treatments of autism
Song et al., 2020	To understand the cancer patient's perspective for using complementary and alternative medicine, or for declining traditional cancer therapy.	Cannabidiol oil was 10th amongst the most used alternative treatments.	20 most prevalent cancers in the U.S. and a list of most frequently utilized complementary and alternative medicine including yoga, herbal, etc.
Smolev et al., 2021	To analyse themes of brachial plexus injury Facebook conversations.	There were 313 posts regarding cannabinoids as a preferred alternative pain management medication.	"traumatic brachial plexus injury" keyword
Meacham, Nobles et al. 2022	To assess and compare active opioid use subreddit and an opioid recovery subreddit: 1) the proportion of posts that mention cannabis, 2) the most common words and phrases in posts that mention cannabis, and 3) motivations for cannabis use in relation to opioid use as described in cannabis-related posts.	Potential use of cannabis for opioid use disorder and management of withdrawal symptoms.	Opioid use and recovery subreddits

Cannabidiol (CBD)

There were six studies in the cannabidiol (CBD) category (58, 77, 79, 86, 87, 105, 106) (Table 7). These studies concentrated on conversations related to the benefits of CBD products, product sentiment (positive, negative, or neutral), the factors that impact on a person's decision to use CBD products, and the trends in therapeutic use of CBD.

Table 7. Papers studying Cannabidiol (CBD)

STUDY	AIM	HEALTH-RELATED EFFECTS/CLAIMS	DATA IDENTIFICATION
Merten, Gordon et al., 2020	To analyse how CBD is portrayed on Pinterest.	Most pins (57.5%) did not make a specific health benefit claim yet 42.5% claimed mental, physical, or both mental and physical health benefits.	'cannabidiol' or 'CBD'
Zenone, Snyder et al., 2021	To analyse the CBD informational pathways which bring consumers to CBD for medical purposes.	Self-directed research was the most common pathway to CBD. The proposed uses of CBD were for cancer, seizure-inducing diseases/conditions, joint/inflammatory diseases, mental health disorders, nervous system diseases, and autoimmune diseases.	'cannabidiol' or 'CBD'
Leas, Nobel et al., 2019	To analyse public interest trends in CBD using Google Trends.	Searches for CBD exceed searches for yoga and around half as much as searches for dieting.	'cannabidiol' or 'CBD' vs other alternative medicine including diet, yoga, etc.
Leas, Hendrickson et al., 2020	To assess if individuals are using CBD for diagnosable conditions which have evidence-based therapies.	Psychiatric conditions were the most cited diagnosable condition, mentioned in 63.9% of testimonials. The second most cited subcategory was orthopedic conditions (26.4%), followed by sleep (14.6%), neurological (6.9%), and gastroenterological (3.9%) conditions.	CBD subreddit posts
Tran & Kavuluru, 2020	To examine social media data to determine perceived remedial effects and usage patterns for CBD.	Anxiety disorders and pain were the two conditions dominating the discussion surrounding CBD, both in terms of general discussion and for CBD as a perceived therapeutic treatment. CBD is mentioned as a treatment for mental issues (anxiety, depression, stress) and physiological issues (pain, inflammation, and others)	CBD subreddit posts

Soleymanpour et al., 2021	To perform content analysis of marketing claims for CBD in Twitter.	Over 50% of CBD tweets appear to be marketing related chatter. Pain and anxiety are the most popular conditions mentioned in marketing messages. Edibles are the most popular product type being advertised, followed by oils.	"cbd", "cbdoil" and "cannabidiol"
Turner, Kantardzic et al. 2021	The objective of this study was to provide a framework for public health and medical researchers to use for identifying and analyzing the consumption and marketing of unregulated substances using CBD as an exemplar.	There was a significant difference in the sentiment scores between the personal and commercial CBD tweets, the mean sentiment score of the commercial CBD scores was higher than that of the personal CBD score. Pain, anxiety, and sleep had the highest positive sentiment score for both personal and commercial CBD tweets.	"CBD", "cannabidiol."

Adverse drug reactions and adverse effects

One paper had a research question that explicitly focused on the detection of adverse events (72) (Table 8) This study explored the prevalence of internet search engine queries relating to the topic of adverse reactions and cannabis use. Seven other studies contained mentions of adverse effects which were associated with cannabis use (44, 46, 48, 50, 52, 59, 74, 78), however these papers were not included under this theme, as their research questions were not centred around the explicit investigation of adverse events.

Table 8. Papers studying adverse events related to cannabis

STUDY	AIM	HEALTH-RELATED EFFECTS/CLAIMS	DATA IDENTIFICATION
Yom-Tov & Lev-Ran, 2017	To check if search engine queries can be used to detect adverse reactions of cannabis use.	A high correlation between the side effects recorded on established reporting systems and those found in the search engine queries. These side effects included anxiety, depression-related symptoms, psychotic symptoms such as paranoia and hallucinations, cough and other symptoms.	Cannabis related keywords

Discussion

Currently, there exist systematic reviews of cannabis and cannabinoids for medical use based on clinical efficacy outcomes from randomized clinical trials (18) and reviews on the use of social media for illicit drug surveillance (107). However, following searches on PROSPERO and the databases listed above, to our knowledge, this paper constitutes the first systematic scoping review examining studies that used user-generated online text to understand the use of cannabis as a medicine in the global community.

Our scoping review found that the use of social media and internet search queries to investigate cannabis as a medicine is a rapidly emerging area of research. Over half of the studies included in this review were published within the last three years, this reflects not only increase community interest in the therapeutic potential of cannabinoids, but also world-wide trends towards cannabis legalization (4, 7-9, 108, 109). Regarding social media platforms, Twitter was the data source in eighteen (42.9%) of the 42 studies, almost three and a half times the number of studies using Reddit (6, 14.3%) and just under three times the number of studies using data from Online forums (5, 11.9%). Three (7.14%) GoFundMe studies and three (7.14%) Google Trends studies were also included in the review. Hence, much of the data in this systematic review

283 comprised posts from the Twitter platform. Several factors may explain this finding, firstly
284 Twitter is real-time in nature, it has a high volume of messages, and it is publicly accessible.
285 These factors makes it a useful data source for public health surveillance (110).

286 Regarding the subjects of the studies, twelve (28.6%) focused on general user-generated content
287 regarding the treatment of health conditions (glaucoma, autism, asthma, cancer, bowel disease,
288 brachial plexus injury, cluster headaches, opioid disorder). These studies were either explicitly
289 designed to investigate cannabis as a medicine or were studies that generated results that
290 incidentally found cannabis mentioned as an alternative or complimentary treatment (either
291 formally prescribed or via self-medication).

292 Qualitative studies featured in the research, but while their contributions are valuable, especially
293 in the context of hypothesis generation, they tend to be limited by their smaller datasets, which
294 frequently comprised manually annotated samples. The recent emergence of powerful machine
295 learning-based natural language processing (NLP) models suggests that it should be possible to
296 automate the continuous processing of far larger datasets using NLP technologies, built upon the
297 insights gained from initial qualitative studies, and even leveraging their annotated data for
298 training purposes. Recent trends in the social science data landscape have shown a convergence
299 between social science and computer science expertise, where the ability to use computational
300 methods has greatly assisted the collection and validation of robust datasets that can form the
301 basis of deeper social science research (111).

302 We found much heterogeneity in approaches applied to analyse user-related content, and
303 inconsistent quality in the methodologies adopted. While we endeavored to include as many
304 studies as possible, some of the publications initially identified as suitable for inclusion were not
305 suitable based on a minimum quality requirements checklist (Appendix 1). This checklist was

306 designed to ensure that selection of data source, choice of platform, data acquisition and
307 preparation, analysis and evaluation delivers data and conclusions that are appropriate for
308 answering the research questions.

309 The utilization of user-generated content for health research is subject to several inherent
310 limitations which include the lack of control that researchers have in relation to the credibility of
311 information, the frequently unknown demographic characteristics and geographical location of
312 individuals generating content, and the fact that social media users are not necessarily
313 representative of the wider community (112). Furthermore, the uniqueness, volume, and salience
314 of social media data has implications that need to be considered when used for health
315 information analysis (113). Volume is usually inversely related to salience: a platform such as
316 Twitter has a very high volume of information, but, apart from its frequency, much of which is
317 not highly pertinent for analyzing an effect; whereas the volume of information contained in a
318 blog will much less but likely more salient for analysis. Notwithstanding these limitations, user-
319 generated content comprises large-scale data that provides access to the unprompted organic
320 opinions and attitudes of cannabis users in their own words and is an effective medium through
321 which to gauge public sentiment. To date, insights regarding cannabis as medicine have gained
322 primarily through surveys or focus groups which have their own limitations regarding the format
323 of data collection and potential bias in participant recruitment. A limitation of this scoping
324 review was the lack of inclusion of a computational database such as *IEEE Xplore* in the search
325 strategy, and the exclusion of the search terms ‘infodemiology’ and ‘infoveillance’.
326 Infodemiology and infoveillance studies explicitly use web-based data for research, and *IEEE*
327 *Xplore* is a repository that contains technical papers and documents relating to computer science.

However, our search was systematic, comprehensive and *IEEE Xplore* is Scopus-indexed, and we expect data loss to be minimal.

Conclusion

Our systematic scoping review reflects a growing interest in the use of user-generated content for public health surveillance. It also demonstrates there is a need for the development of a systematic approach for evaluating the quality of social media studies and highlights the utility of automatic processing and computational methods (machine learning technologies) for large social media datasets. This systematic scoping review has shown that user-generated content as a data source for studying cannabis as a medicine provides another means to understand how cannabis is perceived and used in the community. As such, it is another potential 'tool' with which to engage in pharmacovigilance of, not only cannabis as a medicine, but also other novel therapeutics as they enter the market.

Acknowledgments

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

References

1. Russo EB. History of cannabis and its preparations in saga, science, and sobriquet. Chemistry and Biodiversity. 2007;4(8):1614-48.
2. Bridgeman MB, Abazia DT. Medicinal cannabis: history, pharmacology, and implications for the acute care setting. Pharmacy and therapeutics. 2017;42(3):180-.
3. Grinspoon L, Bakalar JB. Marihuana: the forbidden medicine. New Haven, USA: Yale University Press; 1993.

4. Kalant H. Medicinal use of cannabis: history and current status. Pain Research and Management. 2001;6(2):80-91.
5. Taylor S. Medicalizing cannabis—Science, medicine and policy, 1950–2004: An overview of a work in progress. Drugs: Education, Prevention and Policy. 2008;15(5):462-74.
6. Musto DF. The marihuana tax act of 1937. Archives of General Psychiatry. 1972;26(2):101-8.
7. EMCDDA. Cannabis policy: status and recent developments Lisbon, Portugal: The European Monitoring Centre for Drugs and Drug Addiction; 2021 [cited 2021 August]. Available from: https://www.emcdda.europa.eu/publications/topic-overviews/cannabis-policy/html_en.
8. Shover CL, Humphreys K. Six policy lessons relevant to cannabis legalization. Am J Drug Alcohol Abuse. 2019;45(6):698-706.
9. CPFG. Home Office Circular 1 November 2018: Rescheduling of cannabis-based products for medicinal use in humans. In: Unit CDaA, editor. London, United Kingdom: Crime, Policing, and Fire Group (CPFG); 2018.
10. National Academies of S, Medicine E. The health effects of cannabis and cannabinoids: the current state of evidence and recommendations for research. 2017.
11. Hall W. Medical use of cannabis and cannabinoids: questions and answers for policymaking. 2018.
12. Aguilar S, Gutiérrez V, Sánchez L, Nougier M. Medicinal cannabis policies and practices around the world. International Drug Policy Consortium. 2018(April):32-.
13. TGA. Access to medicinal cannabis products Canberra, Australia: Therapeutic Goods Administration2021 [Available from: Retrieved September, 2021, from https://www.tga.gov.au/access-medicinal-cannabis-products-1
14. National Academies of Sciences EaM. The health effects of cannabis and cannabinoids: the current state of evidence and recommendations for research. 2017.
15. Pearlson G. Medical marijuana and clinical trials. Weed Science: Cannabis Controversies and Challenges Academic Press; 2020. p. 243-60.
16. Sarris J, Sinclair J, Karamacoska D, Davidson M, Firth J. Medicinal cannabis for psychiatric disorders: A clinically-focused systematic review. BMC Psychiatry. 2020;20(1):1-14.
17. Suraev AS, Marshall NS, Vandrey R, McCartney D, Benson MJ, McGregor IS, et al. Cannabinoid therapies in the management of sleep disorders: A systematic review of preclinical and clinical studies. Sleep Medicine Reviews. 2020;53:101339-.
18. Whiting PF, Wolff RF, Deshpande S, Di Nisio M, Duffy S, Hernandez AV, et al. Cannabinoids for medical use: A systematic review and meta-analysis. JAMA - Journal of the American Medical Association. 2015;313(24):2456-73.
19. Schulze-Schiappacasse C, Durán J, Bravo-Jeria R, Verdugo-Paiva F, Morel M, Rada G. Are Cannabis, Cannabis-Derived Products, and Synthetic Cannabinoids a Therapeutic Tool for Rheumatoid Arthritis? A Friendly Summary of the Body of Evidence. JCR: Journal of Clinical Rheumatology. 2021;Publish Ah(00):1-5.
20. Bonomo Y, Souza JDS, Jackson A, Crippa JAS, Solowij N. Clinical issues in cannabis use. British Journal of Clinical Pharmacology. 2018;84(11):2495-8.
21. Owens B. The professionalization of cannabis growing. Nature. 2019;572(7771).
22. Hallinan CM, Bonomo YA. 'The Rise and Rise of Medicinal Cannabis, What Now? Medicinal Cannabis Prescribing in Australia 2017-2022'. International Journal of Environmental Research and Public Health. 2022;ijerph-1806974.
23. Karanges EA, Suraev A, Elias N, Manocha R, McGregor IS. Knowledge and attitudes of Australian general practitioners towards medicinal cannabis: a cross-sectional survey. British Medical Journal Open. 2018;8(7):e022101.

24. Caputi TL. The medical marijuana industry and the use of "research as marketing". American Journal of Public Health. 2020;110(2):174-5.
25. Ayers JW, Caputi TL, Leas EC. The need for federal regulation of marijuana marketing. Jama. 2019;321(22):2163-4.
26. Bonn-Miller MO, Loflin MJE, Thomas BF, Marcu JP, Hyke T, Vandrey R. Labeling accuracy of cannabidiol extracts sold online. JAMA - Journal of the American Medical Association. 2017;318(17):1708-9.
27. Hallinan CM, Eden E, Graham M, Greenwood LM, Mills J, Popat A, et al. Over the counter low-dose cannabidiol: A viewpoint from the ACRE Capacity Building Group. Journal of Psychopharmacology. 2021.
28. Fitzcharles MA, Shir Y, Häuser W. Medical cannabis: Strengthening evidence in the face of hype and public pressure. Cmaj. 2019;191(33):E907-E8.
29. Freeman TP, Hindocha C, Green SF, Bloomfield MAP. Medicinal use of cannabis based products and cannabinoids. BMJ (Online). 2019;365(April):1-7.
30. Graham M, Martin JH, Lucas CJ, Hall W. Translational hurdles with cannabis medicines. 2020(December 2019):1325-30.
31. US Food and Drug Administration. Framework for FDA's real world evidence program. 2018.
32. Nabhan C, Klink A, Prasad V. Real-world Evidence - What Does It Really Mean? JAMA Oncology. 2019;5(6):2019-21.
33. 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;11(10):1-12.
34. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. Online Health Information Seeking Among US Adults: Measuring Progress Toward a Healthy People 2020 Objective. Public Health Reports. 2019;134(6):617-25.
35. Gualtieri LN. The doctor as the second opinion and the internet as the first. 2009. p. 2489-98.
36. Shakeri Hossein Abad Z, Kline A, Sultana M, Noaeen M, Nurmambetova E, Lucini F, et al. Digital public health surveillance: a systematic scoping review. npj Digital Medicine. 2021;4(1):41.
37. Paul MJ, Sarker A, Brownstein JS, Nikfarjam A, Scotch M, Smith KL, et al. Social Media Mining for Public Health Monitoring and Surveillance. Pacific Symposium on Biocomputing. 2016;21(January):468-79.
38. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. PLOS Medicine. 2021;18.
39. Olteanu A, Castillo C, Diaz F, Kıcıman E. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Frontiers in Big Data. 2019;2.
40. Thompson L, Rivara FP, Whitehill JM. Prevalence of Marijuana-Related Traffic on Twitter, 2012-2013: A Content Analysis. Cyberpsychology, Behavior, and Social Networking. 2015;18(6):311-9.
41. McGregor F, Somner JEA, Bourne RR, Munn-Giddings C, Shah P, Cross V. Social media use by patients with glaucoma: what can we learn? Ophthalmic and Physiological Optics. 2014;34(1):46-52.
42. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. Journal of Adolescent Health. 2015;56:139-45.
43. 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 and Asthma Proceedings. 2015;36:e121-e6.
44. Krauss MJ, Sowles SJ, Mylvaganam S, Zewdie K, Bierut LJ, Cavazos-Rehg PA. Displays of dabbing marijuana extracts on YouTube. Drug Alcohol Depend. 2015;155:45-51.
45. Thompson L, Rivara FP, Whitehill JM. Prevalence of Marijuana-Related Traffic on Twitter, 2012-2013: A Content Analysis. Cyberpsychology, Behavior, and Social Networking. 2015;18:311-9.

46. Cavazos-Rehg PA, Sowles SJ, Krauss MJ, Agbonavbare V, Grucza R, Bierut L. A content analysis of tweets about high-potency marijuana. Drug Alcohol Depend. 2016;166:100-8.
47. 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;164:64-70.
48. Mitchell JT, Sweitzer MM, Tunno AM, Kollins SH, Joseph McClernon F. "I use weed for my ADHD": A qualitative analysis of online forum discussions on cannabis use and ADHD. PloS one. 2016;11:1-13.
49. Andersson M, Persson M, Kjellgren A. Psychoactive substances as a last resort-a qualitative study of self-treatment of migraine and cluster headaches. Harm Reduction Journal. 2017;14:1-10.
50. Greiner C, Chatton A, Khazaal Y. Online self-help forums on cannabis: A content assessment. Patient Education and Counseling. 2017;100:1943-50.
51. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Murphy GM, Bierut LJ. Exposure to and Content of Marijuana Product Reviews. Prevention Science. 2018;19:127-37.
52. Meacham MC, Roh S, Chang JS, Ramo DE. Frequently asked questions about dabbing concentrates in online cannabis community discussion forums. International Journal of Drug Policy. 2019;74:11-7.
53. Jia J, Mehran N, Purgert R, Zhang Q, Lee D, Myers JS, et al. Marijuana and Glaucoma: A Social Media Content Analysis. Ophthalmology Glaucoma. 2020.
54. 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 network open. 2020;3:e2020977.
55. Merten JW, Gordon BT, King JL, Pappas C. Cannabidiol (CBD): Perspectives from Pinterest. Substance Use & Misuse. 2020;55(13):2213-20.
56. 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. Journal of Cancer Research and Clinical Oncology. 2020;146:1857-65.
57. 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;17(2):1-16.
58. Zenone M, Snyder J, Crooks VA. What are the informational pathways that shape people's use of cannabidiol for medical purposes? Journal of Cannabis Research. 2021;3(1).
59. Pang RD, Dormanesh A, Hoang Y, Chu M, Allem J-P. Twitter Posts About Cannabis Use During Pregnancy and Postpartum:A Content Analysis. Substance Use & Misuse. 2021;56(7):1074-7.
60. Rhidenour KB, Blackburn K, Barrett AK, Taylor S. Mediating Medical Marijuana: Exploring How Veterans Discuss Their Stigmatized Substance Use on Reddit. Health Communication. 2021:1-11.
61. 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. Journal of hand surgery global online. 2021;3(1):36-40.
62. Zenone M, Snyder J, Caulfield T. Crowdfunding cannabidiol (CBD) for cancer: Hype and misinformation on gofundme. American Journal of Public Health. 2020;110:S294-S9.
63. Leas EC, Nobles AL, Caputi TL, Dredze M, Smith DM, Ayers JW. Trends in Internet Searches for Cannabidiol (CBD) in the United States. JAMA network open. 2019;2:e1913853.
64. Shi S, Brant AR, Sabolch A, Pollom E. False News of a Cannabis Cancer Cure. Cureus. 2019;11(1):e3918-e.
65. Saposnik FE, Huber JF. Trends in web searches about the causes and treatments of autism over the past 15 years: Exploratory infodemiology study. JMIR Pediatrics and Parenting. 2020;3.
66. 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;155:307-11.

67. Nasralah T, El-gayar OF, Wang Y. What Social Media Can Tell Us About Opioid Addicts : Twitter Data Case Analysis. 2019.
68. Mullins CF, Ffrench-O'Carroll R, Lane J, O'Connor T. Sharing the pain: An observational analysis of Twitter and pain in Ireland. Regional Anesthesia and Pain Medicine. 2020;45(8):597-602.
69. Dai H, Hao J. Mining social media data on marijuana use for Post Traumatic Stress Disorder. Computers in Human Behavior. 2017;70:282-90.
70. Turner J, Kantardzic M. Geo-social analytics based on spatio-temporal dynamics of marijuana-related tweets. ACM International Conference Proceeding Series. 2017;Part F1282:28-38.
71. 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 and Tobacco Research. 2017;19:952-9.
72. Yom-Tov E, Lev-Ran S. Adverse Reactions Associated With Cannabis Consumption as Evident From Search Engine Queries. JMIR Public Health and Surveillance. 2017;3:e77.
73. Glowacki EM, Glowacki JB, Wilcox GB. A text-mining analysis of the public's reactions to the opioid crisis. Substance Abuse. 2018;39:129-33.
74. 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;188:364-9.
75. Allem J-P, Escobedo P, Dharmapuri L. Cannabis Surveillance With Twitter Data: Emerging Topics and Social Bots. American Journal of Public Health. 2020;110(3):357-62.
76. van Draanen J, Tao H, Gupta S, Liu S. Geographic Differences in Cannabis Conversations on Twitter: Infodemiology Study. JMIR Public Health and Surveillance. 2020;6:e18540.
77. Soleymanpour M, Saderholm S, Kavuluru R. Therapeutic Claims in Cannabidiol (CBD) Marketing Messages on Twitter. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Bioinformatics and Biomedicine (BIBM), 2021 IEEE International Conference on: IEEE; 2021. p. 3083-8.
78. Allem J-P, 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 formative research. Canada: JMIR Publications; 2022. p. e35027.
79. Turner J, Kantardzic M, Vickers-Smith R. Infodemiological Examination of Personal and Commercial Tweets About Cannabidiol: Term and Sentiment Analysis. Journal of Medical Internet Research. 2022;23(12):e27307.
80. 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. Journal of Medical Internet Research. 2019;21.
81. Tran T, Kavuluru R. Social media surveillance for perceived therapeutic effects of cannabidiol (CBD) products. International Journal of Drug Policy. 2020;77:102688.
82. 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. Journal of Medical Internet Research. 2020;22.
83. Andersson M, Persson M, Kjellgren A. Psychoactive substances as a last resort-a qualitative study of self-treatment of migraine and cluster headaches. Harm Reduction Journal. 2017;14(1):1-10.
84. Cavazos-Rehg PA, Krauss M, Fisher SL, Salyer P, Grucza RA, Bierut LJ. Twitter chatter about marijuana. Journal of Adolescent Health. 2015;56(2):139-45.
85. Greiner C, Chatton A, Khazaal Y. Online self-help forums on cannabis: A content assessment. Patient Education and Counseling. 2017;100(10):1943-50.
86. 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 network open. 2020;3(10):e2020977-e.

87. Merten JW, Gordon BT, King JL, Pappas C. Cannabidiol (CBD): Perspectives from Pinterest. Substance Use and Misuse. 2020;55(13):2213-20.
88. Mitchell JT, Sweitzer MM, Tunno AM, Kollins SH, Joseph McClernon F. "I use weed for my ADHD": A qualitative analysis of online forum discussions on cannabis use and ADHD. PLoS ONE. 2016;11(5):1-13.
89. Pang RD, Dormanesh A, Hoang Y, Chu M, Allem JP. Twitter Posts About Cannabis Use During Pregnancy and Postpartum:A Content Analysis. Substance Use and Misuse. 2021;0(0):1-4.
90. 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. Journal of Cancer Research and Clinical Oncology. 2020;146(7):1857-65.
91. Rhidenour KB, Blackburn K, Barrett AK, Taylor S. Mediating Medical Marijuana: Exploring How Veterans Discuss Their Stigmatized Substance Use on Reddit. Health Communication. 2021;00(00):1-11.
92. Yue L, Chen W, Li X, Zuo W, Yin M. A survey of sentiment analysis in social media. Knowledge and Information Systems. 2019;60(2):617-63.
93. Allem J-P, 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;6(2):e35027.
94. Nasralah T, Spohn R. Social Media Text Mining Framework for Drug Abuse : An Opioid Crisis Case Analysis SOCIAL MEDIA TEXT MINING FRAMEWORK FOR DRUG ABUSE : AN OPIOID CRISIS CASE ANALYSIS A dissertation submitted to Dakota State University in partial fulfillment of the require. 2019.
95. van Draanen J, Tao H, Gupta S, Liu S. Geographic Differences in Cannabis Conversations on Twitter: Infodemiology Study. JMIR Public Health and Surveillance. 2020;6(4):e18540-e.
96. Li M, Kakani N, li C, Park A. Understanding cannabis information on social media: Examining tweets from verified, regular, and suspended users2020. 1-10 p.
97. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Murphy GM, Bierut LJ. Exposure to and Content of Marijuana Product Reviews. Prevention Science. 2018;19(2):127-37.
98. Shi S, Brant AR, Sabolch A, Pollom E. False News of a Cannabis Cancer Cure. Cureus. 2019;11(1):1-11.
99. 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. Journal of Medical Internet Research. 2019;21(8).
100. Glowacki EM, Glowacki JB, Wilcox GB. A text-mining analysis of the public's reactions to the opioid crisis. Substance Abuse. 2018;39(2):129-33.
101. 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;17(2):e0263583.
102. 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 and Asthma Proceedings. 2015;36(6):e121-e6.
103. 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 and Tobacco Research. 2017;19(8):952-9.
104. Saposnik FE, Huber JF. Trends in web searches about the causes and treatments of autism over the past 15 years: Exploratory infodemiology study. JMIR Pediatrics and Parenting. 2020;3(2).
105. Leas EC, Nobles AL, Caputi TL, Dredze M, Smith DM, Ayers JW. Trends in Internet Searches for Cannabidiol (CBD) in the United States. JAMA network open. 2019;2(10):e1913853-e.
106. Tran T, Kavuluru R. Social media surveillance for perceived therapeutic effects of cannabidiol (CBD) products. International Journal of Drug Policy. 2020;77:102688-.

590 107. Kazemi DM, Borsari B, Levine MJ, Dooley B. Systematic review of surveillance by social media
591 platforms for illicit drug use. Journal of Public Health (United Kingdom). 2017;39(4):763-76.
592 108. Grinspoon L, Bakalar JB. Marihuana, the forbidden medicine: Yale University Press; 1997.
593 109. Belackova V, Ritter A, Shanahan M, Chalmers J, Hughes C, Barratt M, et al. Medicinal cannabis in
594 Australia–Framing the regulatory options. Sydney: Drug Policy Modelling Program. 2015.
595 110. Lardon J, Bellet F, Aboukhamis R, Asfari H, Jaulent M, Beyens M, et al. Evaluating Twitter as a
596 complementary data source for pharmacovigilance Expert Opinion on Drug Safety. 2018;17(8):763–74.
597 111. Bode L, Davis-Kean P, Singh L, Berger-Wolf T, Budak C, Chi G, et al. Study designs for quantitative
598 social science research using social media. 2020.
599 112. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on
600 social media: a critical review. npj Digital Medicine. 2020;3(1).
601 113. Abbasi A, Sarker A, Ginn R, Smit K, Oconnor K, Abbasi A. Social Media Analytics for Smart Health.

602 Supporting information

603 Appendix I. Search strategies for each database (Doc)

604 Appendix II. Quality assessment checklist (Doc)

605 Appendix III. Excluded studies (Doc)

606 S1. Database queries (Doc)

607 S2. Database queries results (Xlsx)

608 S3. PRISMA Checklist (Doc)

609