medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Social Media Surveillance for Perceived Therapeutic Effects of Cannabidiol (CBD) Products

Tung Tran, M.S.
Department of Computer Science
University of Kentucky, USA
tung.tran@uky.edu

***** Ramakanth Kavuluru, Ph.D.
Division of Biomedical Informatics, Department of Internal Medicine
Department of Computer Science
University of Kentucky, USA
ramakanth.kavuluru@uky.edu

* **Corresponding Author**

Short Communication

Word Count: 2250
Figures: 1
Tables: 2

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Abstract

Background: CBD products have risen in popularity given CBD's therapeutic potential and lack of legal oversight, despite lacking conclusive scientific evidence for widespread over-the-counter usage for many of its perceived benefits. While medical evidence is being generated, social media surveillance offers a fast and inexpensive alternative to traditional surveys in ascertaining perceived therapeutic purposes and modes of consumption for CBD products.

Methods: We collected all comments from the CBD subreddit posted between January 1 and April 30, 2019 as well as comments submitted to the FDA regarding regulation of cannabis-derived products and analyzed them using a rule-based language processing method. A relative ranking of popular therapeutic uses and product groups for CBD is obtained based on frequency of pattern matches including precise queries that entail identifying mentions of the condition, a CBD product, and some "trigger" phrase indicating therapeutic use. We validated the social media-based findings using a similar analysis on comments to the U.S. Food and Drug Administration's (FDA) 2019 request-for-comments on cannabis-derived products.

Results: CBD is mostly discussed as a remedy for anxiety disorders and pain and this is consistent across both comment sources. Of comments posted to the CBD subreddit during the monitored time span, 6.19% mentioned anxiety at least once with at least 6.02% of these comments specifically mentioning CBD as a treatment for anxiety (i.e., 0.37% of total comments). The most popular CBD product group is oil and tinctures.

Conclusion: Social media surveillance of CBD usage has the potential to surface new therapeutic use-cases as they are posted. Contemporary social media data indicate, for example, that stress and nausea are frequently mentioned as therapeutic use cases for CBD without corresponding evidence, that affirms or denies, in the research literature. However, the abundance of anecdotal claims warrants serious scientific exploration moving forward. Meanwhile, as FDA ponders regulation, our effort demonstrates that social data offers a convenient affordance to surveil for CBD usage patterns in a way that is fast and inexpensive and can inform conventional electronic surveys.

Keywords: cannabidiol; CBD; hemp oil; cannabis; social media surveillance; text mining

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Cannabidiol (CBD) is a well-known cannabinoid compound derived from certain strains of *Cannabis Sativa.* It has been shown to have high potential for therapeutic efficacy and low potential for abuse and dependency in humans (World Health Organization 2017); however, evidence is not substantial enough to warrant wide spread over-the-counter availability of CBD products for perceived therapeutic effects (Cogan 2019, Cohen and Sharfstein 2019). As of now, the U.S. Food and Drug Administration (FDA) has approved a CBD based drug only for epilepsy in June 2018 (U.S. Food and Drug Administration 2018). Nevertheless, public interest in CBD has skyrocketed in recent years. The 2018 United States Farm Bill, which became law on December 20, 2018, included a provision that removed hemp from Schedule 1 controlled substances. As a result, CBD based products, derived primarily from hemp plant extracts, are now ubiquitous in the U.S. marketplace for over-the-counter purchase (backed by major retailers including Walgreens, Kroger, and CVS) and remain relatively unregulated (Reily 2019). Even if there is no long-term harm, aggressive marketing of CBD products (oils, edibles, topicals, and vapes) by the cannabis industry could lead to significant cost burden to consumers as they may regularly buy products for claimed/perceived health benefits without conclusive clinical evidence. Additionally, because of aggressive CBD marketing, consumers may forgo more evidence-based treatments in favor of CBD products.

As more medical evidence is being generated and consolidated, it is critical to keep track of therapeutic purposes that consumers indicate on social media to inform policy and prevention initiatives. In this first of its kind study for CBD, we analyze perceived or expected therapeutic uses for CBD based on social media discussions. Our precise rule-based method is used to ascertain the relative ranking of different therapeutic reasons for the use of CBD according to public perception. We apply a similar method to identify popular ways in which CBD is consumed. Given suitable anonymity properties, we chose the CBD subreddit on the Reddit social platform for this effort. We additionally validate our method on a collection of public comments submitted as part of a recent FDA request for comments (RFC Docket ID**:** FDA-2019-

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

N-1482) regarding regulation of cannabis-derived products. While traditional surveys and literature reviews are powerful tools for assessing potential uses for CBD, we argue that automated surveillance of social media platforms offer a fast and inexpensive alternative that could inform and complement more traditional sources of surveillance. Specifically, it is possible to continually perform this kind of analysis in a live or streaming fashion. Furthermore, this type of automatic analysis on CBD usage and effects may facilitate the construction and deployment of traditional surveys by helping to populate survey choices.

Background

Althouse et al. (2015) first introduced the concept of *novel data streams* (NDS) in the context of public health surveillance and defined NDS as data streams with contents that are generated by users (or patients) themselves, such as social media and web search data, and excludes traditional data sources such as electronic health record, disease registries, and vital statistics. Paul et al. (2016) compared information reported in online forums (i.e., NDS data) against large epidemiology survey data and found significant correlations between the two sources; their results suggest that while online forums are not reliable for predicting long-term trends, they are a reliable source for estimating temporal demographic associations and trends in drug use. A recent study to assess the quality of solicited advice for treatment of chronic pain (including treatment using cannabis products) in Reddit comments found that at least 33.9% had at least some support in the research literature (K. L. Costello 2019), suggesting that health-related discussions on Reddit are, to a reasonable extent, substantial and nuanced. Indeed, several studies have analyzed Reddit communities (subreddits) as a means of exploring temporal trends and deriving insights related to use of cannabis products (Sowles, et al. 2017, Meacham, Paul and Ramo 2018, Costello, Martin III and Edwards Brinegar 2017). Costello et al. (2017) specifically found that drug-related subreddits often contain banter (i.e., off-topic discussions) and argues that banter is an essential part of discussions as they foster disclosure and information-seeking behavior. Enghoff and Aldridge (2019) has since argued that unsolicited online data, including Reddit comment data, is increasing valuable in the digital age for the purpose of informing drug policy. To our knowledge, no studies have explored NDS data to determine perceived remedial effects and usage patterns for CBD.

Materials and Methods

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

We perform our analysis on two collection of comments from the following sources.

1. A collection of all comments posted to the CBD subreddit between January 1, 2019 and April 30, 2019 totaling to 64,099 individual comments. To keep our methodology simple, we only analyzed comments (as opposed to "posts"), which were treated as independent units of discourse irrespective of the complex threading structure. We sourced Reddit comment data from the *pushshift.io* platform[2] which comprehensively archives Reddit comments on a monthly basis (while trailing behind the "live" data by several months). We filtered for comments specifically posted to the CBD reddit within the target date range, which corresponds to the four-month period roughly between when the 2018 Farm Bill became law and the date of the latest comments available on *pushshift.io* at the time of writing.
2. A collection of 3832 machine-processable comments[1] submitted to FDA's RFC on cannabis derived products between April 3, 2019 and July 19, 2019. There were 4272 comments in total[1] by the time the comment period closed on July 19, 2019, and of them 3832 were machine processable comments as ASCII text (mostly from consumers) while the rest included PDF attachments (some paper scans) that were not readily processible, mostly including formal opinions from organizations that are in the stakeholder group for CBD products. We exclusively focused on typed-in comments that were from consumers.

These platforms were chosen for our analyses because of the tendency for comments to be verbose and focused with respect to the topic of cannabis-derived products.

First, we use the MetaMap concept identification and normalization tool (U.S. National Library of Medicine 2019) to identify frequently mentioned psychological and physiological conditions at the concept level. MetaMap is a knowledge-based tool used to identify mentions of medical concepts (e.g., drugs, diseases, and symptoms) in free text and link them to a concept-unique IDs. We specifically considered concepts that are either of semantic type *Mental and Behavioral Dysfunction (MOBD)* or *Disease or Disorder (DSYN)*. Based on the concepts found, we manually curated an exhaustive list of *target conditions*; each target condition is associated with a dictionary of terms (i.e., ways in which a condition is expressed)

[1]https://www.regulations.gov/docket?D=FDA-2019-N-1482
[2]https://files.pushshift.io/reddit/comments/

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

through manual review of disease and disorder concepts frequently identified by MetaMap. For example, for seizure disorders, we look for mentions of terms such as “epilepsy”, “epileptic”, and “seizure/s”.

As the CBD subreddit may contain “off-topic” discussion of ailments not directly related to CBD, we also experimented with increasingly precise queries, based on regular expression patterns to match comments that specifically mention the target condition as well as CBD and some therapeutic phrase. The three type of queries corresponding to different levels of restrictiveness are as follows. For each comment, we

1. search for mentions of the target condition,
2. search for mentions of the target condition and CBD within the same sentence,
3. search for mentions of the target condition, CBD, and some therapeutic phrase indicating treatment where each mention is separated by at most 36 characters (half the observed average sentence length). We match on several variations of the pattern to allow matching on all possible orders, allow for synonyms of CBD including cannabidiol and hemp oil, and allow for queries based on different therapeutic phrases.

Here, therapeutic phrases come from a hand-crafted lexicon of 87 terms including “treats”, “cures”, “helps”, “reduces”, “alleviates”, “relieves”, “eliminates”, “kills”, “stops”, “eases”, “aids”, “soothes”, “inhibits”, “improves”, “destroys”, “reverses”, “suppresses”, “lowers”, “regulates”, “prevents”, “manages”, “fixes”, “better” and their variants including conjugated forms. The lexicon was developed by manually examining samples of comments that matched queries (1) and (2). From the initial list of phrases, we added several more after consulting an online thesaurus for likely synonyms. The queries are performed in a case insensitive fashion and matches containing negation terms (including “never”, “not”, and related contractions) are disregarded. We emphasize that *frequency* or *count* is defined as the number of *unique* comments containing the term or matching a pattern-based query; that is, a term or pattern match is only counted at most once per comment even if it occurs multiple times in some comments. Comments mentioning multiple conditions will increment the match count for each condition separately. For example, a comment containing the statement “CBD helps my anxiety and pain” would increment the match count for both *anxiety* and *pain* at all levels of restrictiveness. By manually examining 100 randomly sampled

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

matches at the strictest (third) query type, we estimate our method to be 96% accurate (more precisely, 96% positive predictive value rate) with a 95% confidence interval of [90.16%, 98.43%]. However precise, we note that the frequency of pattern matches is an underestimate of the true frequency.

To explore the modes of consumption for CBD products, we curated a list of popular CBD products based on a manual review of popular CBD e-commerce websites. We group them into five broad product groups including Oils/Tinctures, Vapes, Edibles, Pills/Capsules, and Topicals. Less obvious terms were obtained by querying similar word vectors as induced by the distributional semantics software Word2Vec (Google Code Archive 2013) on the comment data; for example, by querying for words similar to “vape”, “vaping”, we obtained additional terms such as “dab” and “wax” which are associated with a less well-known mode of inhalation. We note that mentions of vaping, for example, are not all CBD-related; some posts are simply discussing the default practice of vaping *nicotine liquid*. As such, we additionally perform more precise queries to match for mentions of CBD products that explicitly contain as a prefix “CBD”, “cannabidiol”, or “hemp” such as “CBD tea” or “hemp lotion.”

Results and Discussion

We report our analysis of top ten mentioned conditions in Table 1, sorted by matching at the strictest level on Reddit comments, where the second column is a list of terms for the target condition along with their individual frequencies (in parentheses). The next six columns are frequencies for matches of queries at varying levels of restrictiveness, based on pattern rules, with an example at the last column. We present a plot of the proportions for each condition along with their 95% confidence intervals in Figure 1. Based on these results, anxiety disorders and pain are the two conditions dominating much of the discussion surrounding CBD, both in terms of general discussion and as a perceived therapeutic treatment. This is consistent for both comments posted to the CBD subreddit and comments submitted to the FDA’s 2019 RFC on cannabis-derived products.

Despite having an order of magnitude more Reddit comments than comments to the FDA, we note that there is not a dramatic difference in the number of term/pattern matches between the two platforms. This may indicate that comments to the FDA, as expected, are highly focused on the perceived therapeutic effects of CBD products while Reddit comments are more likely to include off-topic discussions. Given the

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

match counts/frequencies are comment-unique (and underestimates the true frequency), it is possible to assess, for example that 6.19% (3968 out of 64099) of comments posted to the CBD subreddit mention anxiety, and that at least 6.02% (239 out of 3968) of these comments explicitly discuss CBD as a potential or perceived remedy for anxiety (0.37% of total comments); for pain, these percentages are 3.22% and 5.56% respectively (0.17% of total comments). On the other hand, 27.00% of comments to the FDA RFC mention some form of anxiety disorder, and at least 15.39% of these comments explicitly mention CBD as a potential therapeutic solution (4.16% of total comments); for pain, these percentages are much more prominent at 47.57% and 16.18% respectively (7.60% of total comments).

From Table 1, it can be surmised that Reddit comments tend to focus more on mental conditions while FDA comments tend to focus more on physiological conditions. To assess this, we performed the original experiment on groups of conditions (from Table 1) organized based on whether they are in the mental (anxiety, depression, stress) or physiological (pain, inflammation, headache, sleep disorder, seizure disorders, nausea, and cancer) category. Reddit comments had mentions of mental conditions in at least 4373 comments (6.82% with a 95% confidence interval of [6.63%, 7.02%]) and physiological conditions in at least 3262 comments (5.09% with a 95% confidence interval of [4.92%, 5.26%]). On the other hand, FDA comments had mentions of mental conditions in at least 1191 comments (31.02% with a 95% confidence interval of [29.63%, 32.56%]) and physiological conditions in at least 2260 comments (58.98% with a 95% confidence interval of [57.41%, 60.52%]). The reported distributions seem to affirm that FDA comments tend to focus heavily on physiological conditions (compared with mental conditions), while Reddit comments tend to focus only slightly more on mental conditions (compared with physiological conditions). The difference in anonymity provided by these platforms and the greater stigma associated with disclosure of mental health conditions may be one explanation for the discrepancy; while FDA comments may be submitted anonymous, most commenters (82%) chose to submit comments with a partial or full name.

We use superscript (A) in Table 1 to indicate conditions covered in a recent review of human studies assessing the potential of CBD (White 2019) and superscript (B) in Table 1 to indicate conditions covered in a recent survey study of CBD users (Corroon and Phillips 2018). Among conditions frequently discussed on social media that are **not** discussed in research literature are stress and nausea. Less

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

frequently discussed conditions (not shown in Table 1) for which there is little or no research evidence (as observed from PubMed searches) include ADHD and autism, with users making comments such as, “I will say from my personal experience that hemp flower and oil have really helped my ADHD.”

We similarly report popular modes of consumption in Table 2. We found that CBD oil and tinctures are most popular, either as food additives or directly administered sublingually, with vaping being the second most popular mode of consumption. Approximately 13% of the Reddit comments mention oil or tinctures, and 25% of these are explicitly mentioned as a CBD product (3.25% of total comments). In FDA comments, oil and tinctures are likewise most popular at 1442 mentions; however, vapes were mentioned the least number of times at a frequency of 87. Other modes of consumption, including topicals, pills/capsules, and edibles, are mentioned in at similar levels of frequency at 319, 295, and 254 matches respectively.

Limitations and Ethics

Given the precise nature of our methodology (and rule-based methods in general), our results in fact lead to underestimates of proportions of Reddit posters perceiving therapeutic potential. However, we contend that our method is sufficient as the goal is to obtain a relative ranking among popularly discussed conditions and not necessarily to obtain absolute estimates of CBD users using it for a particular therapeutic reason. As a limitation, our analyses and estimates do not necessarily reflect the perception and usage patterns of the general population of CBD users, who may or may not comment in these specific forums. Such an estimate can only be obtained by a conventional survey targeting a truly random sample of consumers and not just those who post on Reddit or submit comments to the FDA. However, our effort has the potential to surface new therapeutic use-cases (as they are posted), which can then be included as options in more traditional surveys that can be timed at regular intervals as per resource constraints.

The ethics of using social media data in research is still a contentious topic. Given their private yet public nature, some researchers argue that social media data should be treated as public data that can used without ethical considerations while others argue that it should be regarded as research with human participants and thus treated with the appropriate ethical stipulations (Enghoff and Aldridge 2019). Herein,

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

we address ethical concerns related to our methodology. First, we note that, unlike many social media platforms, Reddit users typically use pseudonyms or handles not related to their real identity. As of the time of this writing, only a username and password and no other identifying information are required for registration to the platform. Second, we note that data used in our research is *openly and publicly available* in that no form of registration or admittance is required to access them. In example comments shown in Table 1, we have changed minor details to eliminate any potential for identification. Lastly, besides providing anonymized examples of comments for context, we report only aggregate data. Thus we believe that our study conforms to common guidelines for ethical research using online social media data (Markham and Buchanan 2012, Gonzalez-Hernandez 2019).

We note that online groups and communities may feel ownership over the content they produce. While effort has been taken to ensure anonymity, the analyses in this study were performed without input from individual users, content moderators, or community leaders. While permission cannot be obtained post-factum, in future studies of a similar nature, we recommend reaching out to community leaders and content owners with offers of research collaboration early on. In doing so, researchers exercise best practices for social media data research.

Conclusion and Future Work

CBD's fast-growing popularity fueled by the current relatively unregulated landscape warrants a serious and continuous exploration of perceived therapeutic claims by consumers. We took a first step in that direction in this effort by mining CBD related social media chatter. Specifically, we analyzed social media comments to the Reddit platform and consumer comments to the 2019 FDA RFC regarding cannabis product regulation to attain a relative ranking of perceived therapeutic uses for CBD. We analyzed these comments at varying levels of granularity with respect to a target condition, including assessing whether the comment simply mentioned the target condition or if there was an explicit mention of therapeutic effect related to CBD usage. Additionally, using a similar method, we obtain a relative ranking of popular CBD products as measured by discussion frequency. To our knowledge, this is the first effort to mine CBD related online content. In future efforts, we will explore the use of smarter and less mechanical methods, such as machine learning classifiers, to recognize mentions of conditions and products. We believe such an approach will be less likely to underestimate the true frequencies. Moreover, we will

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

explore the use of fuzzy matching (or spelling correction as a preprocessing step) as a clever way to improve the sensitivity of the proposed method. As FDA ponders regulation, we believe our effort demonstrates that social data offers a convenient affordance to surveil for CBD usage patterns in a way that is fast and inexpensive and can be deployed in a live fashion, offering unique complementary advantages to more traditional surveys.

Acknowledgments

This work is supported by the U.S. National Cancer Institute through NIH grant R21CA218231. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

References

Althouse, Benjamin M, Samuel V Scarpino, Lauren Ancel Meyers, John W Ayers, Marisa Bargsten, Joan Baumbach, John S Brownstein, et al. 2015. "Enhancing disease surveillance with novel data streams: challenges and opportunities." *EPJ Data Science* 4 (1): 1--8.

Cogan, Peter S. 2019. "On healthcare by popular appeal: critical assessment of benefit and risk in cannabidiol based dietary supplements." *Expert review of clinical pharmacology* (Taylor \& Francis) 12 (6): 501-511.

Cohen, Pieter A, and Joshua Sharfstein. 2019. "The Opportunity of CBD—Reforming the Law." *New England Journal of Medicine* (Mass Medical Soc) 381 (4): 297--299.

Corroon, Jamie, and Joy A Phillips. 2018. "A cross-sectional study of cannabidiol users." *Cannabis and cannabinoid research* 3 (1): 152--161.

Costello, Kaitlin L, John D Martin III, and Ashlee Edwards Brinegar. 2017. "Online disclosure of illicit information: Information behaviors in two drug forums." *Journal of the Association for Information Science and Technology* 68 (10): 2439--2448.

Costello, Kaitlin Light. 2019. "Quality of medical advice for chronic pain on social platforms." *Proceedings of the 82nd Annual Meeting of the Association for Information Science and Technology.* 381--385.

Enghoff, Oskar, and Judith Aldridge. 2019. "The value of unsolicited online data in drug policy research." *International Journal of Drug Policy.*

Gonzalez-Hernandez, Graciela. 2019. "On the Ethics of Using Social Media Data for Health Research." *Innovations in Health Information from the Director of the National Library of Medicine.* June 25. Accessed November 18, 2019. https://nlmdirector.nlm.nih.gov/2019/06/25/on-the-ethics-of-using-social-media-data-for-health-research/comment-page-1/.

Google Code Archive. 2013. *Word2vec: Tool for computing continuous distributed representations of words.* July 29. Accessed July 1, 2019. https://code.google.com/archive/p/word2vec/.

Markham, Annette, and Elizabeth Buchanan. 2012. "Ethical Decision-Making and Internet Research." *Recommendations from the AoIR Ethics Working Committee.* 1--19.

Meacham, Meredith C, Michael J Paul, and Danielle E Ramo. 2018. "Understanding emerging forms of cannabis use through an online cannabis community: an analysis of relative post volume and subjective highness ratings." *Drug and Alcohol Dependence* 188: 364--369.

Paul, Michael J, Margaret S Chisolm, Matthew W Johnson, Ryan G Vandrey, and Mark Dredze. 2016. "Assessing the validity of online drug forums as a source for estimating demographic and temporal trends in drug use." *Journal of Addiction Medicine* 10 (5): 324--330.

Reily, Laura. 2019. "CBD-infused food and beverages are still illegal under US law. So why are they everywhere?" *Chicago Tribune.* June 24. Accessed July 1, 2019. https://www.chicagotribune.com/marijuana/sns-cbd-infused-food-still-illegal-20190724-yp3d7n3fpjd2zb4eyczwoepnla-story.html.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Sowles, Shaina J, Melissa J Krauss, Lewam Gebremedhn, and Patricia A Cavazos-Rehg. 2017. "“I feel like I've hit the bottom and have no idea what to do”: Supportive social networking on Reddit for individuals with a desire to quit cannabis use." *Substance Abuse* 38 (4): 477--482.

U.S. Food and Drug Administration. 2018. "FDA approves first drug comprised of an active ingredient derived from marijuana to treat rare, severe forms of epilepsy." *FDA Newsroom.* June 25. Accessed July 1, 2019.

U.S. National Library of Medicine. 2019. *MetaMap - A Tool For Recognizing UMLS Concepts in Text.* June 24. Accessed July 1, 2019. https://metamap.nlm.nih.gov/.

White, C Michael. 2019. "A Review of Human Studies Assessing Cannabidiols (CBD) Therapeutic Actions and Potential." *The Journal of Clinical Pharmacology* 59 (7): 923--934.

World Health Organization. 2017. "CANNABIDIOL (CBD) Pre-Review Report: Agenda Item 5.2." *Thirty-ninth Meeting of the Expert Committee on Drug Dependence.* Geneva. 1--27.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

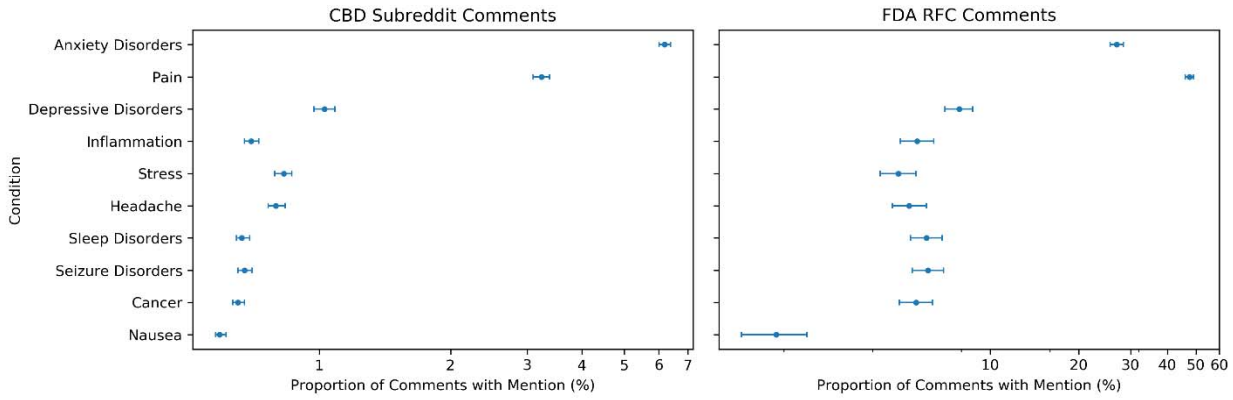

Figure 1. 95% confidence intervals around the proportion of total comments mentioning each condition.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Table 1. Most frequently mentioned conditions in the CBD subreddit. Superscript [(A)] marks conditions discussed in a review of studies regarding CBD therapeutic efficacy (White 2019). Superscript [(B)] marks conditions discussed in a recent survey study of CBD users (Corroon and Phillips 2018).

		Reddit Comments			FDA RFC Responses			
Conditions	**Term Count (Reddit Only)**	**Mentions Count**	**Mentions with CBD Count**	**Mentions with CBD and Treatment Phrase Count**	**Mentions Count**	**Mentions with CBD Count**	**Mentions with CBD and Treatment Phrase Count**	**Example Comments with Mentions including Target Condition, CBD, and Treatment Phrase**
Anxiety Disorders[(A)(B)]	anxiety (3596) anxious (376) panic attacks (323) social anxiety (138) anxiety attack (60) general anxiety (41) GAD (39) panic disorder (33) anxiety/depression (32) anxiety symptoms (15) generalized anxiety disorder (12) anxiousness (9) general anxiety disorder (6)	3968	872	239	1033	326	159	I use **CBD** for **anxiety** and personally feel like I've been getting **better** results with this vape setup than I ever did taking a full spectrum tincture.
Pain [(A)(B)]	pain (2052) chronic pain (189) back pain (127) nerve pain (56) painful (38) breakthrough pain (29) ache (19)	2068	419	115	1823	770	295	I find that for me, vaping **CBD** in the morning and evening **helps** my **back pain** and sciatica.
Depressive Disorders [(B)]	depression (592) depressed (63) anxiety/depression (32) miserable (22) depressing (8) misery (5)	667	110	36	301	59	30	Thank you for the review, I just bought some **CBD** flower and I'm hoping that it will **help** with my **depression** and get me back to be productive.
Inflammation [(B)]	inflammation (307) inflammatory reaction (2)	309	87	31	217	65	28	The **CBD reduces** my **inflammation** so having my inflammation reduced for a period of time sense to make it so my pain doesn't spike as high long-term.
Stress	stress (347) stressful (66) stressed (60) stressing (11) stresses (8)	465	58	16	187	32	14	Because autism seems so complex and I only saw **CBD** being talked about for **stress** I thought it just **lowered** stress but it really reduces the feelings of overstimulation.
Headache [(A)(B)]	headaches (286) migraines (222)	431	78	12	204	65	27	**CBD** has been the best thing yet in **reducing** nicotine cravings and **headaches** I have found.
Sleep Disorders [(A)(B)]	insomnia (206) sleep issues (20) sleep disorders (13) sleep problems (12) sleeping problems (10) problems sleeping (3) sleepless (3) sleeplessness (2)	265	52	12	233	40	21	**CBD took care of** my chronic **insomnia** issues, and I no longer need Zolpimist for sleep.
Seizure Disorders [(A)(B)]	seizures (199) epilepsy (109) epileptic (20)	277	63	9	236	79	37	Also, anecdotally, I'm told by a doctor that gives CBD to epileptic children that full-spectrum is more effective than **CBD** isolate (for **eliminating seizures**, at least).
Cancer [(B)]	cancer (233) cancerous (12)	244	41	9	215	39	9	I know when my cousin was using **CBD** oil to **treat** his **cancer** the doctors and nurses were almost rude to him.
Nausea	nausea (128) nauseous (21) nauseated (7) nauseating (4)	157	24	9	73	6	2	My aunt was diagnosed with CHS seven years ago and aside from hot showers the only thing that actually **relieves** her **nausea** is **CBD**.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted February 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Table 2. Popular modes of consumption based on product group mention frequency in the CBD subreddit.

Product Groups	Terms	Mentions Count	Mentions with CBD Prefix	Mentions with CBD Prefix Count
Oil/Tinctures	oil (6461) tincture (2687) drops (1143) droplets (12)	8609	CBD oil (1701) hemp oil (374) CBD tincture (193) CBD drops (29)	2215
Vapes	vape (3684) juice (955) vape juice (394) dab (360) wax (340) vape pen (278) pods (190) ejuice (180) vape cartridge (124) eliquid (101) vape liquid (40) vape tank (19)	4775	CBD vape (277) CBD vape juice (88) CBD juice (41) CBD ejuice (26) CBD dabs (24) CBD pods (20) CBD eliquid (19) CBD vape pen (19) CBD wax (13) CBD vape cartridges (10) CBD vape liquid (8) hemp pods (6) hemp vape (4)	419
Edibles	edibles (748) gummies (688) coffee (337) candy (249) tea (245) chocolate (122) drinks (96) honey (71) cookies (70) brownies (38) syrup (34) beverage (30) snacks (30) smoothies (16) chews (11)	2429	CBD gummies (81) CBD edibles (56) CBD coffee (20) CBD tea (14) hemp tea (9) hemp edibles (8) CBD drinks (4) CBD honey (4) CBD candy (4) hemp honey (3) hemp gummies (2) CBD snacks (2) CBD beverages (2)	209
Pills/Capsules	capsules (726) pills (445) caps (127) tablets (93) softgels (15)	1270	CBD capsules (92) CBD pills (15) CBD caps (11)	128
Topicals	cream (228) salve (166) balm (163) topicals (115) lotion (102) patches (76) gel (46) ointment (13) topical cream (12) body butter (12)	787	CBD balm (22) CBD cream (17) CBD salve (12)	96