

The Twittersphere of Cannabis Use During Pregnancy

Keywords: cannabis, pregnancy, Twitter, topical context, census

Extended Abstract

Problem. Cannabis use during pregnancy is on the rise, and its effects and long-term consequences on fetal development and maternal health are topics of active research [1, 2]. As a result, patients increasingly seek medical advice from social media, but these platforms are not regulated and may contain anecdotal discussions and misinformation [3]. Furthermore, these informal discussions can contribute to normalizing cannabis use during pregnancy, which may lead to more pregnant individuals using cannabis despite the known risks. Thus, health-care professionals must engage in these discussions and provide evidence-based information to counteract misinformation and ensure that pregnant individuals make informed health decisions.

Research gap. Current studies, that investigate how “cannabis use during pregnancy” is being discussed on social media, often rely on small samples with data collected over one year [4]. Additionally, they focus on manual labeling and do not provide explanations to make sense of popularity differences across different regions. Here, we bridge these gaps by collecting and analyzing discussions on Twitter to verify whether real-time online measures can complement official indicators such as the population of drug consumers, females, or people living below the poverty line.

Pre-processing. As shown in Figure 1, we first collect tweets related to “cannabis during pregnancy” (referred to as *CanPreg*) tweeted from 2012-01-01 to 2021-12-31. We rely on the archival search functionality from the Twitter Academic API v2 using 30 cannabis- and 17 pregnancy-related keywords. Then, retweets, duplicated tweets, and clear non-related tweets were removed to maintain original content. Next, we automatized the removal of non-related tweets using S-BERT [5] semantic search. Shortly, we provided five queries (e.g., cannabis during pregnancy, smoking weed while pregnant, the effects of cannabis on newborns) to S-BERT to get a semantic similarity score between each tweet and query. All tweets with a similarity greater than a threshold that we infer for each query by labeling a sample of 100 randomly-selected tweets via crowd-sourcing were selected. Similarly, Botometer [6] was used to remove authors who are likely bot accounts. All accounts with a cap¹ score lower than a threshold that is determined via crowd-sourcing were selected. Finally, a geo-location was assigned to users using the information from their bio-descriptions or the most common location of their CanPreg tweets. A location was assigned to those tweets that do not have a geo-tag by assigning their author’s location. In order to make fair comparisons across regions, we account for sample size effects by normalizing all CanPreg counts by the totals² Note that, for technical reasons the total counts is available from 2015-01-01. Finally, survey and census data³ were collected and each indicator was normalized by the population size of each state reported in 2020.

Analysis. To start our analysis, the volume of tweets across the globe were quantified and its correlation with census and survey data were measured, i.e., drug abuse, deaths, health insurance, global and women population, poverty, and social capital. Second, we inferred the

¹The cap score can be interpreted as the probability that an account with this score or greater is a bot

²All tweets (and unique users who) ever posted within the same time-frames.

³CDC (www.cdc.gov) and Census data (data.census.gov).

tone of these tweets with BERTopic [7] to better understand the topical context in which users discussed this subject. Third, we characterized users talking about “cannabis use during pregnancy” based on the content they often produce, and verified whether their characteristics correlate with their demographics.

Results. Our main findings are summarized as follows. (1) We provided a modular methodology to curate topical tweets and their authors, see Figure 1. Using this methodology, we found 102K valid tweets and 83K valid users of which approximately 58% of them were located across 209 countries. Most tweets were located in the United States followed by the United Kingdom, Canada, Australia, and France. (2) Using BERTopic, we found 274 topics (i.e., clusters of tweets). Figure 2 shows the 8 most popular topics in terms of number of tweets. We see that Topic 0 refers to all CanPreg tweets on “breast feeding”, and Topic 4 to “thc” (Tetrahydrocannabinol, the principal psychoactive component of cannabis) and “fetus”. (3) We found significant positive correlations between Twitter signals and certain official indicators from the Census and survey data in the USA. For instance, the more white people who experienced an abortion we observed an increase in tweets about CanPreg. Conversely, we did not find any significant correlations with social capital, children or elderly populations, those with health insurance, deaths and live births. Due to space limitations, we have omitted the results outside the USA.

Implications. This study demonstrates how cannabis use during pregnancy is portrayed through social media platforms, specifically Twitter. Further, it shows how changes to the legal status influence the volume and tone of posts related to this topic and how the content and origin of supportive posts relate to census and survey data. Finally, we demonstrated the potential signal of social media as complementary information to traditional data collection techniques. Our findings help inform policy strategies to public health agencies, care providers, and other stakeholders. Moreover, they suggest future avenues for research.

References

1. Corsi, D. J. *et al.* Maternal cannabis use in pregnancy and child neurodevelopmental outcomes. *Nature medicine* **26**, 1536–1540 (2020).
2. Richardson, K. A., Hester, A. K. & McLemore, G. L. Prenatal cannabis exposure-The “first hit” to the endocannabinoid system. *Neurotoxicology and Teratology* **58**, 5–14 (2016).
3. Meacham, M. C., Paul, M. J. & Ramo, D. E. 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 (2018).
4. Pang, R. D., Dormanesh, A., Hoang, Y., Chu, M. & Allem, J.-P. Twitter posts about cannabis use during pregnancy and postpartum: a content analysis. *Substance Use & Misuse* **56**, 1074–1077 (2021).
5. Reimers, N. & Gurevych, I. *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (2019), 3982–3992.
6. Sayyadiharikandeh, M., Varol, O., Yang, K.-C., Flammini, A. & Menczer, F. *Detection of novel social bots by ensembles of specialized classifiers* in *Proceedings of the 29th ACM international conference on information & knowledge management* (2020), 2725–2732.

7. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).

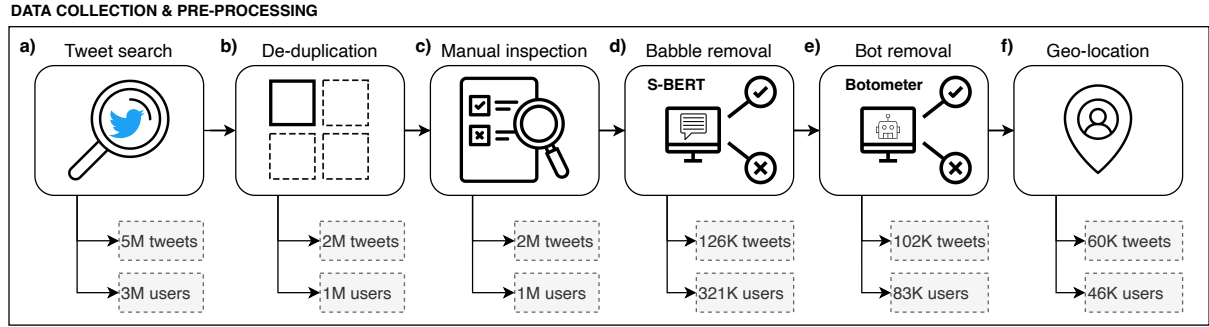


Figure 1: **Pipeline.** (a) English tweets related to “cannabis use during pregnancy” using the Twitter Search Academic API v2 and 47 keywords were collected. (b) Retweets were removed to focus on original generated content. (c) Tweets outside the scope of the topic of interest were manually removed. (d) The detection of unrelated tweets were automatized using S-BERT. (e) The Botometer was used to remove all authors who are likely to represent bot accounts. (f) Lastly, a location was assigned to each tweet with no geo-tag by selecting the most common location of its author. This location can be extracted from the author’s CanPreg tweets, or from the author’s bio. Similarly, to those authors with no location from their bio, we assign the most common location from their tweets.

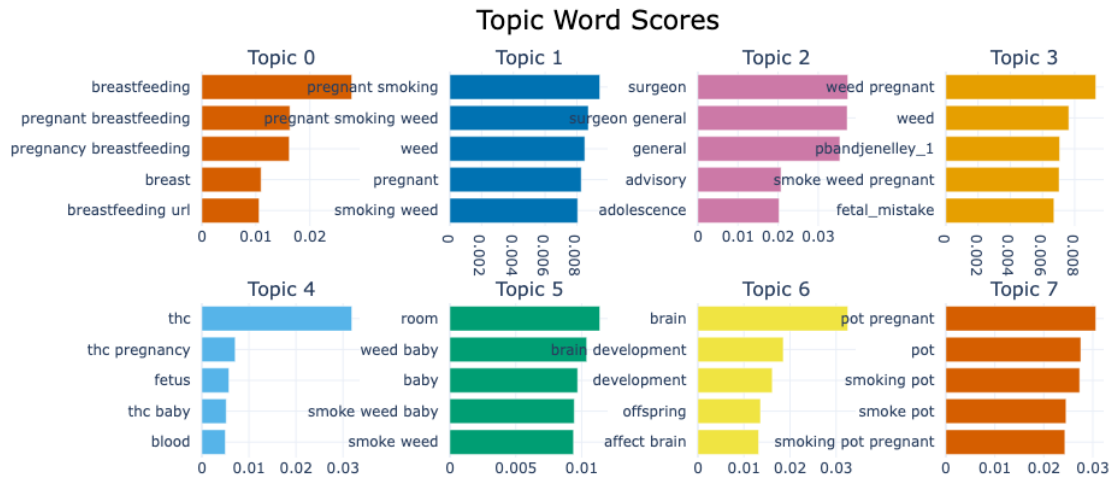


Figure 2: **Topic modeling.** 8 most common topics found by BERTopic, and the density of word-occurrence across tweets of the same cluster (topic). For example, Topic 0 clusters all tweets related to breast-feeding, and Topic 6 clusters all tweets related to brain development.

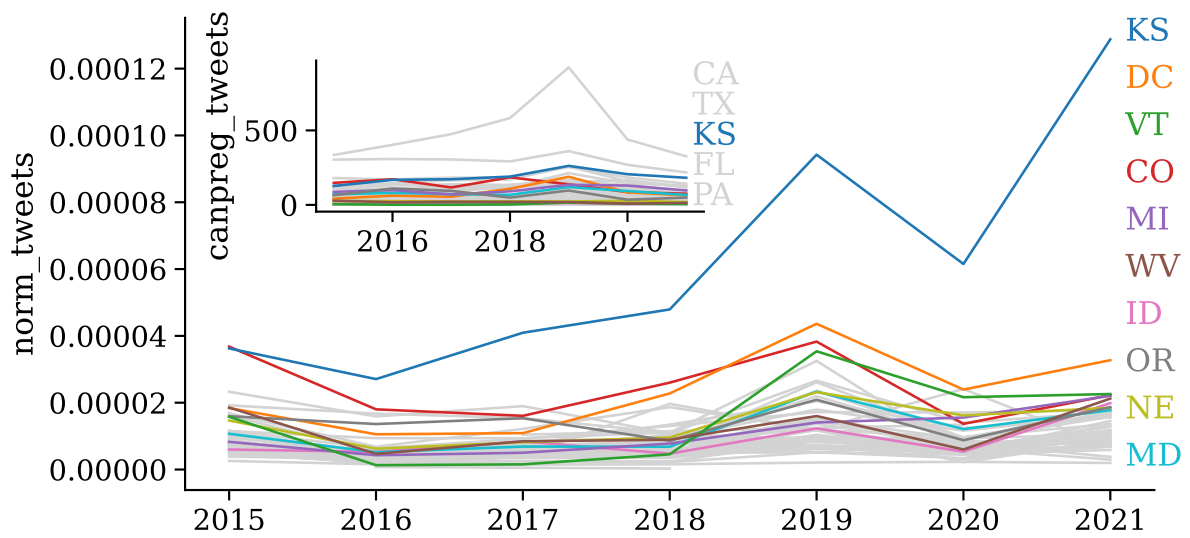


Figure 3: **Number of tweets per year across the most active states in the USA.** Number of CanPreg tweets across the 5 most active states in 2021 (inset). Fraction of CanPreg tweets (normalized by the total number of tweets ever posted in each year), and 10 most active states in 2021: Kansas, Washington DC, Vermont, Colorado, Michigan, West Virginia, Idaho, Oregon, Nebraska, and Maryland.

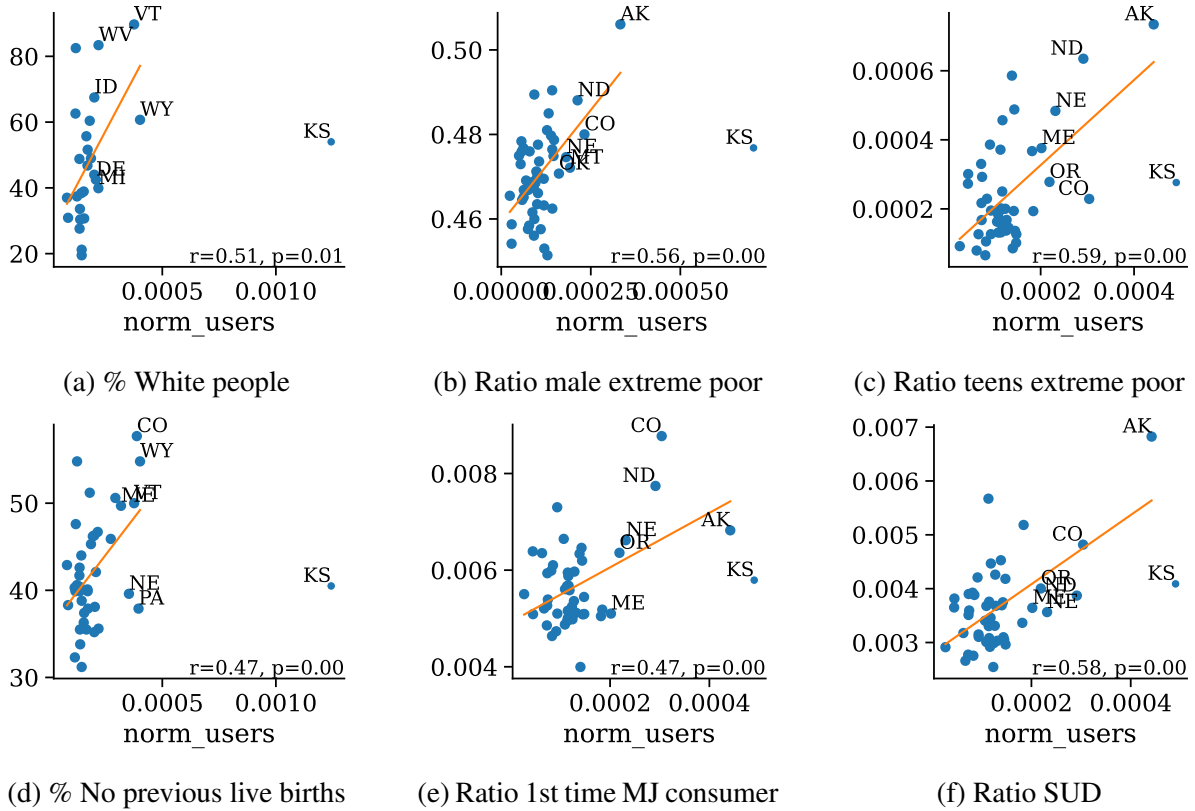


Figure 4: Twitter vs. survey data in the USA. A positive and significant correlation between number of Twitter users and survey indicators in each panel can be observed. The more users twitting about CanPreg, the larger proportion of the following were observed: (a) abortions in the population identifying as white, (b) men below the poverty line, (c) 16-17 year-olds people below the poverty line, (d) people without any previous live births, (e) people who consumed cannabis for the first time in the last year, and (f) people with substance use disorder. Kansas (KS) was considered as an outlier when fitting the linear regression.