



Social Big Data as a Tool for Understanding and Predicting the Impact of Cannabis Legalization

Sean D. Young^{1,2*}, Howard Padwa³ and Erin E. Bonar⁴

¹ University of California Institute for Prediction Technology, Department of Informatics, University of California, Irvine, Irvine, CA, United States, ² Department of Emergency Medicine, University of California, Irvine, Irvine, CA, United States, ³ Integrated Substance Abuse Program, University of California, Los Angeles, Los Angeles, CA, United States, ⁴ Department of Psychiatry, Addiction Center, University of Michigan, Ann Arbor, MI, United States

Keywords: cannabis, social media, big data, drug abuse, prevention

INTRODUCTION

After alcohol and tobacco, cannabis is the most commonly used substance in the United States (U.S.), with 8.9% of individuals ages 12 and older reporting past-month use (1). Cannabis policy is currently undergoing a historic change in the United States. Though cannabis is illegal under federal law, 29 states and the District of Columbia (D.C.) now allow medical use, and since 2012, eight states and D.C. have passed laws permitting non-medical use (2).

In California, voters approved non-medical use of cannabis through a ballot initiative in November 2016. With a population of nearly 40 million, California has a larger population than the other states with legalized non-medical use combined (3). Because of its size, California may offer an unprecedented opportunity to examine the consequences of legalization, and to glean lessons about legalization's impacts that will be essential as policymakers and regulators in other states (and potentially at the federal level) consider cannabis legalization in the future. In this article, we highlight our viewpoint on some of the major lessons about cannabis legalization that can be learned in California, and how big data can be used to generate knowledge about cannabis legalization's impact on public health and safety.

EVALUATING LEGALIZATION'S IMPACT

We believe there are reasons to be both optimistic and concerned about legalization's impact. First, cannabis legalization holds promise as a way to reduce the criminal justice-related costs of prohibition (e.g., incarceration for cannabis-related offenses, enforcement costs), decrease the cannabis black market, facilitate the regulation of cannabis products for purity and safety, generate tax revenues (4), and increase access for individuals who use cannabis without experiencing significant negative consequences or impaired functioning. However, legalization may increase cannabis use, putting more individuals at risk for cannabis use disorders and other mental health, medical, and psychosocial problems associated with the drug (5). Legalization may increase rates of cannabis consumption in ways that increase risk to public health and safety, such as driving while under the influence (6), and overconsumption of edibles or concentrates leading to psychiatric distress that may require emergency care (7, 8). These risks are of particular concern since the potency of cannabis—as measured by THC levels—has been consistently rising over the past two decades (9).

OPEN ACCESS

Edited by:

Aek Palakorn Achananuparp,
Singapore Management
University, Singapore

Reviewed by:

Thomas Lefèvre,
Université Paris 13, France

*Correspondence:

Sean D. Young
syoung5@uci.edu

Specialty section:

This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

Received: 06 April 2018

Accepted: 10 September 2019

Published: 04 October 2019

Citation:

Young SD, Padwa H and Bonar EE
(2019) Social Big Data as a Tool for
Understanding and Predicting the
Impact of Cannabis Legalization.
Front. Public Health 7:274.
doi: 10.3389/fpubh.2019.00274

Given these potential benefits and harms, there has been significant interest in analyzing survey, health system, and administrative data from the first states to legalize non-medical use in order to evaluate legalization's impact. A major focus of this research has been the impact of legalization on rates of cannabis use, the frequency of use, and use by individuals who are at particularly high risk for health or psychosocial problems because of cannabis (10–15). Legalization's impact on negative outcomes related to increased use—such as cannabis-related hospital and emergency department visits, poison control center calls, motor vehicle crashes, and cannabis use disorder treatment episodes—has also received significant attention (11, 15–19). Shifts in cannabis markets, production, sales, and tax revenue are also being monitored (15, 17, 20). The potential benefits of legalization have garnered less attention, though quasi-experimental research suggests that cannabis legalization may be associated with reductions in opioid prescribing and opioid-related deaths (21, 22).

We believe this research is helping develop the knowledge base about critical health and policy questions related to cannabis legalization, but the data sources have limitations. There is a considerable time lag between policy implementation and being able to collect and analyze data on the impact of policy changes, making it difficult to fully understand legalization's consequences in a timely manner. Many measures of cannabis use on existing surveys only ask about prevalence or frequency of use, but do not collect information about amounts of cannabis consumed, the potency of cannabis being used, or what other substances (e.g., alcohol, tobacco) are co-ingested with cannabis (23). Available data typically rely on self-report, which may be imprecise because of difficulty measuring and reporting cannabis consumption in absence of accurate information about potency or “dose.” Information concerning legalization's adverse effects collected from hospitals, poison control centers, law enforcement, and substance use disorder treatment providers often have small sample sizes and short time frames, though these data will become more robust as they are collected in more places and over longer periods of time (24). Analyses of sales data are generating significant insights into how legalization is impacting the price of cannabis and the nature of products being sold, but it remains challenging to evaluate the impact these products have on the individuals purchasing and consuming them (23).

Another issue is that the datasets being used to inform analyses of legalization's impact are designed to measure and evaluate legalization's potential negative outcomes (e.g., increased use among at-risk individuals, hospitalizations, accidents), but are ill equipped to capture information concerning legalization's potential benefits. Though some research has examined the relationship between legalization and reductions in negative outcomes (21), existing datasets do not collect information about the benefits of legalization—such as increased access to a substance that many individuals find pleasurable and can use with little ill effects. To comprehensively evaluate legalization's impact such measures of its benefits will be essential.

Finally, existing methods of collecting data on cannabis trends—such as surveys, medical data, and interviews—are

expensive, rely on cumbersome and time-consuming processes to secure funding, and require significant amounts of time and staff resources. With these traditional methods, researchers may not be able to identify the societal impact of cannabis policy changes for many years. It is imperative that researchers can begin to explore the impacts of legalization prospectively in order to learn about potential benefits and to address drawbacks of the policy and growing concerns before they become significant public health problems.

HOW BIG DATA CAN ENHANCE OUR UNDERSTANDING OF LEGALIZATION'S IMPACT

Methods to analyze “big data” might be leveraged to advance our understanding of cannabis legalization's impact. As big data have penetrated our daily existence, large amounts of medical, environmental, genomic, and public health data, paired with publicly available data from social media and search engines can provide critical information about public health problems. However, new methods are needed that are capable of collecting and analyzing these data. These approaches may be useful in surveillance efforts to help public health and policy researchers understand and predict the implications of changing trends in cannabis policies (25).

For example, one area of big data research that may be particularly helpful involves the study of “social big data,” such as data from social media, wearable devices, and online search data. One in four people worldwide are publically documenting their activities, intentions, moods, and social interactions on social sites each day (26). They are increasingly doing so, generating 400 million “tweets” per day on Twitter (27) and 4.75 billion content items posted each day on Facebook (28). Most of these platforms support user profiles, tagging, time-stamping, and/or geolocation capabilities, making key demographic and contextual information available for analysis. Many platforms also provide data on users' social network connections, allowing access to new and valuable information on how social groups influence attitudes, behaviors, and health (29). For example, HIV researchers have studied the content people display on their social networking profiles and used this information to inform HIV research (30–32). These studies have found that social media posts contain information about people health behaviors, such as their drug use and risk behaviors, and that these conversations can be mapped on a US map and used to help inform public health and surveillance efforts. Similarly, social media data regarding cannabis use has implications for interventions (33–36), such as identifying trends in cannabis use or abuse across counties or states, as well as identifying new ways that cannabis is being used as described in social media.

Much of the available social data exists in an unstructured format, such as free-text social media posts which would require researchers to read and analyze them (37–39). However, computer science/machine learning methods could be used to train machines to learn the patterns identified by human domain experts in detecting whether posts are related to cannabis or not.

This process could lead to instant identification of millions of real-time social media posts about cannabis (40, 41). While sifting through these posts could take researchers years to complete, machines could do so within seconds.

Once posts have been identified as being related to cannabis or not, data could be tagged for location to triangulate location data (i.e., identifying potential hot spots) with actual cannabis outcomes data for use in models attempting to predict cannabis-related outcomes, such as trends in cannabis use, or motor vehicle crashes related to driving under the influence. For example, researchers at the University of California Institute for Prediction Technology (UCIPT) partnered with the California Highway Patrol (CHP) and the crowdsourced app, Waze, to study whether Waze data might predict reported crashes, before these events were reported to the CHP. Waze data on reported incidents were labeled as to whether they were car accidents or potential car accidents and merged with actual CHP data, then plotted on a map of California together. Machine learning models identified whether Waze reports of accidents were reported prior to CHP reported accidents. Results suggested that Waze reported accidents ~3 min faster than CHP reports (42). This provided a model for how social data might be used to monitor events in real-time so that public health officials and first responders could intervene faster than with current methods. Cannabis use researchers might apply these social big data methods similarly to gain an understanding of real-time cannabis-related outcomes before they are even reported.

Machine learning methods can be used not only for monitoring public health issues and trends, but also as a tool for prediction. For example, social media and online search data have been used to predict future outbreaks of syphilis, rates of heart disease, and many other public health events (43–45), which can then inform adjustments to policies or community-level interventions. These methods may be especially helpful for monitoring and predicting events in areas where limited data exist or where it is expensive and takes a long time to gain data on a topic that needs to be understood quickly, such as the implications of cannabis use. Similarly, cannabis-related data triangulated between social media and other mhealth sources (e.g., geolocation, analyses of movement, ambient noise), might be used to predict presence of cannabis use disorders or related co-occurring conditions (e.g., anxiety, major depression), perhaps identifying systems (e.g., school, healthcare), or communities

for treatment resource allocation. In addition, applying social big data to the individual can inform personalized medicine approaches to delivering mhealth interventions to prevent further negative consequences which could be instantaneously initiated through social media. Social big data could also yield insights regarding the nature of non-hazardous cannabis use, perhaps by identifying features of social networks and individuals that can be leveraged to increase safer use of the drug among legal users. Furthermore, machine learning methods could potentially be utilized to analyze social media data concerning the potential benefits of legalization—such as pleasure resulting from enhanced access or increased use—that may be difficult to measure using other surveys and other existing data collection methods.

CONCLUSION

We believe that big data modeling approaches can be extremely useful for researchers and policymakers attempting to learn about the implications of cannabis legalization. If applied in real time to a large population, such as that of California, big data-based modeling approaches could provide an additional tool learn about cannabis reforms' benefits and costs, and inform the development of evidence-based public health and public policy in the age of legalization.

AUTHOR CONTRIBUTIONS

SY conceived of the concept, wrote the section on big data, and reviewed final draft. HP and EB helped provide input to the concept, wrote the introduction and concluding sections, and reviewed manuscripts drafts.

FUNDING

SY has received consulting funds from a company involved in cannabis sales.

ACKNOWLEDGMENTS

We wish to thank the National Institute of Allergy and Infectious Diseases (NIAID) for funding this research (SY, 7R01AI132030 and R56AI125105). This manuscript is the opinion of the authors and does not necessarily reflect the views of the sponsor.

REFERENCES

1. NATIONAL Survey on Drug Use and Health. *Results from the 2016 National Survey on Drug Use and Health: Detailed Tables, SAMHSA, CBHSQ*. (2016). Available online at: <https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.htm#lotsect1pe> (accessed March 21, 2018).
2. Carliner H, Brown QL, Sarvet AL, Hasin DS. Cannabis use, attitudes, and legal status in the U.S.: a review. *Prev Med*. (2017) 104:13–23. doi: 10.1016/j.jypmed.2017.07.008
3. Bureau UC. *Census.gov*. Available online at: <https://www.census.gov/en.html> (accessed March 21, 2018).
4. Hall W, Lynskey M. The challenges in developing a rational cannabis policy. *Curr Opin Psychiatry*. (2009) 22:258–62. doi: 10.1097/YCO.0b013e3283298f36
5. Committee on the Health Effects of Marijuana. *An Evidence Review and Research Agenda, Board on Population Health and Public Health Practice, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. The Health Effects of Cannabis and Cannabinoids: The Current State of Evidence and Recommendations for Research*. Washington, DC: National Academies Press (2017). doi: 10.17226/24625
6. Bondallaz P, Favrat B, Chtioui H, Fornari E, Maeder P, Giroud C. Cannabis and its effects on driving skills. *Forensic Sci Int*. (2016) 268:92–102. doi: 10.1016/j.forsciint.2016.09.007

7. Alzghari SK, Fung V, Rickner SS, Chacko L, Fleming SW. To dab or not to dab: rising concerns regarding the toxicity of cannabis concentrates. *Cureus*. (2017) 9:e1676. doi: 10.7759/cureus.1676
8. MacCoun RJ, Mello MM. Half-baked—the retail promotion of marijuana edibles. *N Engl J Med*. (2015) 372:989–91. doi: 10.1056/NEJMp1416014
9. ElSohly MA, Mehmedic Z, Foster S, Gon C, Chandra S, Church JC. Changes in cannabis potency over the last 2 decades (1995–2014): analysis of current data in the United States. *Biol Psychiatry*. (2016) 79:613–9. doi: 10.1016/j.biopsych.2016.01.004
10. Cerdá M, Wall M, Feng T, Keyes KM, Sarvet A, Schulenberg J, et al. Association of state recreational marijuana laws with adolescent marijuana use. *JAMA Pediatr*. (2017) 171:142–9. doi: 10.1001/jamapediatrics.2016.3624
11. Colorado Department of Public Health & Environment, Monitoring Health Concerns Related to Marijuana in Colorado: 2016. *Changes in Marijuana Use Patterns, Systematic Literature Review, and Possible Marijuana-Related Health Effects*. Denver (2016). Available online at: <https://www.colorado.gov/pacific/cdphe/marijuana-health-report> (accessed March 21, 2018).
12. Kerr DCR, Bae H, Phibbs S, Kern AC. Changes in undergraduates' marijuana, heavy alcohol and cigarette use following legalization of recreational marijuana use in Oregon. *Addiction*. (2017) 112:1992–2001. doi: 10.1111/add.13906
13. Miller AM, Rosenman R, Cowan BW. Recreational marijuana legalization and college student use: early evidence. *SSM Population Health*. (2017) 3:649–57. doi: 10.1016/j.ssmph.2017.08.001
14. Rusby JC, Westling E, Crowley R, Light JM. Legalization of recreational marijuana and community sales policy in Oregon: impact on adolescent willingness and intent to use, parent use, and adolescent use. *Psychol Addict Behav*. (2018) 32:84–92. doi: 10.1037/adb0000327
15. Washington State Office of Financial Management, Forecasting and Research Division. *Monitoring Impacts of Recreational Marijuana Legalization: 2015 Update Report*. Olympia, WA (2015). Available online at: https://www.ofm.wa.gov/sites/default/files/public/legacy/reports/marijuana_impacts_update_2015.pdf (accessed March 21, 2018).
16. Kim HS, Hall KE, Genco EK, Van Dyke M, Barker E, Monte AA. Marijuana tourism and emergency department visits in Colorado. *N Engl J Med*. (2016) 374:797–8. doi: 10.1056/NEJMc1515009
17. Oregon Public Health Division. *Marijuana Report: Marijuana Use, Attitudes and Health Effects in Oregon*. Portland, OR: Oregon Health Authority (2016). <http://www.oregon.gov/oha/ph/PreventionWellness/marijuana/Documents/oha-8509-marijuana-report.pdf> (accessed March 21, 2018).
18. Sokoya M, Eagles J, Okland T, Coughlin D, Dauber H, Greenlee C, Winkler AA. Patterns of facial trauma before and after legalization of marijuana in Denver, Colorado: a joint study between two Denver hospitals. *Am J Emerg Med*. (2017) 36:780–3. doi: 10.1016/j.ajem.2017.10.014
19. Wang GS, Hall K, Vigil D, Banerji S, Monte A, VanDyke M. Marijuana and acute health care contacts in Colorado. *Prev Med*. (2017) 104:24–30. doi: 10.1016/j.ypmed.2017.03.022
20. Smart R, Caulkins JP, Kilmer B, Davenport S, Midgette G. Variation in cannabis potency and prices in a newly legal market: evidence from 30 million cannabis sales in Washington state. *Addiction*. (2017) 112:2167–77. doi: 10.1111/add.13886
21. Livingston MD, Barnett TE, Delcher C, Wagenaar AC. Recreational cannabis legalization and opioid-related deaths in Colorado, 2000–2015. *Am J Public Health*. (2017) 107:1827–9. doi: 10.2105/AJPH.2017.304059
22. Wen H, Hockenberry JM. Association of medical and adult-use marijuana laws with opioid prescribing for medicaid enrollees. *JAMA Intern Med*. (2018) 178:673–9. doi: 10.1001/jamainternmed.2018.1007
23. Kilmer B, Pacula RL. Understanding and learning from the diversification of cannabis supply laws. *Addiction*. (2017) 112:1128–35. doi: 10.1111/add.13623
24. Addiction Technology Transfer Center Network, Advocates for Human Potential, National Council for Behavioral Health. *2017 National Cannabis Summit: Science, Policy, and Best Practices Summary of Proceedings*. Los Angeles, CA: Addiction Technology Transfer Center (2017). Available online at: [http://www.ahpnet.com/AHPNet/media/AHPNetMediaLibrary/News/CannabisProceedings_2017_Final-\(003\)_1.pdf](http://www.ahpnet.com/AHPNet/media/AHPNetMediaLibrary/News/CannabisProceedings_2017_Final-(003)_1.pdf)
25. Levy S, Weitzman ER. Building a learning marijuana surveillance system. *JAMA Pediatrics*. (2016) 170:193. doi: 10.1001/jamapediatrics.2015.3489
26. *Social Networking Reaches Nearly One in Four Around the World - eMarketer*. Available online at: <https://www.emarketer.com/Article/Social-Networking-Reaches-Nearly-One-Four-Around-World/1009976> (accessed April 4, 2018).
27. Tsukayama H. *Twitter Turns 7: Users Send Over 400 Million Tweets per Day*. Washington Post (2013). Available online at: https://www.washingtonpost.com/business/technology/twitter-turns-7-users-send-over-400-million-tweets-per-day/2013/03/21/2925ef60-9222-11e2-bdea-e32ad90da239_story.html (accessed April 4, 2018).
28. Facebook Studio Blog. *Introducing Hashtags on Facebook [Latest News and Updates]*. (2013). Available online at: <https://www.facebook-studio.com/news/item/introducing-hashtags-on-facebook>
29. Young SD, Jaganath D. Online social networking for HIV education and prevention: a mixed-methods analysis. *Sex Transm Dis*. (2013) 40:162–7. doi: 10.1097/OLQ.0b013e318278bd12
30. Moreno MA, Parks MR, Zimmerman FJ, Brito TE, Christakis DA. Display of health risk behaviors on MySpace by adolescents: prevalence and associations. *Arch Pediatr Adolesc Med*. (2009) 163:27–34. doi: 10.1001/archpediatrics.2008.528
31. Young AL, Quan-Haase A. Information revelation and internet privacy concerns on social network sites: a case study of facebook. In: *Proceedings of the Fourth International Conference on Communities and Technologies C&T '09*. New York, NY: ACM (2009). p. 265–74. doi: 10.1145/1556460.1556499
32. Young SD. Behavioral insights on big data: using social media for predicting biomedical outcomes. *Trends Microbiol*. (2014) 22:601–2. doi: 10.1016/j.tim.2014.08.004
33. Cabrera-Nguyen EP, Cavazos-Rehg P, Krauss M, Bierut LJ, Moreno MA. Young adults' exposure to alcohol- and marijuana-related content on Twitter. *J Stud Alcohol Drugs*. (2016) 77:349–53. doi: 10.15288/jsad.2016.77.349
34. Daniulaityte R, Chen L, Lamy FR, Carlson RG, Thirunarayan K, Sheth A. “When ‘bad’ is ‘good’”: identifying personal communication and sentiment in drug-related tweets. *JMIR Public Health Surveill*. (2016) 2:e162. doi: 10.2196/publichealth.6327
35. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. Marijuana-related posts on Instagram. *Prev Sci*. (2016) 17:710–20. doi: 10.1007/s11121-016-0669-9
36. 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. doi: 10.1016/j.drugalcdep.2015.08.020
37. Blei DM, McAuliffe JD. Supervised topic models. In: *Proceedings of the 20th International Conference on Neural Information Processing Systems NIPS'07*. Vancouver, BC: Curran Associates Inc. p. 121–8. Available online at: <http://dl.acm.org/citation.cfm?id=2981562.2981578> (accessed April 4, 2018).
38. Garrett R, Liu S, Young SD. The relationship between social media use and sleep quality among undergraduate students. *Inf Commun Soc*. 21:163–73. doi: 10.1080/1369118X.2016.1266374
39. Liu S, Zhu M, Young SD. Monitoring Freshman College experience through content analysis of tweets: observational study. *JMIR Public Health Surveillance*. (2018) 4:e5. doi: 10.2196/publichealth.7444
40. Young SD, Yu W, Wang W. Toward automating HIV identification: machine learning for rapid identification of HIV-related social media data. *J Acquir Immune Defic Syndr*. (2017) 74(Suppl. 2):S128–31. doi: 10.1097/QAI.0000000000001240
41. Cavazos-Rehg P, Krauss M, Grucza R, Bierut L. Characterizing the followers and tweets of a marijuana-focused twitter handle. *J Med Internet Res*. (2014) 16:e157. doi: 10.2196/jmir.3247

42. Young SD, Wang W, Chakravarthy B. Crowdsourced traffic data as an emerging tool to monitor car crashes. *JAMA Surg.* (2019) 154:777–8. doi: 10.1001/jamasurg.2019.1167
43. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on twitter predicts county-level heart disease mortality. *Psychol Sci.* (2015) 26:159–69. doi: 10.1177/0956797614557867
44. Young SD, Mercer N, Weiss RE, Torrone EA, Aral SO. Using social media as a tool to predict syphilis. *Prevent Med.* (2018) 109:58–61. doi: 10.1016/j.ypmed.2017.12.016
45. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with Web search. *PNAS.* (2010) 107:17486–90. doi: 10.1073/pnas.1005962107

Conflict of Interest: SY has received consulting funds from Grassp, which may have affected experiences and/or views related to this manuscript.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Young, Padwa and Bonar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.