

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Drug and Alcohol Dependence Reports

journal homepage: www.elsevier.com/locate/dadr

Examining stigma relating to substance use and contextual factors in social media discussions

Annie T. Chen^{a,*}, Shana Johnny^b, Mike Conway^{c,d}^a University of Washington School of Medicine, Biomedical Informatics and Medical Education, Box 358047, 850 Republican St., Seattle, WA 98109, United States^b University of Washington School of Nursing, Box 357260, Seattle, WA 98195, United States^c University of Utah, Department of Biomedical Informatics, 421 Wakara Way #140, Salt Lake City, UT 84108, United States^d University of Melbourne, School of Computing and Information Systems, Parkville VIC 3052, Australia

ARTICLE INFO

Keywords:

Stigma
Alcohol
Cannabis
Opioids
Social media
Contextual factors
Affect

ABSTRACT

Background: Stigma associated with substance use can have severe negative consequences for physical and mental health and serve as a barrier to treatment. Yet, research on stigma processes and stigma reduction interventions is limited.

Aim: We use a social media dataset to examine: 1) the nature of stigma-related experience related to substance use; and 2) salient affective and temporal factors in the use of three substances: alcohol, cannabis, and opioids.

Methods: We harvested several years of data pertaining to three substances – alcohol, cannabis, and opioids – from Reddit, a popular social networking platform. For Part I, we selected posts based on stigma-related keywords, performed content analysis, and rendered word clouds to examine the nature of stigma associated with these substances. In Part II, we employed natural language processing in conjunction with hierarchical clustering and visualization to explore temporal and affective factors.

Results: In Part I, internalized stigma was most commonly exhibited. Anticipated and enacted stigma were less common in posts relating to cannabis compared to the other two substances. Work, home, and school were important contexts in which stigma was observed. Part II showed that temporal markers were prominent; post authors shared stories of substance use journeys, and timelines of their experience with quitting and withdrawals. Shame, sadness, anxiety, and fear were common, with shame being more prominent in alcohol-related posts.

Conclusion: Our findings highlight the importance of contextual factors in substance use recovery and stigma reduction and offer directions for future interventions.

1. Introduction

Substance use disorder can have severe consequences. Aside from health consequences, individuals who experience a substance use disorder may also contend with stigma, or being treated unfavorably, being labeled, stereotyped, or treated with prejudice, due to characteristics perceived as negative (Palamar et al., 2011; Witte et al., 2019), which may in turn be internalized and experienced as self-stigma, involving negative thoughts, feelings, and sense of self-worth (Kulesza et al., 2013; Matthews et al., 2017; Wogen and Restrepo, 2020).

Stigma associated with substance use can have severe negative consequences for physical and mental health (Kulesza et al., 2014); affect many aspects of life, including employment, housing, and social relationships (Kulesza et al., 2014; Livingston et al., 2012); and serve as a barrier to obtaining treatment (Hammarlund et al., 2018; Paquette et al., 2018; Yang et al., 2017). Moreover, stigma can interfere with substance

use recovery, as individuals who believe that they are stigmatized or who experienced internalized stigma may have less self-esteem and recovery capital (Ashford et al., 2019; Bozdağ and Çuhadar, 2022).

However, there is limited research on the effect of substance use stigma on stigmatized persons and how to address it (Brown, 2011; Corrigan et al., 2017; Kulesza et al., 2013, 2017; Smith et al., 2016). There is also a need for clearer definitions and improved measurements of stigma constructs.. (Brown, 2011; Kulesza et al., 2013; Smith et al., 2016).

In this study, we explore the potential for social media to augment our understanding of experiences of stigma relating to substance use. Social media, including discussion forums, can be rich sources of information about health-related topics, particularly from patient and lay perspectives (Chen, 2012; Hartzler and Pratt, 2011). The anonymity afforded by discussion forums may be beneficial in the case of stigmatized conditions and behaviors, including substance use (Barratt, 2011;

* Corresponding author at: University of Washington School of Medicine, Biomedical Informatics and Medical Education, Box 358047, 850 Republican St., Seattle, WA 98109.

E-mail addresses: atchen@uw.edu (A.T. Chen), mike.conway@unimelb.edu.au (M. Conway).

<https://doi.org/10.1016/j.dadr.2022.100061>

Received 4 December 2021; Received in revised form 2 April 2022; Accepted 29 April 2022

2772-7246/© 2022 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

De Choudhury and De, 2014), and previous research has employed social media to understand the experiences of those affected by stigmatized conditions (Allen et al., 2016). For example, qualitative research of alcohol-related discussion forums has reported three main themes: sharing, supporting, and sobriety (Coulson, 2014). A qualitative study of online discussion content on buprenorphine/naloxone examined facilitators of self-management (Brown and Altice, 2014).

Other studies of content relating to substance use in social media have focused on automatically characterizing substance use patterns in larger datasets. For example, extant research has employed natural language processing methods to predict addiction phase, relapse, and recovery (MacLean et al., 2015), and transitions from casual drug discussion forums to recovery forums (Lu et al., 2019). Extant literature has also employed text and behavior in the form of likes, status update topics, and word usage to predict substance use (Ding et al., 2017, 2018).

However, to our knowledge, research has yet to explore the experiences of stigma relating to substance use in social media. In this work, we examine substance use discussions on Reddit, a social networking platform with a large user base that features discussion forums, or subreddits, in which users discuss a variety of topics. Reddit has been employed in other research relating to substance use (e.g., Brett et al., 2019; Chancellor et al., 2019; Chen et al., 2015), and is currently the 14th most popular website in the world in terms of traffic (Alexa Internet, Inc., 2022).

Our work explores two primary research questions: 1) what is the nature of the stigma experienced by individuals who use substances; and 2) what affective and temporal factors are salient in the use of three substances: alcohol, cannabis, and opioids? Public perception can differ for different substances (Cunningham et al., 1993; Johnson-Kwochka et al., 2021). For example, greater stigma is associated with heroin as opposed to cannabis use (Brown, 2015). The stigma associated with substance use can also vary due to perceived dangerousness, social acceptance, and legality (Janulis et al., 2013). Consideration of differences in the nature of stigma and its effects is imperative in developing effective interventions.

2. Materials and methods

2.1. Harvesting the data, data selection and pre-processing

We downloaded content from Reddit authored between January 1, 2013 and December 31, 2019 using the pushshift.io API (Baumgartner et al., 2020). We included subreddits dedicated to use and recovery of the substances of interest.

2.2. Part I. Annotation of stigma constructs in social media data

In Part I, we aimed to characterize the experiences of persons relating to substance use stigma, whether or not they were in recovery. We downloaded content from subreddits relating to one of three substances: alcohol, cannabis, and opioids. As we were interested in experiences of stigma whether or not Redditors were seeking to quit substance use, we included both subreddits that were recovery-focused, and those that were not. We sampled posts from this dataset to identify posts with a higher probability of stigma-related content. The keywords included both adjectives that might connote stigma, such as “shame,” “untrustworthy,” and “disappoint,” labels such as “crackhead,” “junkie,” and “alcoholic,” and persons who might be involved in stigma-related experiences, such as significant others, parents, and co-workers (see Supplementary Material for details). This keyword list was derived from extant literature and iteratively refined. We focused only on the initiating posts in each subreddit, as these posts often contain richer descriptions of Redditors’ experiences (MacLean et al., 2015).

We performed a directed content analysis of posts iteratively sampled based on pre-determined categories of interest (Hsieh and Shannon, 2005). A coding guide was developed based on extant literature and refined through successive rounds of annotation to account for how

stigma manifests in social media data. We identified whether a post mentioned one or more of the substances of interest (alcohol, cannabis, opioids), and three types of stigma based on the Stigma Framework (Earnshaw et al., 2013; Livingston et al., 2012). Enacted stigma describes experiences of stereotyping, prejudice, and discrimination due to a stigmatized attribute (Smith et al., 2016). Anticipated stigma is the expectation that one might experience stereotyping, prejudice, and discrimination due to a stigmatized attribute. Internalized stigma is the application of negative feelings and beliefs about a stigmatized group to oneself.

The three authors independently annotated constructs of interest in the sampled data, then discussed and reconciled differences either by consensus or majority vote. Inter-annotator agreement was measured using the F-measure, a measure used as a surrogate for Cohen’s Kappa (Cohen, 1960; Hripcsak and Rothschild, 2005). The overall pairwise F-measure on the iterations, prior to reconciliation, varied between 0.67 and 0.71, indicating substantial agreement among annotators (Viera et al., 2005). Inter-annotator agreement per iteration was calculated using eHost, the annotation tool that we used (South et al., 2012). As it may be possible to identify users based on direct quotes, in the presentation of our findings, to protect posters’ identities, we present synthetic quotations based on the annotated data (Moreno et al., 2013).

We rendered word clouds using the Python *wordcloud* package to facilitate comparison of the stigma types in the context of different substances. The size of words was rendered based on term frequency – inverse document frequency (TF-IDF) to accentuate distinguishing factors in the corpus (Soucy, 2005). To reduce obvious connections such as “drinking,” “smoking,” and “weed,” and render more nuanced depictions of factors relating to the specific context, we retained words that were present in no more than 45% of all documents for a given substance and stigma type. The cutoff of 45% was determined empirically by testing different proportions to find a proportion that resulted in word clouds depicting non-obvious and substantively pertinent themes in the data.

2.3. Part II. Topic modeling and visualization of temporal and affective factors in topics

In Part II, we explored temporal and affective factors relating to substance use stigma using a two-step process. We first performed topic modeling to identify main topics of conversation relating to each of the three substances, and then we performed a hierarchical cluster analysis of these topics based on temporal and affective factors. The resulting heatmap visualization enabled us to characterize similarities and differences between topics in terms of these dimensions. This part of the study primarily drew from recovery-focused subreddits so temporal and affective factors experienced by persons working on recovery could be explored.

As a starting point for this analysis, we identified a subset of posts based on keywords potentially associated with stigma. The keywords that we used are available in the Supplemental Material. We pre-processed the data by stripping HTML tags, lowercasing, and tokenizing the text using spaCy (<https://spacy.io/>).

We identified common topics in the posts using topic modeling. There are various methods for topic modeling, including Latent Dirichlet Allocation and non-negative matrix factorization (NMF). We employed NMF because the desired outcome of our first step was that each document (a Reddit post) be assigned to a single topic which is essentially document clustering, a task in which NMF performs well (Casalino et al., 2018; Sharaff and Nagwani, 2016). We performed NMF using Ho’s implementation of NMF (Ho, 2018) using the scikit-learn package (Pedregosa et al., 2011).

As part of the topic modeling process, it is necessary to decide the number of topics. Our ultimate goal was to employ a heatmap visualization to examine the expression of affective and temporal factors in multiple topics and with different substances. We set out to select: 1)

Table 1
Temporal keywords and sample affective keyword categories.

Category	Sample Keywords
Time	year, month, week, day, morning, afternoon, evening, night
Shame	ashamed, upset, embarrass, self-conscious, guilt, humiliate
Sadness	sad, unhappy, despondent, miserable, depressing, sorrow, grief
Anxiety	worry, edgy, impatient, anxious, insecure, upset, concern, disquieted
Negative-Fear	fearful, trepidation, frighten, scare, intimidated, afraid, apprehensive
Loneliness	lonely, alone, isolated

a number that would be consistent irrespective of substance; and 2) a number large enough to demonstrate some diversity of topics, but also small enough to facilitate examination of broad trends in the topics. We examined solutions for varying numbers of topics (2–7) and decided that employing four topics worked optimally based on these criteria. We labeled resulting topics manually by reviewing the keywords and posts assigned to the topic given that the nuanced nature of the topics may make it difficult to develop suitable names from keywords alone. Each label was intended to capture the core concepts that were embodied in the posts and keywords, while avoiding less central concepts.

We estimated the prevalence of temporal and affective factors in the text based on the proportion of keywords from WordNet-Affect (Strapparava and Valitutti, 2004) (Table 1). For each topic, we calculated the proportion of posts assigned to that topic containing keywords belonging to each affective and temporal factor. The selection of affect categories was based on extant literature relating to emotion, substance use, and stigma. For example, shame is common in the experience of internalized stigma (Brown-Johnson et al., 2015), anxiety and depression are correlates of self-stigma (Brown et al., 2015), and categories such as forgiveness, optimism, and confusion were intended to capture cognitive processes that might occur with stigma reduction. With one exception, the keywords for each affect type were drawn from WordNet-Affect (Strapparava and Valitutti, 2004). As loneliness was an important construct to include but not present in the lexicon, we created a keyword list for this category in consultation with prior work on loneliness in social media (Guntuku et al., 2019). The keywords for time were based on a manual examination of our corpus and prior work (Chen et al., 2015).

The next step of our analysis involved hierarchical clustering, an exploratory data analysis technique that facilitates the discovery of groups of similar objects, which is often used to visualize biomedical data (Škuta et al., 2014). The result of the clustered data is often displayed as a matrix visualization, or clustered heatmap, in which the colors of each cell correspond to its value, and a dendrogram, or tree structure, showing the relationships between the clusters. Rows/columns that are more similar to one another are ordered closer to each other. In this study, we clustered the data by *topic* and *contextual factor* (either an affective or temporal factor), using the estimates of affective and temporal keywords for each topic. The resulting heatmap visualization thus illustrates the salience of affective and temporal factors in the data, similarities of topics to one another, and the similarities of contextual factors to one another. We employed the matplotlib and seaborn packages to perform hierarchical clustering and render a heatmap clustered on both rows and columns, using average linkage as the linkage method, and Euclidean distance as the similarity metric.

3. Results

3.1. Part I. Characterization of stigma in social media data

We identified subreddits focused on the substances of interest: r/cripplingalcoholism, r/stopdrinking, r/trees, r/Petioles, r/leaves, r/opiates, and r/OpiatesRecovery (see Supplementary Material for more detail). We sampled the posts using stigma-related keywords, and anno-

Table 2
Proportion of stigma types in annotated social media data.

Stigma Types	Alcohol (n = 705)	Cannabis (n = 572)	Opioids (n = 292)
	n /%	n /%	n /%
Internalized Stigma	335 (48%)	215 (38%)	90 (31%)
Anticipated Stigma	135 (19%)	72 (13%)	61 (21%)
Enacted Stigma	137 (19%)	61 (11%)	54 (19%)

%. Percentage of posts for that substance that contain the construct (e.g. 48% of posts pertaining to alcohol were annotated with internalized stigma).

tated stigma constructs in these posts. The proportion of each stigma type for each substance is depicted in Table 2. Some posts mentioned more than one substance. Since the sample for each substance was extracted using keywords, the proportions indicated in Table 2 do not reflect proportions in the corpus as a whole, and are merely provided to characterize the body of content that we drew upon for content analysis. We describe the stigma types expressed in this data using synthetic quotes.

Internalized stigma was most frequently expressed. People often reflected on their past with despair: “How do I forgive myself for all of the terrible things that I’ve done while drunk?” They struggled with shame, worthlessness, and a desire to become better: “I want to become a person I’m not ashamed of being,” and wrote of being a failure and ruining their lives: “I’ve given up everything in my life for heroin.”

Anticipated stigma often manifests in the form of people being concerned about how others might view them, “I’m constantly worrying what people think of me because of my drug use,” or consequences that they might suffer at work: “No one would guess that I’m the addict that I am. If I were to be found out, I would lose my license.” They might also hide their behavior: “I kept my drinking a secret from everyone because I didn’t want to disappoint them.” Anticipated stigma was more common in opioid-related posts as compared to the other two substances.

Common actors in enacted stigma were friends, family, and the post author, with the post author most often being the person stigmatized: “My best friend told me everyone was talking about me drinking and making a fool of myself.” People often spoke of lost and strained relationships: “I’ve lost so many friends and family due to my drug use.” There were also examples of the post author exhibiting enacted stigma: “I realized how lazy and stupid my stoner friends were.” Levels of enacted stigma were lower in the posts with cannabis.

We rendered word clouds to compare the nature of the different types of stigma in the use of the different substances (Fig. 1). People (e.g., family, parents, and friends) feature prominently, as do contexts such as home, school, and work. Temporal words such as “day,” “night,” and “years,” words used by posters to explain their substance use over time, were more salient with alcohol than the other substances. Alcohol use is also featured in the posts about cannabis (note that words such as “weed” and “smoking” and other high-frequency keywords were excluded from the word clouds). Shame and anxiety feature more prominently in the cannabis-related posts, and the opioid-related posts had a stronger focus on particular substances and the physical aspects of recovery, such as relapse and withdrawal.

3.2. Part II. Temporal and affective factors in recovery-focused topics

3.2.1. Sample characteristics

In Part II, we identified a subset of posts using a subreddit and keyword sampling procedure similar to Part I, but we narrowed the subreddits of focus so the posts would be more recovery-focused (Table 3). There was substantial variability in post length, as some posts were short statements, e.g., “relapsed after x days,” while others were comprised of narratives of a person’s experiences.

3.2.2. Topic modeling

We performed topic modeling of the posts using non-negative matrix factorization. Table 4 shows each cluster’s main topic and keywords;

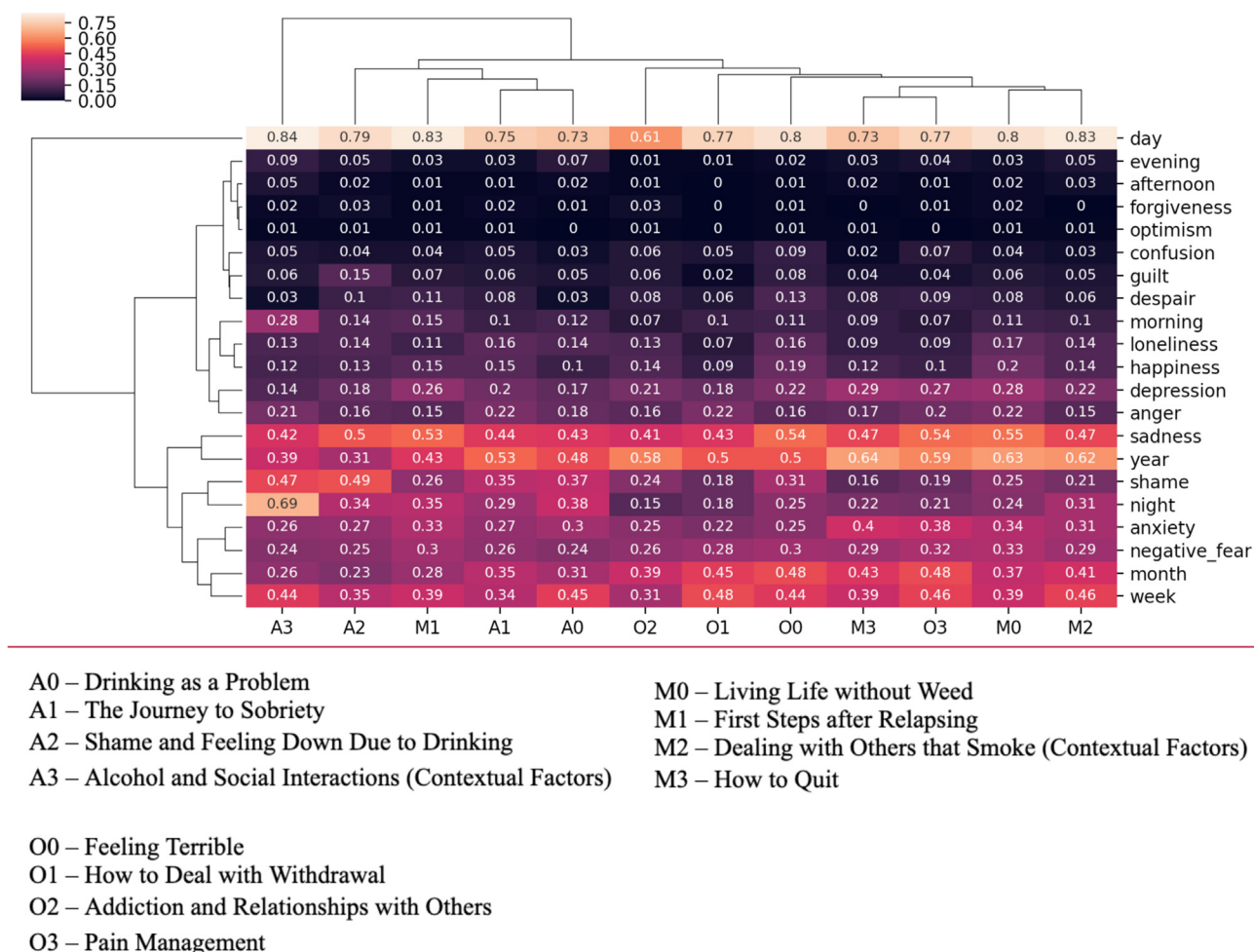


Fig. 2. Clustered heatmap of topics vs. contextual factors.

The first opioid-related topic was the negative feelings associated with opioid use. In particular, post authors often referred to themselves as “losers” and “failures.” The next topic addressed the challenges of coping with withdrawal symptoms. The third topic had to do with addiction, recovery, and dealing with others in the recovery process. Last, one topic focused on dealing with chronic pain.

Overall, there were similarities among the topics across the three substances. The posts often reflect on the past and the journey to stop using. Social interactions were an important part of substance use recovery and stigma reduction. The cannabis- and opioid-related content seem to include more concrete content on how to quit in terms of timeline and phases of withdrawal.

3.2.3. Examining contextual factors and topic clusters

We performed hierarchical clustering to examine how contextual factors and topics were related and rendered a clustered heatmap visualization (Fig. 2). The cells in the heatmap are colored based on the proportion of the factor with the sample on a scale from 0 to 1, with darker cells representing lower proportions, and lighter cells representing higher proportions.

Though there are similarities in the experiences of any given substance, and hence they are often positioned closer to one another, there are also similarities across substances (Fig. 1). For example, How to Quit (M3) and Pain Management (O3) focus on concrete aspects of quitting, and in these posts, asking for and providing advice on quitting was salient. Depression and anxiety in the context of withdrawal were also commonly discussed. Posts relating to The Journey to Sobriety (A1) and First Steps after Relapsing (M1) tended to contain temporal markers due

to the recounting of people’s stories of their attempts at recovery, with some being successful (at the time of writing) and others not.

We observe three main branches in the dendrogram regarding temporal and affective factors. The word “day” is common across all clusters, likely due to multiple reasons: not only is “day” a standard unit of time, but it is also a way for people to quantify their recovery. For example, they might say that they are on “day X” or “X days clean.” The rest of the temporal and affective factors formed two branches, with the bottom set of factors representing negative affect (sadness, anxiety, fear, and shame) and temporal factors (night, year, month, and week), suggesting the salience of temporal elements in these topics. The darker colors of the middle branch of the dendrogram show that the corresponding factors were not as salient in these data.

The heatmap visualization also suggests relationships between the topics in terms of similarity and temporal and affective involvement. Overall, we can see that shame appears more frequently with topics originating from the alcohol-related subreddits. The keyword “night” was extremely frequent with the “Alcohol and Social Interactions” topic, and moderately frequent in the other alcohol and some cannabis-related topics. Sadness was present in over 40% of the posts, irrespective of the topic. Sadness, anxiety, and fear also occurred more frequently than other affect types.

4. Discussion

The two parts of our study provide complementary insights concerning the experiences of substance use stigma. Part I showed that, internalized stigma was most often expressed among the three stigma types, as

post authors often share what they have learned or are struggling with. Anticipated and enacted stigma were also present, albeit to a lesser degree. Qualitative analysis and word clouds revealed a variety of actors, including family, friends, and others, as well as different contexts of use, including work, home, and school.

Part II of our study examined temporal and affective factors in recovery-focused content. Temporal markers were common, as post authors often reflect on their journey or provided a timeline of their experience of quitting/withdrawal. These experiences may serve as a reference point to individuals trying to quit, signposting potential barriers and high-risk situations that these individuals may encounter. Consistent with the greater presence of internalized stigma in Part 1, shame was more common in the alcohol-related posts, and negative emotions such as sadness, anxiety, and fear were also common. As emotion regulation can be an important mediator in the effect of stigma on mental health (Burton et al., 2018), it is important to address emotion regulation in stigma reduction interventions. Overall, there was convergence in the two parts of the study regarding the importance of interactions with others, affective and temporal factors, and challenges of withdrawal with opioid use.

Our work has implications for developing stigma reduction interventions for stigmatized persons. First, stigma reduction interventions have more often focused on negative beliefs held by the public or medical students and other professional groups (structural stigma), instead of challenging self-stigmatizing beliefs (Dickstein et al., 2010; Livingston et al., 2012). The results of this study show that internalized stigma, and stigma involving others (anticipated and enacted stigma), are discussed on Reddit. Insight concerning these experiences can help us better understand the stigma associated with substance use and inform the design of stigma reduction interventions.

Moreover, the findings from Part I showed that diverse contexts, including work, home, and school, were represented; these may be focal points to consider for stigma reduction interventions. Research has also observed that social, situational, and contextual factors such as peer substance use, parental support, and school characteristics affect substance use patterns (Andrade, 2013; Pilatti et al., 2020; Wolf et al., 2019), underlining the need for targeted interventions to incorporate contextual factors.

This study also identified an important function served by these discussion forums – to enable Redditors to ask questions about dealing with stigma or barriers to quitting substances, such as having friends who also use substances. Stigma relating to substance use poses a significant challenge to those recovering from substance use disorder; discussion forums and other social networks such as Alcoholics Anonymous can provide an important support role for those in need. However, post authors expressed concerns regarding the fit of Alcoholics Anonymous for their needs, such as the religious aspect of the group. Issues have also been raised about online support groups (Malik and Coulson, 2010). Rather than suggesting that any single modality suits all persons in need, we suggest that there is a need to develop alternative support mechanisms to complement existing face-to-face and online support mechanisms.

These results also suggest directions for information presentation. First, in online discussion forums, readers might be offered the possibility of focusing on posts with one of the three stigma types, or even anticipated or enacted stigma having to do with a specific actor, such as friends, family, or healthcare providers. Doing so could reduce the amount of content that readers need to sift through to find content of interest. Alternatively, visual exploratory systems displaying aggregated timelines of people's experiences of substance use recovery can assist persons engaged/considering recovery to better understand challenges that they face and strategies to meet these challenges.

4.1. Limitations and future directions

There are various limitations of this work. First, as we employed a keyword list to sample posts, the topics that we extracted do not neces-

sarily reflect the main themes of the subreddits. Second, in this study, we focused on temporal and affective factors; there may be other contextual factors that play important roles in substance use and stigma relating to substance use that we did not examine. For example, future work might consider the types of stigma that are expressed in work, school, and domestic situations. Lastly, in this study we focused on stigma relating to three substances that we expected to exhibit different temporal and affective associations, due to variation in their legality and public perception. Exploration of the stigma associated with other substances is a potentially important topic for future research.

5. Conclusions

This study analyzed narrative data on substance use collected from Reddit, a popular social networking platform. Our research characterized stigma exhibited in these subreddits and analyzed differences in the emphases of the stigma associated with three substances. Then we employed topic modeling and hierarchical clustering to identify themes, and explore similarities and dissimilarities in the temporal and affective experience of using these substances. This work contributes to extant knowledge in two primary ways. First, this study expands our knowledge of differences in the subjective experience of the stigma associated with different substances and provides recommendations for targeted stigma reduction interventions. Second, we demonstrate a novel method of exploring contextual factors that play a role in substance use and stigma relating to it.

Contributors

Annie T. Chen contributed to the conceptualization, corpus annotation, data analysis portions of the study, and writing and revision of article reporting the findings of this study. Shana Johnny contributed to corpus annotation and provided feedback on the article. Mike Conway contributed to the conceptualization and corpus annotation parts of the research, and provided feedback on multiple versions of the manuscript. All authors have approved the final copy of this manuscript.

Declarations of interest

None.

Role of Funding Source

Nothing declared.

Acknowledgments

We would like to thank Mengke Hu and Tian Xu for their assistance with earlier phases of this research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dadr.2022.100061.

References

- Alexa Internet, Inc, 2022. Alexa. <https://www.alexa.com/topsites>.
- Allen, C., Vassilev, I., Kennedy, A., Rogers, A, 2016. Long-term condition self-management support in online communities: a meta-synthesis of qualitative papers. *J. Med. Internet Res.* 18 (3), e61. doi:10.2196/jmir.5260.
- Andrade, F.H, 2013. Can the school context moderate the protective effect of parental support on adolescents' alcohol trajectories in urban Chicago? *Drug Alcohol Depend.* 133 (2), 330–337. doi:10.1016/j.drugalcdep.2013.06.010.
- Ashford, R.D., Brown, A.M., Canode, B., McDaniel, J., Curtis, B, 2019. A mixed-methods exploration of the role and impact of stigma and advocacy on substance use disorder recovery. *Alcohol Treat. Q.* 37 (4), 462–480. doi:10.1080/07347324.2019.1585216.
- Barratt, M.J, 2011. Discussing illicit drugs in public internet forums: visibility, stigma, and pseudonymity. In: Proceedings of the 5th International Conference on Communities and Technologies, pp. 159–168. doi:10.1145/2103354.2103376.

- Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., Blackburn, J., 2020. The pushshift reddit dataset. *ArXiv 2001.08435 [Cs]*. <http://arxiv.org/abs/2001.08435>.
- Bozdağ, N., Çuhadar, D., 2022. Internalized stigma, self-efficacy and treatment motivation in patients with substance use disorders. *J. Subst. Use* 27 (2), 174–180. doi:10.1080/14659891.2021.1916846.
- Brett, E.I., Stevens, E.M., Wagener, T.L., Leavens, E.L.S., Morgan, T.L., Cotton, W.D., Hébert, E.T., 2019. A content analysis of JUUL discussions on social media: using Reddit to understand patterns and perceptions of JUUL use. *Drug Alcohol Depend.* 194, 358–362. doi:10.1016/j.drugalcdep.2018.10.014.
- Brown, S.A., 2011. Standardized measures for substance use stigma. *Drug Alcohol Depend.* 116 (1), 137–141. doi:10.1016/j.drugalcdep.2010.12.005.
- Brown, S.A., 2015. Stigma towards marijuana users and heroin users. *J. Psychoact. Drugs* 47 (3), 213–220. doi:10.1080/02791072.2015.1056891.
- Brown, S., Kramer, K., Lewno, B., Dumas, L., Sacchetti, G., Powell, E., 2015. Correlates of self-stigma among individuals with substance use problems. *Int. J. Ment. Health Addict.* 13 (6), 687–698. doi:10.1007/s11469-015-9559-9.
- Brown, S.-E., Altice, F.L., 2014. Self-management of buprenorphine/naloxone among online discussion board users. *Subst. Use Misuse* 49 (8), 1017–1024. doi:10.3109/10826084.2014.888449.
- Brown-Johnson, C.G., Cataldo PhD, J.K., Orozco, N., Lisha, N.E., Hickman, N., Prochaska, J.J., 2015. Validity and reliability of the internalized stigma of smoking inventory: an exploration of shame, isolation, and discrimination in smokers with mental health diagnoses. *Amer. J. Addict./Amer. Acad. Psychiatr. Alcohol. Addict.* 24 (5), 410–418. doi:10.1111/ajad.12215.
- Burton, C.L., Wang, K., Pachankis, J.E., 2018. Does getting stigma under the skin make it thinner? Emotion regulation as a stress-contingent mediator of stigma and mental health. *Clinic. Psychol. Sci.* 6 (4), 590–600. doi:10.1177/2167702618755321.
- Casalino, G., Castiello, C., Del Buono, N., Mencar, C., 2018. A framework for intelligent Twitter data analysis with non-negative matrix factorization. *Int. J. Web Inf. Syst.* 14 (3), 334–356. doi:10.1108/IJWIS-11-2017-0081.
- Chancellor, S., Nitzburg, G., Hu, A., Zampieri, F., De Choudhury, M., 2019. Discovering alternative treatments for opioid use recovery using social media. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19, pp. 1–15. doi:10.1145/3290605.3300354.
- Chen, A.T., 2012. Exploring online support spaces: using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups. *Patient Educ. Couns.* 87 (2), 250–257. doi:10.1016/j.pec.2011.08.017.
- Chen, A.T., Zhu, S.-H., Conway, M., 2015. What online communities can tell us about electronic cigarettes and hookah use: a study using text mining and visualization techniques. *J. Med. Internet Res.* 17 (9), e220. doi:10.2196/jmir.4517.
- Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas* 20 (1), 37–46. doi:10.1177/001316446002000104.
- Corrigan, P., Schomerus, G., Shuman, V., Kraus, D., Perlick, D., Harnish, A., Kulesza, M., Kane-Willis, K., Qin, S., Smelson, D., 2017. Developing a research agenda for understanding the stigma of addictions Part I: lessons from the Mental Health Stigma Literature. *Am. J. Addict.* 26 (1), 59–66. doi:10.1111/ajad.12458.
- Coulson, N.S., 2014. Sharing, supporting and sobriety: a qualitative analysis of messages posted to alcohol-related online discussion forums in the United Kingdom. *J. Subst. Use* 19 (1–2), 176–180. doi:10.3109/14659891.2013.765516.
- Cunningham, J.A., Sobell, L.C., Chow, V.M., 1993. What's in a label? The effects of substance types and labels on treatment considerations and stigma. *J. Stud. Alcohol* 54 (6), 693–699. doi:10.15288/jsa.1993.54.693.
- De Choudhury, M., De, S., 2014. Mental health discourse on reddit: self-disclosure, social support, and anonymity. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, p. 10.
- Dickstein, B.D., Vogt, D.S., Handa, S., Litz, B.T., 2010. Targeting self-stigma in returning military personnel and veterans: a review of intervention strategies. *Milit. Psychol.* 22 (2), 224–236. doi:10.1080/08995600903417399.
- Ding, T., Bickel, W.K., Pan, S., 2017. Multi-view unsupervised user feature embedding for social media-based substance use prediction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2275–2284. doi:10.18653/v1/D17-1241.
- Ding, T., Hasan, F., Bickel, W.K., Pan, S., 2018. Interpreting social media-based substance use prediction models with knowledge distillation. In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 623–630. doi:10.1109/ICTAI.2018.00100.
- Earnshaw, V.A., Smith, L.R., Chaudoir, S.R., Amico, K.R., Copenhaver, M.M., 2013. HIV Stigma mechanisms and well-being among PLWH: a test of the HIV stigma framework. *AIDS Behav.* 17 (5), 1785–1795. doi:10.1007/s10461-013-0437-9.
- Guntuku, S.C., Schneider, R., Pelullo, A., Young, J., Wong, V., Ungar, L., Polsky, D., Volpp, K.G., Merchant, R., 2019. Studying expressions of loneliness in individuals using Twitter: an observational study. *BMJ Open* 9 (11), e030355. doi:10.1136/bmjopen-2019-030355.
- Hammarlund, R., Crapanzano, K., Luce, L., Mulligan, L., Ward, K., 2018. Review of the effects of self-stigma and perceived social stigma on the treatment-seeking decisions of individuals with drug- and alcohol-use disorders. *Subst. Abuse Rehabil.* 9, 115–136. doi:10.2147/SAR.S183256.
- Hartzler, A., Pratt, W., 2011. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. *J. Med. Internet Res.* 13 (3). doi:10.2196/jmir.1728.
- Ho, G., 2018. Understanding hate speech on Reddit through text clustering. [Eigenfoo https://eigenfoo.xyz/reddit-clusters/](https://eigenfoo.xyz/reddit-clusters/).
- Hripcsak, G., Rothschild, A.S., 2005. Agreement, the F-measure, and reliability in information retrieval. *J. Amer. Med. Inform. Assoc.* 12 (3), 296–298. doi:10.1197/jamia.M1733.
- Hsieh, H.-F., Shannon, S.E., 2005. Three approaches to qualitative content analysis. *Qual. Health Res.* 15 (9), 1277–1288. doi:10.1177/1049732305276687.
- Janulis, P., Ferrari, J.R., Fowler, P., 2013. Understanding public stigma toward substance dependence. *J. Appl. Soc. Psychol.* 43 (5), 1065–1072. doi:10.1111/jasp.12070.
- Johnson-Kwochka, A., Aalsma, M.C., Monahan, P.O., Salyers, M.P., 2021. Development and examination of the attribution questionnaire-substance use disorder (AQ-SUD) to measure public stigma towards adolescents experiencing substance use disorders. *Drug Alcohol Depend.* 221, 108600. doi:10.1016/j.drugalcdep.2021.108600.
- Kulesza, M., Larimer, M.E., Rao, D., 2013. Substance use related stigma: what we know and the way forward. *J. Addict. Behav. Ther. Rehabil.* 2 (2). doi:10.4172/2324-9005.1000106.
- Kulesza, M., Ramsey, S., Brown, R., Larimer, M., 2014. Stigma among individuals with substance use disorders: does it predict substance use, and does it diminish with treatment? *J. Addict. Behav. Ther. Rehabil.* 3 (1), 1000115. doi:10.4172/2324-9005.1000115.
- Kulesza, M., Watkins, K.E., Ober, A.J., Osilla, K.C., Ewing, B., 2017. Internalized stigma as an independent risk factor for substance use problems among primary care patients: rationale and preliminary support. *Drug Alcohol Depend.* 180, 52–55. doi:10.1016/j.drugalcdep.2017.08.002.
- Livingston, J.D., Milne, T., Fang, M.L., Amari, E., 2012. The effectiveness of interventions for reducing stigma related to substance use disorders: a systematic review. *Addiction* 107 (1), 39–50. doi:10.1111/j.1360-0443.2011.03601.x.
- Lu, J., Sridhar, S., Pandey, R., Hasan, M.A., Mohler, G., 2019. Investigate transitions into drug addiction through text mining of Reddit data. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2367–2375. doi:10.1145/3292500.3330737.
- MacLean, D., Gupta, S., Lembke, A., Manning, C., Heer, J., 2015. Forum77: an analysis of an online health forum dedicated to addiction recovery. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW), pp. 1511–1526. doi:10.1145/2675133.2675146.
- Malik, S., Coulson, N.S., 2010. 'They all supported me but I felt like I suddenly didn't belong anymore': an exploration of perceived disadvantages to online support seeking. *J. Psychosom. Obstet. Gynecol.* 31 (3), 140–149. doi:10.3109/0167482X.2010.504870.
- Matthews, S., Dwyer, R., Snoek, A., 2017. Stigma and self-stigma in addiction. *J. Bioeth. Inq.* 14 (2), 275–286. doi:10.1007/s11673-017-9784-y.
- Moreno, M.A., Goniou, N., Moreno, P.S., Diekema, D., 2013. Ethics of social media research: common concerns and practical considerations. *Cyberpsychol. Behav. Soc. Netw.* 16 (9), 708–713. doi:10.1089/cyber.2012.0334.
- Palamar, J.J., Kiang, M.V., Halkitis, P.N., 2011. Development and psychometric evaluation of scales that assess stigma associated with illicit drug users. *Subst. Use Misuse* 46 (12), 1457–1467. doi:10.3109/10826084.2011.596606.
- Paquette, C.E., Syvertsen, J.L., Pollini, R.A., 2018. Stigma at every turn: health services experiences among people who inject drugs. *Int. J. Drug Policy* 57, 104–110. doi:10.1016/j.drugpo.2018.04.004.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pilatti, A., Bravo, A.J., Pautassi, R.M., 2020. Contexts of alcohol use: a latent class analysis among Argentinean college students. *Drug Alcohol Depend.* 209, 107936. doi:10.1016/j.drugalcdep.2020.107936.
- Sharaff, A., Nagwani, N.K., 2016. Email thread identification using latent Dirichlet allocation and non-negative matrix factorization based clustering techniques. *J. Inf. Sci.* 42 (2), 200–212. doi:10.1177/0165551515587854.
- Škuta, C., Bartůněk, P., Svozil, D., 2014. InChIlib – interactive cluster heatmap for web applications. *J. Cheminform.* 6 (1), 44. doi:10.1186/s13321-014-0044-4.
- Smith, L.R., Earnshaw, V.A., Copenhaver, M.M., Cunningham, C.O., 2016. Substance use stigma: reliability and validity of a theory-based scale for substance-using populations. *Drug Alcohol Depend.* 162, 34–43. doi:10.1016/j.drugalcdep.2016.02.019.
- Soucy, P., 2005. Beyond TFIDF weighting for text categorization in the vector space model. In: In Proceedings of the Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005), pp. 1130–1135.
- South, B., Shen, S., Leng, J., Forbush, T., DuVall, S., Chapman, W., 2012. A prototype tool set to support machine-assisted annotation. In: BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing, pp. 130–139 <https://www.aclweb.org/anthology/W12-2416>.
- Strapparava, C., Valitutti, A., 2004. WordNet-affect: an affective extension of WordNet. In: Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), pp. 1083–1086.
- Viera, A.J., Garrett, J.M., 2005. Understanding interobserver agreement: the kappa statistic. *Fam. Med.* 37 (5), 360–363.
- Witte, T.H., Wright, A., Stinson, E.A., 2019. Factors influencing stigma toward individuals who have substance use disorders. *Subst. Use Misuse* 54 (7), 1115–1124. doi:10.1080/10826084.2018.1560469.
- Wogen, J., Restrepo, M.T., 2020. Human rights, stigma, and substance use. *Health Hum. Rights* 22 (1), 51–60.
- Wolf, J.P., Lipperman-Kreda, S., Bersamin, M., 2019. "It just depends on the environment": patterns and decisions of substance use and co-use by adolescents. *J. Child Adolesc. Subst. Abuse* 28 (3), 143–149. doi:10.1080/1067828X.2019.1637316.
- Yang, L.H., Wong, L.Y., Grivel, M.M., Hasin, D.S., 2017. Stigma and substance use disorders: an international phenomenon. *Curr. Opin. Psychiatry* 30 (5), 378–388. doi:10.1097/YCO.0000000000000351.