medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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: 1995

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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 August 22, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 its therapeutic potential and lack of legal oversight, despite lacking conclusive scientific evidence for wide spread 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.

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. 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. 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.

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

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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, 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 (Reiley, 2019). Figure 1 presents the billboard sign of a local CBD shop that recently opened in Lexington, Kentucky, clearly indicating a health benefit perspective. 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.

Figure 1: A local CBD shop's billboard in Lexington, KY

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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.

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-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.

Material and Methods

We perform our analysis on a collection of all comments posted to the CBD subreddit between January 1, 2019 and April 30, 2019 totaling to 64,099 individual comments. We validate our method on a secondary collection of 3832 machine-processable comments[1] submitted to FDA's RFC on cannabis derived products between April 3, 2019 and July 19, 2019. 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

[1] There were 4272 comments in total (https://www.regulations.gov/docket?D=FDA-2019-N-1482) by the time the comment period closed on July 16, 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.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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.

concept level. Based on the concepts found, we manually curate an exhaustive list of target conditions (TCs); each TC is associated with a dictionary of terms (i.e., ways in which a condition is expressed) 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 TC as well as CBD and some therapeutic trigger 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 TC,
2. search for mentions of the TC and CBD within the same sentence,
3. search for mentions of the TC, CBD, and some therapeutic trigger phrase indicating treatment where each mention is separated by at most 36 characters (half the observed average sentence length). For example, when analyzing mentions of epilepsy, one regular expression pattern used is “\b(cbd)\b[^\\/\.,;\-\(\)]{1,36}\b(treats)\b[^\\/\.,;\-\(\)]{1,36}\b(epilepsy)\b” which matches occurrences of “cbd”, trigger phrase “treats”, and “epilepsy” in the given order separated by at most 36 non-punctuation characters. We match on several variations of the aforementioned 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 trigger phrases.

Here, therapeutic trigger phrases come from a hand-crafted dictionary 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 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. By manually

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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.

examining 100 randomly sampled 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. As vaping, for example, may not necessarily be associated with CBD, we additionally perform more precise queries to match for mentions of these products that explicit 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 posts, where the second column is a list of terms for the TC 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. 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.

Given the 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; for pain, these percentages are 3.22% and 5.56% respectively. On the other hand, 27.00% of comments to the FDA RFC mention some form of anxiety

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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.

disorder, and at least 15.39% of these comments explicitly mention CBD as a potential therapeutic solution; for pain, these percentages are much more prominent at 47.57% and 16.18% respectively. Overall, it can be observed that Reddit comments tend to focus more on mental conditions (anxiety and stress) while FDA comments tend to focus more on physiological conditions (pain and headache).

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 query hits between the two platforms. This may indicate that comments to the FDA, as expected, are highly focused on CBD products and their perceived therapeutic effects while Reddit comments are more likely to include off-topic discussions.

We use superscript (A) to indicate conditions covered in a recent review of human studies assessing the potential of CBD (White, 2019) and superscript (B) 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 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.

Conclusion and Limitations

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 posts to the Reddit platform and 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

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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.

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. Given the precise nature of our methodology (and rule-based methods in general), our results are in fact underestimates of the true population. However, we posit 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. The latter estimate can only be obtained by a conventional survey targeting a truly random sample of consumers and not just those who post on Reddit. 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. 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.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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

1. World Health Organization. CANNABIDIOL (CBD) Pre-Review Report Agenda Item 5.2. In: Expert Committee on Drug Dependence Thirty-ninth Meeting Geneva, November 6-10, 2017.
2. Cogan PS. On healthcare by popular appeal: critical assessment of benefit and risk in cannabidiol based dietary supplements. Expert Review of Clinical Pharmacology. 2019;12(6):501-511. doi:10.1080/17512433.2019.1612743.
3. Cohen PA, Sharfstein J. The Opportunity of CBD — Reforming the Law. New England Journal of Medicine. 2019;381(4):297-299. doi:10.1056/nejmp1906409.
4. U.S. Food and Drug Administration. FDA approves first drug comprised of an active ingredient derived from marijuana to treat rare, severe forms of epilepsy. *FDA Newsroom*. https://www.fda.gov/news-events/press-announcements/fda-approves-first-drug-comprised-active-ingredient-derived-marijuana-treat-rare-severe-forms/. Published June 26, 2018. Accessed July 1, 2019.
5. Reiley L. CBD-infused food and beverages are still illegal under U.S. law. So why are they everywhere? *The Washington Post*. https://www.washingtonpost.com/business/2019/06/24/cbd-infused-food-beverages-are-still-illegal-under-us-law-so-why-are-they-everywhere. Published June 24, 2019. Accessed July 1, 2019.
6. U.S. National Library of Medicine. MetaMap - A Tool For Recognizing UMLS Concepts in Text. https://metamap.nlm.nih.gov/. Accessed July 1, 2019.
7. Google Code Archive. Word2vec: Tool for computing continuous distributed representations of words. https://code.google.com/archive/p/word2vec/. Published July 29, 2013. Accessed July 1, 2019.
8. White CM. A Review of Human Studies Assessing Cannabidiols (CBD) Therapeutic Actions and Potential. *The Journal of Clinical Pharmacology*. 2019;59(7):923-934. doi:10.1002/jcph.1387.
9. Corroon J, Phillips JA. A Cross-Sectional Study of Cannabidiol Users. *Cannabis and Cannabinoid Research*. 2018;3(1):152-161. doi:10.1089/can.2018.0006.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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 Signal Count**	**Mentions Count**	**Mentions with CBD Count**	**Mentions with CBD and Treatment Signal Count**	**Example Comments with Mentions including Target Condition, CBD, and Treatment Trigger**
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 **control** 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 Ambien 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 many, many times 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 other friend was using **CBD** oil to **treat** her **cancer** the doctors and nurses were almost rude to her.
Nausea	nausea (128) nauseous (21) nauseated (7) nauseating (4)	157	24	9	73	6	2	My older brother was diagnosed with CHS roughly 3-4 years ago and aside from steamy hot showers the only thing that actually **relieves** his **nausea** and vomiting is **CBD**.

medRxiv preprint doi: https://doi.org/10.1101/19004929; this version posted August 22, 2019. 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