

Original Paper

Using Machine Learning of Online Expression to Explain Recovery Trajectories: Content Analytic Approach to Studying a Substance Use Disorder Forum

Ellie Fan Yang^{1*}, PhD; Rachel Kornfield^{2*}, PhD; Yan Liu³, PhD; Ming-Yuan Chih⁴, PhD; Prathusha Sarma⁵, PhD; David Gustafson⁶, PhD; John Curtin⁶, PhD; Dhavan Shah⁶, PhD

¹School of Communication and Mass Media, Northwest Missouri State University, Maryville, MO, United States

²Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

³School of Journalism and Communication, Shanghai University, Shanghai, China

⁴College of Health Science, University of Kentucky, Lexington, KY, United States

⁵Apple Inc, Cupertino, CA, United States

⁶Department of Psychology, University of Wisconsin-Madison, Madison, WI, United States

*these authors contributed equally

Corresponding Author:

Ellie Fan Yang, PhD

School of Communication and Mass Media

Northwest Missouri State University

800 University Dr

239 Wells Hall

Maryville, MO, 64468

United States

Phone: 1 6605621618

Email: eyang@nwmissouri.edu

Abstract

Background: Smartphone-based apps are increasingly used to prevent relapse among those with substance use disorders (SUDs). These systems collect a wealth of data from participants, including the content of messages exchanged in peer-to-peer support forums. How individuals self-disclose and exchange social support in these forums may provide insight into their recovery course, but a manual review of a large corpus of text by human coders is inefficient.

Objective: The study sought to evaluate the feasibility of applying supervised machine learning (ML) to perform large-scale content analysis of an online peer-to-peer discussion forum. Machine-coded data were also used to understand how communication styles relate to writers' substance use and well-being outcomes.

Methods: Data were collected from a smartphone app that connects patients with SUDs to online peer support via a discussion forum. Overall, 268 adult patients with SUD diagnoses were recruited from 3 federally qualified health centers in the United States beginning in 2014. Two waves of survey data were collected to measure demographic characteristics and study outcomes: at baseline (before accessing the app) and after 6 months of using the app. Messages were downloaded from the peer-to-peer forum and subjected to manual content analysis. These data were used to train supervised ML algorithms using features extracted from the Linguistic Inquiry and Word Count (LIWC) system to automatically identify the types of expression relevant to peer-to-peer support. Regression analyses examined how each expression type was associated with recovery outcomes.

Results: Our manual content analysis identified 7 expression types relevant to the recovery process (emotional support, informational support, negative affect, change talk, insightful disclosure, gratitude, and universality disclosure). Over 6 months of app use, 86.2% (231/268) of participants posted on the app's support forum. Of these participants, 93.5% (216/231) posted at least 1 message in the content categories of interest, generating 10,503 messages. Supervised ML algorithms were trained on the hand-coded data, achieving F_1 -scores ranging from 0.57 to 0.85. Regression analyses revealed that a greater proportion of the messages giving emotional support to peers was related to reduced substance use. For self-disclosure, a greater proportion of the messages expressing universality was related to improved quality of life, whereas a greater proportion of the negative affect expressions was negatively related to quality of life and mood.

Conclusions: This study highlights a method of natural language processing with potential to provide real-time insights into peer-to-peer communication dynamics. First, we found that our ML approach allowed for large-scale content coding while retaining moderate-to-high levels of accuracy. Second, individuals' expression styles were associated with recovery outcomes. The expression types of emotional support, universality disclosure, and negative affect were significantly related to recovery outcomes, and attending to these dynamics may be important for appropriate intervention.

(*J Med Internet Res* 2023;25:e45589) doi: [10.2196/45589](https://doi.org/10.2196/45589)

KEYWORDS

supervised machine learning; online peer support forum; expression effects; content analysis; substance use disorder; mobile phone

Introduction

Background

An estimated 19.7 million Americans have a substance use disorder (SUD) in their lifetime, yet few receive treatment [1]. Furthermore, recovery outcomes are poor even among those receiving treatment, with more than two-thirds returning to the use of substances within months of leaving treatment [2,3]. Given that SUDs are serious and stigmatized conditions, bolstering online peer-to-peer communication can play a crucial role in recovery, offering an accessible and comfortable social context within which to discuss challenges, give and receive social support, and build a positive social identity around recovery [4,5]. Smartphone-based support apps are increasingly used to sustain peer support for recovery and prevent relapse [6,7]. These apps can also unobtrusively collect rich data about participants' communication with peers, including the content of messages exchanged and surrounding metadata.

The idea of connecting individuals' patterns of expression to their recovery outcomes predates the emergence of app-based communication; for example, prior work has applied content analysis methods to therapists' interactions with patients in recovery for SUDs, finding that certain patient utterances, such as relaying motivations for change, precede reductions in substance use [8]. Through unobtrusively collecting a wealth of text data, online peer-to-peer support forums now offer a potential avenue to clarify the relationship between communication and recovery. Simultaneously, recent developments in natural language processing and machine learning (ML) provide new opportunities to take advantage of the huge volume of unstructured text in online support forums and other health care applications [9-16]. However, few studies have merged online text with survey-based measures of patients' health and behaviors to understand how communication styles within a peer-to-peer forum relate to SUD recovery outcomes [17].

This study examined communication styles within an app-based peer-to-peer forum for primary care patients with SUDs. In the next subsection, we review the forms of communication indicated in prior literature as potentially relevant in peer-to-peer communication and that could affect patients' recovery course in the context of SUDs.

Types of Expression in Online Support Forums

It remains unclear what types of peer-to-peer communication, especially what types of expression, are key to activating

recovery benefits. Social support is a dominant expression type of online recovery support forums, but individuals also frequently share their thoughts, feelings, and experiences with others through these forums (ie, *self-disclosure*). Self-disclosure may play an important role in recovery by soliciting social support [18]; allowing individuals to work through, and understand, their challenges [19]; or deepening relationships with others [20]. Thus, expressing both social support and self-disclosure is likely to play a role in helping individuals to adjust to recovery, sustain motivation, and enhance their sense of connectedness and general well-being. In the following paragraphs, we summarize what the literature suggests about how specific forms of social support and self-disclosure may factor in health and well-being outcomes in recovery.

Across previous research, emotional support and informational support are the 2 most studied types of communication in peer-to-peer contexts [21,22]. Emotional support refers to messages that help meet recipients' emotional needs, such as by enhancing feelings of being understood, respected, or cared for [18]. It is one of the most common forms of communication across numerous types of online support forums (eg, recovery, cancer, and grief [23]). Although much social support literature focuses on benefits that come from *receiving* emotional support, evidence suggests that *providing* emotional support can also be beneficial because it may help the providers to build relationships and develop a sense of purpose [20,24]; for example, expressing emotional support had a salutary effect on mental well-being among patients with cancer with high communication competence [25] and led to effective coping strategies [26]. Accordingly, giving emotional support is likely to be positively related to recovery outcomes.

Another type of social support communication in the context of SUDs is informational support, which refers to suggestions or information aimed at helping recipients to manage their recovery [27]. This could involve furnishing information about treatment options, coping skills, or solutions to day-to-day problems [27,28]. In addition to helping the receiver to navigate recovery, some evidence suggests that giving informational support may foster self-reflection and reappraisal of one's own problems [29]. Consistent with this, patients with alcohol use disorder had reduced relapse likelihood when they provided informational support [28,30]. However, giving informational support can be challenging because support providers must *match* their support to the recipient's particular challenge [21,31]. Nonetheless, we expect that providing peer-to-peer

informational support will result in improved recovery outcomes.

Beyond social support, self-disclosure, which is defined as “the act of revealing personal information to others” [32], such as feelings, thoughts, and experiences [33], also plays an important role in support forums. A functional perspective suggests that self-disclosure allows communicators to meet diverse strategic goals, depending on the audiences reached and the disclosure context [34,35]. Self-disclosure in a supportive social context may help individuals to process experiences and emotions for effective coping [36,37] as well as elicit reciprocal disclosure [38], with multiple forms of self-disclosure being potentially relevant to recovery.

The first common type of self-disclosure is negative affect expression [18]. This reflects that stress, depression, and frustration often accompany illness [39]. Consistent with the *expressive writing* paradigm, some experimental work suggests that expressing difficult feelings allows individuals to make sense of these experiences and find solutions [40,41]. In social contexts such as peer-to-peer forums, expressing one’s challenges may also lead to receiving more social support [18,42]. However, other studies found that negative emotional disclosures do not always generate peer responsiveness [43,44]. Expressing negative affect can also signal that an individual is facing a more challenging recovery course, preceding relapse [45,46]. Therefore, the relationship between negative affect disclosure and recovery outcomes is unclear.

Another well-studied form of self-disclosure involves generating commitments to behavior change. Within motivational interviewing, “change talk” refers to “self-expressed language that is an argument for change” [47], including conveying the ability, need, desire, or reasons to shift behaviors; commitment to change; or referencing actions linked to behavioral adjustments. Putting one’s argument for change into their language use may increase awareness about discrepancies between one’s goals and current behaviors, motivating action [48]. As such, therapy often aims to create contexts for patients to engage in change talk [49,50]. In SUDs, such talk has been linked to persistence in treatment and reduced substance use [51], reduced alcohol-related negative consequences [52], and increased abstinence from drug use [49]. However, change talk has predominantly been examined in face-to-face therapist-client sessions, leaving it unclear to what extent a shift of language use occurs in online peer-to-peer forums and whether spontaneous peer-to-peer change talk relates to recovery outcomes.

A third type of self-disclosure involves the expression of insights. Reflecting on, writing about, and sharing one’s experiences can facilitate cognitive processing that ultimately fosters an awareness of triggering factors and that may potentially support one in taking new actions to cope with, and address, stressors [19,53]. Insightful disclosure is related to individuals’ adjustment to recovery in various health contexts, including recovering from breast cancer [53] and anorexia [54], suggesting that it may be positively related to an individual’s recovery outcomes.

Additional forms of self-disclosure capture how the communicator reflects on feelings about others within the support group itself; for example, gratitude expression, a common feature of peer-to-peer SUDs forums [18], involves individuals sharing their feelings of thankfulness or appreciation toward those in their life [55]. Some health interventions have sought to promote gratitude expression [56] in the belief that it can enhance recovery outcomes.

Similarly, given that peer support brings together individuals with a common condition or identity, a core function of online peer support groups can be promoting a feeling of shared experience [57,58]. This feeling of “universality” describes “the idea that people have the same experiences or report similar experiences, circumstances, or feelings” [59]. Although less studied in online peer support, expressions of universality may validate and normalize shared identities [60,61], playing an important role for self-acceptance and well-being, particularly in the face of societal stigma [62], such as exists for SUDs. On this account, universality expressions are likely to be positively associated with recovery outcomes.

Thus, the benefits of peer-to-peer communication for recovery can reflect several communicative processes: an expression of emotional and informational support as well as the sharing of emotions, insights, gratitude, and a sense of connection with others. To understand at a nuanced level how these dynamics relate to recovery benefits, this study assessed their relationship with 3 outcomes: substance use, mood, and quality of life. Substance use directly captures behavior change, but mood and quality of life provide a broad view of day-to-day well-being, with improved mood and quality of life potentially supporting long-term sustainment of recovery [63].

Guided by the prior literature, this study examined the relationships between the forms of expression and recovery outcomes. Building on past work [10-13], we first performed a manual content analysis to identify and code the dominant forms of social support and self-disclosure expressed in the forum and then used the manually coded data to train a series of supervised ML algorithms that we applied to the remaining uncoded data. We then assessed the relationships between 7 expression types and 3 recovery indicators: substance use, mood, and quality of life. Overall, the objective of this study was to (1) deepen our understanding of the forms of peer-to-peer communication that underlie the recovery process, (2) demonstrate a method of generating scalable insights about patients’ status and needs, and (3) inform adaptive interventions for recovery support.

Methods

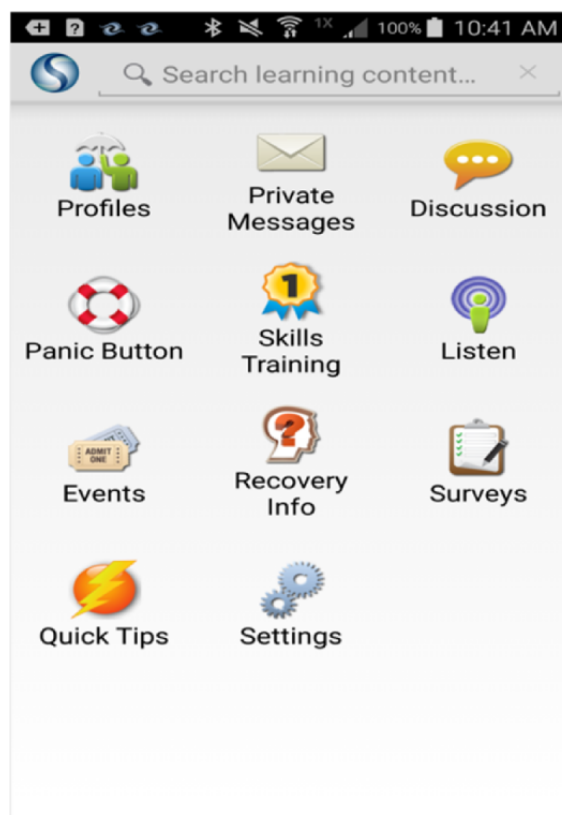
Data Collection and Recruitment

Participants were recruited by primary care providers from 3 federally qualified health centers (FQHCs) in the United States in 2014 and offered access to a smartphone-based health app called *Seva* that was designed to connect patients with SUDs to web-based peers as well as offer other informational and communication resources. In a clinical trial, an earlier version of this app reduced the likelihood of risky drinking by half [6].

The app offers various functions, including motivational journal writing, private messaging, games and relaxation activities, meeting and event directories, and the peer-to-peer discussion

forum (Figure 1). The contents of the app, including the forum discussions, are in English.

Figure 1. Home page of the Seva app.



This study draws on text data from these peer-to-peer discussion forums. Specifically, the *discussion* forum is displayed alongside other features that the user can access on demand through the home page. Within this feature, users may select *groups*, including one that represents all of Seva, as well as site-specific groups that are limited to those enrolled at each of the 3 study sites. Within each group, conversations are listed in chronological order, with the most recent first. Users may click on a conversation to see the initial post as well as all replies it received; thus, responses are nested under an original post. Users responding to an existing post trigger a notification to the original poster. Users are all identified to one another by a pseudonym, and the content of the discussions is only available to the research team and participants in the study. In contrast to the discussions that are visible to those within the same site or everyone on the study (depending on the group), Seva also facilitates private messaging among a dyad, as well as personal journaling features. Although not considered here, these communication features have been examined in other research [64].

Participants were recruited by primary care providers from 3 FQHCs in the United States in 2014. Clinicians identified their potentially eligible patients aged >18 years who met the criteria for a SUD diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [65], who they thought may benefit from the study and who had no current psychotic disorder severe enough to prevent participation and no acute medical problem requiring immediate inpatient

treatment. After providing informed consent in English, participants received training on how to use Seva, including its discussion forum. The participants also received compensation in the form of payment for the mobile phones and data plans that allow the operation of Seva. A maximum of 100 patients per clinic were recruited through this approach. More details about the app and study design are provided elsewhere [65,66]. To protect participants' privacy, app use data, including the content of messages posted to the peer-to-peer discussion forum, were stored in a secure encrypted database accessible only to the researchers. The information shared by the participants was anonymized before analysis.

In total, 268 patients were included in the study. Over 6 months, 86.2% (231/268) of these participants posted 10,548 messages on the peer-to-peer forum. We focus here on 93.5% (216/231) of these participants who posted at least 1 message, including a form of social support or self-disclosure. Those who did not post within these categories typically posted only 1 or 2 introductory or greeting messages. Two waves of survey data were collected, capturing demographic characteristics and study outcomes: one at the baseline (before accessing the app) and one after 6 months of using the app. Of the 216 patients in this analysis, all completed the baseline survey, and 180 (83.3%) completed the 6-month survey.

Ethics Approval

The study received approval from the Health Sciences Institutional Review Board at the University of

Wisconsin–Madison (2012-0937-CP020) and has been registered at ClinicalTrials.gov (NCT01963234).

Informed Consent

All participants provided informed consent, which allows secondary analysis.

Measurement

Human Content Coding

To support our analyses, we first developed a codebook to capture the major types of messages on the forum that may play a role in the recovery process. The codebook was developed from data from an earlier version of the system used in the clinical trial [18,31] and reflected an iterative approach integrating both deductive and inductive techniques [67]. First, we reviewed the literature on peer-to-peer support to deductively define expression types that are central to peer-to-peer communication. As described previously in the literature review, informational and emotional support are well established within peer-to-peer forums. This study adopted coding definitions based on the study by Cohen and Wills [68], renaming *cognitive support* as *informational support*. We developed additional coding categories inductively, based on 3 coders independently reviewing a subset of the messages to observe forms of self-disclosure that emerged in peer-to-peer discussions. On the

basis of the codes identified within the data, we conducted a targeted literature review to guide decisions of how to formalize these as coding categories (eg, those literature guided us to code expressions of negative affect and gratitude). These codes were consolidated within a preliminary shared codebook.

Across several rounds of coding, 3 coders overlapped in coding subsets of the data and then met to discuss and adjudicate coding discrepancies and to refine the codebook to better encompass the data, resulting in a formalized codebook for which the coders achieved satisfactory interrater reliability (IRR) for each code (Cohen $\kappa \geq 0.70$) in an overlapping subset of 100 messages per code, consistent with the minimum threshold for computing IRR in prior literature [69]. Coding discrepancies in the overlapping subsets were adjudicated by consensus. Given the substantial effort involved in manual coding, after calculating IRR, the remaining uncoded messages were divided evenly among the coders. Table 1 displays the operationalization of the 7 communication types, along with example messages and the IRR achieved by the coders. Codes were not mutually exclusive, meaning that the same message may include multiple expression types. As Seva is a research forum that is not accessible to the public, only research team members can search the text of messages. Some of the quotes presented in the example message column have been paraphrased.

Table 1. Operationalization, example messages, and interrater reliabilities for the 7 expression types.

Expression type	Operationalization	Example message	Cohen κ (human coding)
Emotional support	Messages that foster feelings of comfort and lead the recipient to believe that they are understood, admired, respected, and loved and that others are available to provide caring and security; the writer may convey that they feel for the recipient and wish, or offer prayers and encouragement, for their well-being	“That’s awesome you’re feeling better” [Participant 1]	0.73
Informational support	Messages that relay information, knowledge, and advice to help the receiver understand their world and to adjust to changes within it; the writer may convey how they think it would be useful or appropriate to think about, or respond to, a given situation or may suggest resources, including websites, features of the smartphone app, or other sources	“Exercise helps for anxiety” [Participant 3]	0.78
Negative affect	Messages that reveal the writer’s negative emotions toward events and experiences (eg, fear, anxiety, sadness, anger, or guilt)	“Sometimes I feel like I’m going crazy. Just needed to vent” [Participant 12]	0.84
Change talk	Messages that make an argument for personal behavior change, including conveying ability, need, desire, or reasons to change; commitment to change; and referencing actions linked to behavior change intentions	“I have to put down and stop alcohol and drugs” [Participant 15]	0.70
Insightful disclosure	Messages suggesting that the writer has gained a better or deeper understanding of their experiences over time	“I realized how important it is to take care of myself” [Participant 22]	0.73
Gratitude	Messages giving thanks to specific others or expressing general appreciation	“thanks [username]. [smiley emoji]” [Participant 25]	0.94
Universality	Messages that express recognition of first-hand experience of similar situations or feelings as the recipient	“I too have felt the same, depressed on so many different levels” [Participant 34]	0.72

ML Approach

As manual coding for the total 14,393 messages collected in the forum is labor-intensive and time-consuming, this study improved the efficiency of the coding process by applying an ML approach. The 3 coders applied the final codebook to the

full set of training data that would be used to develop the ML classifiers. The full codebook was first applied to all 2590 messages exchanged by patients with SUDs within the prior iteration of the app within which the codebook was developed [18,31]. Using this human-coded data as training data, we developed a series of machine classifiers to predict the presence

or absence in each message of the aforementioned 7 expression types (emotional support, informational support, negative affect, change talk, insightful disclosure, gratitude, and universality). We iteratively tested algorithm performance, examining F_1 -scores to assess when additional manual coding was needed to expand the training set and improve algorithm performance, based on a minimum acceptable threshold of 0.55 based on the F_1 -scores provided in prior literature [70-74]. Of the total 14,393 messages that participants exchanged in the study data set (from the deployment within FQHCs), we coded up to 3515 (24.42%) additional messages. Of the 14,393 messages, the total number of human-coded messages ranged from 4467 (31.04%; for negative affect and insightful disclosure) to 6105 (42.42%; for emotional support, informational support, gratitude, and universality). This selective coding process allowed us to reduce the burden on the human coders, leveraging their effort where training data were most needed to improve performance. To assess the ML algorithms, we prepared 1 training set and 4 test sets for each of the 7 expression types, all derived from the human-coded message data set. The training set was drawn from the corpus of coded messages from the study data set that was posted across the entire observation window and was randomly selected to reduce the likelihood of biases owing to changes in the conversation dynamics over time. The human-coded labels were compared with the labels predicted by the algorithms, generating classifier performance statistics across the test sets.

Specifically, drawing on prior work [73], our supervised ML process drew on the linguistic features of messages extracted using the Linguistic Inquiry and Word Count (LIWC) system, which automatically computes the rates at which messages include 90 linguistic features (eg, mentions of *love*, *nice*, and *sweet* are coded as instances of *positive emotion*) [75]. These 90 features were extracted from all training data and test data, and the labels of the training data were used to predict the labels of the test data. Using *scikit-learn* in Python (Python Software Foundation) [76], we used boosted decision tree algorithms—a type of supervised ML where each internal node represents a *test* on an attribute, and each branch represents the outcome of that test, with the performance boosted by making each *tree* dependent on prior trees by fitting the residual of the trees that preceded it. This approach has performed well in similar data sets [72,73,77]. For content categories with imbalanced classes (eg, far more messages did not mention gratitude than those that did), we compensated for the imbalance by oversampling from the minority class using the Synthetic Minority Oversampling Technique [78]. Specifically, after splitting the data set into training and testing sets, we generated synthetic samples from the minority class to extend the training data set. These synthetic samples are created to closely resemble but not exactly replicate existing samples within the minority class.

The other Python packages used in the study are *pandas* [79], *NumPy* [80], *pickle* [81], and *Natural Language Toolkit* [82].

Survey Measures

Substance use (ie, any drinking or drug use) and well-being outcomes (ie, positive mood and quality of life) were self-reported through baseline and 6-month surveys.

Any Substance Use

At both baseline and the 6-month follow-up, the survey asked participants to report whether they had consumed any alcoholic beverages in the past 30 days. In addition, the survey asked whether participants had used any illegal or street drugs or abused any prescription medications in the past 30 days. To capture whether participants engaged in any substance use, we created a variable combining these 2 binary measurements, that is, recoding as *Yes* if they engaged either in drinking or drug use.

Mood

The Positive and Negative Affect Schedule–Short Form (PANAS-SF) was used to measure mood [83]. The scale consists of 10 items, of which 5 measure positive affect, and 5 measure negative affect. Cronbach α values for the positive affect and negative affect dimensions were .83 and .80, respectively. This study adopts a single dimension by reversing the negative items (Cronbach α =.80), following Kim and Hatfield [84] and Furlong et al [85], producing a measure of mood as a single dimension. We also constructed positive and negative affect subscales so that we could test these components.

Quality of Life

The Patient-Reported Outcomes Measurement Information System (PROMIS) Global-10 measure consists of 10 items assessing general functioning and well-being across physical and mental health domains [86]. PROMIS assesses the general health care–related quality of life. Consistent with the PROMIS scoring guideline that combines the 2 dimensions of health [86], 3 of the 10 items were recoded such that all items were scored on a scale from 1 to 5, with higher scores indicating higher functioning, after which all 10 items were averaged (Cronbach α =.73). We also estimated reliability for the physical and mental health subscales (Cronbach α for physical health=.68 and Cronbach α for mental health=.81).

The zero-order correlations for the 7 expression types were also assessed, as presented in Table 2. The correlations ranged from –0.24 (emotional support and negative affect) to 0.54 (insightful disclosure with both change talk and universality), with all falling below the threshold of 0.80 that might raise concerns about multicollinearity if entered simultaneously in regression models [87].

Table 2. Zero-order correlations for the 7 expression types (n=216)^a.

	Emotional support	Informational support	Negative affect	Change talk	Insightful disclosure	Gratitude	Universality disclosure
Emotional support	1	0.3	-0.24	-0.19	-0.21	-0.10	-0.09
Informational support	0.3	1	0.07	0.08	0.2	-0.20	0.27
Negative affect	-0.24	0.07	1	0.3	0.41	0.15	0.35
Change talk	-0.19	0.08	0.3	1	0.54	0.32	0.27
Insightful disclosure	-0.21	0.2	0.41	0.54	1	0.31	0.54
Gratitude	-0.10	-0.20	0.15	0.32	0.31	1	0.18
Universality	-0.09	0.27	0.35	0.27	0.54	0.18	1

^aWe also calculated the mean and SD of the 7 types of expression: emotional support (mean 0.41, SD 0.23), informational support (mean 0.23, SD 0.20), negative affect (mean 0.07, SD 0.10), change talk (mean 0.07, SD 0.11), insightful disclosure (mean 0.16, SD 0.19), gratitude (mean 0.27, SD 0.20), and universality disclosure (mean 0.08, SD 0.10).

Statistical Analysis

The 2 waves of survey data and the machine-coded discussion forum data were merged based on the participants' unique identifiers. For the discussion forum data, the count of messages falling within each expression category was aggregated for each user over the 6-month study period. As enrollment was conducted on a rolling basis, time on the study was relative to each participant's start date. Given that the raw counts of messages in each category can be biased by the participants' uneven message production (ie, some participants produce many more messages than others), we created a proportion score for each expression type, dividing the count of messages within each expression type by the total number of messages posted by the participant. Therefore, the value of the emotional support measure ranges from 0 to 1 (a 1-unit change in this measure). Hierarchical ordinary least squares (OLS) and logistic regression models were used to evaluate the relationships between the proportion of an individual's messages in each category and the 3 recovery outcomes as assessed at 6 months, controlling for the baseline value of the outcome. All models also controlled for baseline alcohol and other drug use, age, sex, education, and race because they were significant predictors of SUD recovery

in prior research [88,89]. SPSS software (version 25.0; IBM Corp) was used for the regression analyses.

Results

Participant and Message Characteristics

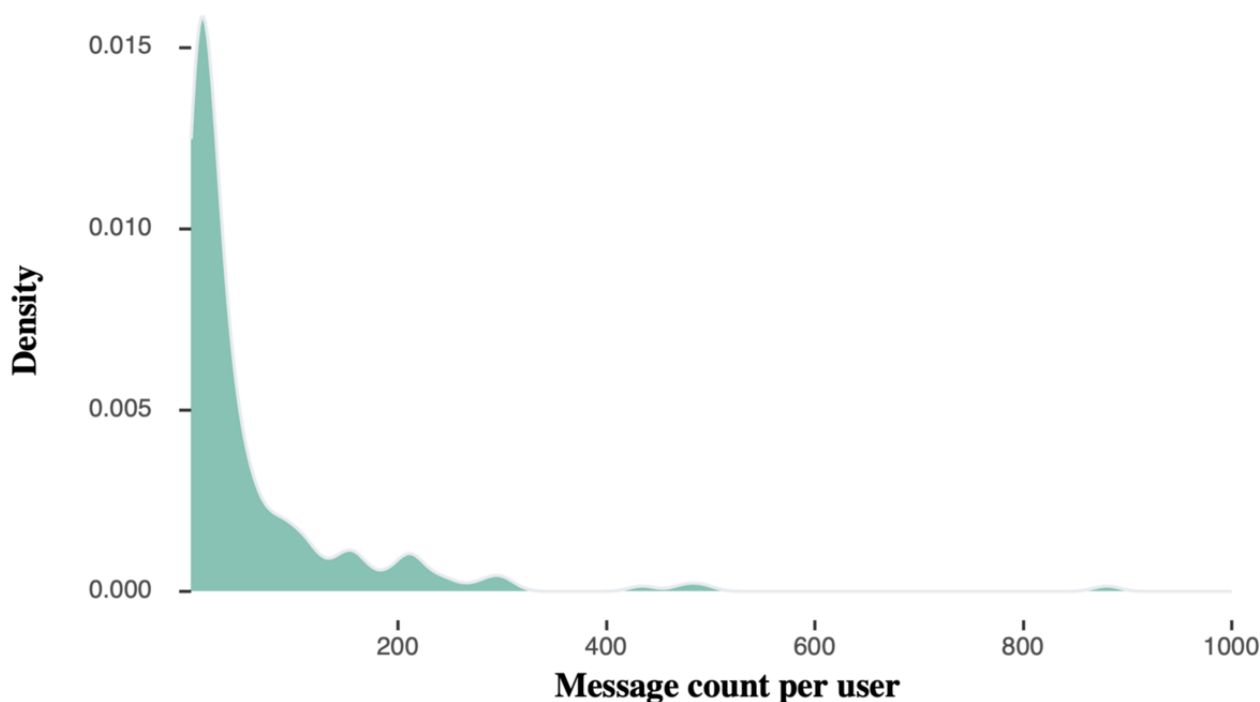
Table 3 summarizes participants' demographics and outcome measurements from the 2 waves. Despite some attrition of participants (36/216, 16.7%) between the baseline and 6-month surveys, we did not find demographic differences between those who remained on the study and those who did not.

A total of 10,503 messages were posted by the participants (n=216) over 6 months, most frequently in the first 2 months. The average length of messages in the study data set was 146.37 (SD 75.00) characters, with the median length being 139.00 (IQR 31.00-170.00). Of the 10,503 messages, 1576 (15.01%) were original messages (starting a new thread), with the rest being replies to an existing thread. The users each produced on average 72 (SD 164) messages over 6 months on the study. Figure 2 displays the density distribution of message numbers per user. The density is derived on Gaussian kernels from the R graphing package *ggplot*.

Table 3. Descriptive statistics of participants at baseline and 6-month surveys.

Demographics	Baseline (n=216)	6-month survey (n=180)
Education, n (%)		
Less than college	108 (50)	87 (48.3)
Some college	64 (29.6)	57 (31.7)
More than college	44 (20.3)	36 (20)
Sex, n (%)		
Male	114 (52.8)	94 (52.2)
Female	102 (47.2)	86 (47.8)
Race, n (%)		
People of color	73 (33.8)	60 (33.3)
White	143 (66.2)	120 (66.7)
Age (years), mean (SD)	41.8 (10.60)	42.3 (10.73)
Outcomes		
Any drinking (yes), n (%)	68 (31.5)	40 (18.5)
Drug use (yes), n (%)	60 (27.8)	27 (12.5)
Positive mood, mean (SD)	3.22 (0.76)	3.54 (0.71)
Quality of life, mean (SD)	28.48 (9.70)	30.02 (7.02)

Figure 2. Distribution of message count by Seva app user.



ML Performance

Table 4 presents the final algorithm performance indexes for each expression type, as averaged across the 4 test sets. The boosted decision tree ML algorithms achieved F_1 -scores ranging from 0.57 to 0.85, area under the receiver operating characteristic curve values ranging from 0.72 to 0.86, sensitivity

scores ranging from 0.75 to 0.95, and specificity scores ranging from 0.65 to 0.79.

On the basis of the machine classifiers applied to the entire study data set, emotional support was the most common expression type (4788/10,503, 45.59%), followed by gratitude (2972/10,503, 28.3%), informational support (2514/10,503, 23.94%), insightful disclosure (1514/10,503, 14.41%), universality (901/10,503, 8.58%), negative affect (731/10,503,

6.96%), and change talk (707/10,503, 6.73%). [Table 5](#) displays the average proportion of messages in each category in each study month for each expression type. In linear growth curve models assessing change in monthly expression in each category

(not shown), only 2 expression types had slopes significantly different from 0: emotional support ($\beta=.012$; $P=.03$) and universality ($\beta=.008$; $P=.007$), with both increasing over time.

Table 4. F_1 -scores, area under the receiver operating characteristic curve (AUROC) values, and sensitivity and specificity scores for boosted decision tree algorithm performance for the 7 expression types.

Expression types	F_1 -score	AUROC	Sensitivity	Specificity
Emotional support	0.65	0.72	0.75	0.70
Informational support	0.71	0.76	0.79	0.72
Negative affect	0.57	0.79	0.86	0.73
Change talk	0.85	0.86	0.95	0.74
Insightful disclosure	0.60	0.76	0.86	0.65
Gratitude	0.70	0.78	0.81	0.79
Universality	0.59	0.79	0.83	0.75

Table 5. Proportion of messages posted within each of the 7 expression categories by study month.

	Month 1 (n=216), average proportion (SD)	Month 2 (n=190), average proportion (SD)	Month 3 (n=163), average proportion (SD)	Month 4 (n=136), average proportion (SD)	Month 5 (n=115), average proportion (SD)	Month 6 (n=97), average proportion (SD)
Emotional support	0.40 (0.31)	0.40 (0.32)	0.47 (0.34)	0.45 (0.35)	0.48 (0.33)	0.44 (0.33)
Informational support	0.20 (0.26)	0.25 (0.28)	0.22 (0.26)	0.25 (0.28)	0.25 (0.29)	0.26 (0.30)
Negative affect	0.06 (0.12)	0.09 (0.17)	0.08 (0.18)	0.08 (0.17)	0.11 (0.22)	0.07 (0.18)
Change talk	0.07 (0.17)	0.07 (0.17)	0.06 (0.16)	0.08 (0.17)	0.11 (0.20)	0.08 (0.18)
Insightful disclosure	0.14 (0.21)	0.18 (0.27)	0.19 (0.26)	0.18 (0.24)	0.18 (0.27)	0.15 (0.21)
Gratitude	0.26 (0.24)	0.25 (0.25)	0.27 (0.29)	0.28 (0.31)	0.27 (0.28)	0.25 (0.29)
Universality	0.06 (0.12)	0.09 (0.18)	0.07 (0.13)	0.09 (0.17)	0.12 (0.22)	0.11 (0.22)

Regression Analysis

A series of hierarchical OLS and logistic regression models were run to predict substance use, mood, and quality of life at 6 months based on the 7 types of expression within the peer-to-peer discussion forum. We controlled for the baseline measurement of each outcome, baseline drinking and drug use, and demographic covariates as described previously. For substance use, [Table 6](#) shows that expressing a higher proportion of emotional support was significantly ($P=.03$) associated with a decreased likelihood of substance use. After controlling for baseline substance use, a 1-unit increase in emotional support proportion is associated with an 88% decrease in the odds of reporting substance use at 6 months (odds ratio 0.12, CI 0.016-0.83; $P=.03$). The standardized coefficient estimates suggest that compared with the other 6 types of expression, emotional support had the strongest negative association with the change of any drinking or drug use at 6 months (standardized $\beta=-.49$; $P=.03$).

For well-being, we found that expressing a greater proportion of messages containing negative affect was significantly associated with lower positive mood score at 6 months ([Table 7](#)), with a 1-unit increase in the proportion of negative affect expression within a message being associated with a 1.49-point decline in the mood score at 6 months as measured on the

PANAS-SF compared with baseline ($\beta=-1.49$; $P=.007$). Considering the typical use of the PANAS-SF as 2 dimensions (ie, positive mood and negative mood), we also applied OLS regression to the positive and negative subscales separately. We found a significant negative relationship between negative affect expression and positive mood ($P=.02$), whereas the relationship between negative affect expression and negative mood did not achieve significance ($P=.09$). These results are presented in [Multimedia Appendix 1](#). This model, including expression types and baseline outcome control as the predictors, explained 62% of the variance in mood change ($F_{13,164}=7.73$; $P<.001$; $R^2=0.62$). Similarly, greater negative affect expression was significantly associated with lower reported quality of life at 6 months compared with baseline, with a 1-unit increase in the proportion of negative affect expressions being associated with an 11-point reduction on the PROMIS scale ($\beta=-11.00$; $P=.05$; [Table 8](#)). Conversely, the proportion of universality expressions was found to be positively associated with quality of life ($\beta=11.83$; $P=.04$), with a 1-unit increase associated with an 11-point increase on the PROMIS scale compared with baseline. The model fit the data well ($F_{13,160}=7.64$; $P<.001$; $R^2=0.62$).

In summary, we found significant relationships between recovery outcomes and a subset of the expression types examined. The proportion of emotional support expression was

positively associated with a reduced likelihood of substance use but not well-being outcomes at 6 months. We did not find any significant relationships between the proportion of informational support expressions and recovery outcomes. The proportion of negative affect expressions was not significantly related to substance use but was linked with decreased mood and quality

of life. We found no significant relationships between recovery outcomes and the proportion of expressions that included change talk, insightful disclosure, or gratitude. Finally, the proportion of universality expressions was associated with positive mood but not substance use or quality of life.

Table 6. Hierarchical logistic regression models using expression types to predict whether individuals engage in any drinking or drug use at 6 months.

	Model 1 ^a				Model 2 ^b			
	Unstandardized β estimates	Standardized β estimates	OR ^c (95% CI)	<i>P</i> value	Unstandardized β estimates	Standardized β estimates	OR (95% CI)	<i>P</i> value
Demographics								
Age	-.020	-.240	0.97 (0.94-1.01)	.19	-.020	-.230	0.98 (0.94-1.10)	.22
Education	-.100	-.050	0.91 (0.44-1.88)	.80	-.100	-.050	0.90 (0.43-1.90)	.79
Sex: male	-.007	.000	0.99 (0.48-2.04)	.99	-.090	-.040	0.92 (0.43-1.97)	.82
Race: White	.170	.080	1.19 (0.54-2.59)	.67	.050	.020	1.05 (0.45-2.46)	.92
Baseline drinking and drug use	1.720	.860	5.56 (2.65-11.68)	<.001	1.820	.090	6.14 (2.84-13.26)	<.001
Expression types								
Emotional support	N/A ^d	N/A	N/A	N/A	-2.150	-.490	0.12 (0.016-0.83)	.03
Informational support	N/A	N/A	N/A	N/A	1.740	.350	5.70 (0.56-0.44)	.79
Negative affect	N/A	N/A	N/A	N/A	-.690	-.070	0.50 (0.00-58.066)	.75
Change talk	N/A	N/A	N/A	N/A	-.530	-.060	0.59 (0.01-42.24)	.81
Insightful disclosure	N/A	N/A	N/A	N/A	.060	.010	1.06 (0.07-15.33)	.96
Gratitude	N/A	N/A	N/A	N/A	.750	.140	2.12 (0.17-26.26)	.56
Universality	N/A	N/A	N/A	N/A	-.590	-.060	0.56 (0.01-45.08)	.79

^aThe intercept for model 1 is -0.90 (log likelihood=187.81, Cox & Snell $R^2=0.15$, and McFadden $R^2=0.14$).

^bThe intercept for model 2 is -0.44 (log likelihood=151.67, Cox & Snell $R^2=0.18$, and McFadden $R^2=0.17$).

^cOR: odds ratio.

^dN/A: not applicable.

Table 7. Hierarchical linear regression models using expression types to predict positive mood at 6 months^a.

	Model 1 ^b			Model 2 ^c			Model 3 ^d		
	Unstandardized β estimates	Standardized β estimates	P value	Unstandardized β estimates	Standardized β estimates	P value	Unstandardized β estimates	Standardized β estimates	P value
Demographics									
Age	.002	.030	.73	-.007	-.100	.13	-.008	-.120	.08
Education	-.090	.110	.40	-.080	-.060	.35	-.060	-.040	.50
Sex: male	.180	.110	.11	.140	.100	.12	.120	.090	.19
Race: White	-.060	.120	.60	-.080	-.050	.40	-.070	-.050	.48
Baseline positive mood	N/A ^e	N/A	N/A	.580	N/A	<.001	.580	.610	<.001
Baseline drinking and drug use	N/A	N/A	N/A	.100	N/A	.29	.120	.080	.22
Expression types									
Emotional support	N/A	N/A	N/A	N/A	N/A	N/A	-.150	-.050	.54
Informational support	N/A	N/A	N/A	N/A	N/A	N/A	-.100	-.030	.72
Negative affect	N/A	N/A	N/A	N/A	N/A	N/A	-1.490	-.210	<.001
Change talk	N/A	N/A	N/A	N/A	N/A	N/A	.840	.110	.12
Insightful disclosure	N/A	N/A	N/A	N/A	N/A	N/A	.220	.060	.51
Gratitude	N/A	N/A	N/A	N/A	N/A	N/A	.020	.005	.95
Universality	N/A	N/A	N/A	N/A	N/A	N/A	-.180	-.030	.75

^aMale, White, and baseline any drinking and drug use were dummy coded as Male=1, White=1, and Had any drinking and drug use=1. Education is an ordinal scale.

^bThe intercept for model 1 is 3.46 ($R^2=0.16$; $F^2=0.19$).

^cThe intercept for model 2 is 1.96 ($R^2=0.59$; $F^2=1.44$).

^dThe intercept for model 3 is 2.09 ($R^2=0.62$; $F^2=1.63$).

^eN/A: not applicable.

Table 8. Hierarchical linear regression models using expression types to predict quality of life (QOL) at 6 months^a.

	Model 1 ^b			Model 2 ^c			Model 3 ^d		
	Unstandardized β estimates	Standardized β estimates	P value	Unstandardized β estimates	Standardized β estimates	P value	Unstandardized β estimates	Standardized β estimates	P value
Demographics									
Age	-.13	-.19	.01	-.11	-.17	.01	-.13	-.19	.006
Education	1.21	.09	.26	.97	.07	.29	1.40	.10	.13
Sex: male	1.67	.12	.12	1.50	.11	.10	1.53	.11	.11
Race: White	1.49	.10	.19	1.25	.08	.20	1.32	.09	.20
Baseline QOL	N/A ^e	N/A	N/A	.58	.52	<.001	.58	.53	<.001
Baseline drinking and drug use	N/A	N/A	N/A	-.23	-.02	.81	-.27	-.02	.78
Expression types									
Emotional support	N/A	N/A	N/A	N/A	N/A	N/A	.37	.01	.88
Informational support	N/A	N/A	N/A	N/A	N/A	N/A	-1.99	-.06	.48
Negative affect	N/A	N/A	N/A	N/A	N/A	N/A	-11.00	-.16	.05
Change talk	N/A	N/A	N/A	N/A	N/A	N/A	-2.04	-.03	.71
Insightful disclosure	N/A	N/A	N/A	N/A	N/A	N/A	-1.54	-.04	.65
Gratitude	N/A	N/A	N/A	N/A	N/A	N/A	-1.02	-.03	.73
Universality	N/A	N/A	N/A	N/A	N/A	N/A	11.83	.17	.04

^aMale, White, and baseline any drinking and drug use were dummy coded as Male=1, White=1, and Had any drinking and drug use=1. Education is an ordinal scale.

^bThe intercept for model 1 is 32.99 ($R^2=0.26$; $F^2=0.35$).

^cThe intercept for model 2 is 16.28 ($R^2=0.59$; $F^2=1.44$).

^dThe intercept for model 3 is 17.31 ($R^2=0.62$; $F^2=1.63$).

^eN/A: not applicable.

Discussion

Principal Findings

Peer-to-peer communication within SUD apps can provide a wealth of data about individuals' experiences and likely health trajectories, but a review by hand and the coding of messages can quickly become prohibitively effortful and time consuming. This study makes a methodological contribution by demonstrating that supervised ML can be applied to perform large-scale automated content analysis in a peer-to-peer discussion group context. The ML performance was consistent with that achieved using similar methods in prior literature [70-74], with variation across content categories ranging from moderate to good performance. In turn, this content analysis allowed us to examine the dominant ways in which individuals engage in app-based peer-to-peer communication over 6 months and to explore how each communication style relates to the user's recovery course. Our results add to a growing body of evidence suggesting that informal online peer-to-peer conversations provide a valuable source of information about participants' future health trajectories [12,90] and point to giving

emotional support and expressing a sense of universality as potential predictors of reduced substance use and improved quality of life, respectively. By contrast, negative affect expressions were associated with lower quality of life and mood. As discussed in the following paragraphs, these findings deepen our understanding of how peer-to-peer communication may bolster recovery as well as suggest potential opportunities to improve how app-based SUD interventions adapt to better meet users' needs.

Our findings add to a growing body of work suggesting that the expression of social support, particularly emotional support, is a dominant expression type of peer-to-peer forums and can potentially benefit individuals in their recovery. Of course, because these data are correlational, the reverse is also possible, meaning that those experiencing positive recovery outcomes express themselves in particular ways. Expressing emotional support was negatively associated with substance use after 6 months of accessing the app, which is consistent with other work in the domain of SUD recovery [91]. Indeed, this was the only communication type that was associated with this behavioral outcome. We did not find evidence of a similar role

for informational support, which could reflect that informational support can be more challenging or potentially burdensome to give than emotional support [31,92]. These findings suggest a need to distinguish between giving informational support and giving emotional support when examining social support in peer-to-peer communication.

Our research also looks beyond social support to consider self-disclosure. Although less often examined than social support in peer-to-peer forums, self-disclosure can be an important method of soliciting support and can provide a pathway to better understand one's own recovery process or to connect to other group members through finding areas of shared experience and understanding. Whereas some past literature has proposed potential benefits that come from openly expressing negative emotions, our findings are more consistent with those of prior research, which found that high levels of self-disclosure of negative affect predict health risks [93], likely capturing individuals' underlying struggles and challenges. Finally, although past work has noted that experiences of oneness or universality are common in peer support groups [60], this work establishes that individual differences in expressing such feelings distinguish those more likely to have greater well-being at follow-up. Experiences of universality and belonging may be important in the SUD context to counteract the feelings of alienation and loneliness that individuals frequently experience [57,59].

Our analysis of the monthly proportions of message types shows that both emotional support and universality expression tended to increase as a proportion of total messages as the patients engaged with this support forum over time. The patterns of increasing emotional support are consistent with prior work suggesting that participants in support groups often move from primarily receiving to giving support as their health trajectories become more stable, and they feel more confident with their role in the group [94]. Longer-term participation may also allow for finding commonalities and building a sense of belonging, as reflected in an increasing proportion of universality expressions.

These findings have implications for the design of online support forums. First, although our findings are correlational, it may be worth considering whether designers can make support forums more conducive to certain communication styles that are associated with improvements in recovery outcomes; for example, our findings diverge for emotional support giving and informational support giving, which may warrant a greater focus on buttressing emotional support giving. Systems could attempt to address this by offering guidance to support givers, perhaps encouraging emotional support delivery or giving training for those who do want to give informational support so that they can also find applications to their own recovery process. For universality, structured activities may support finding commonalities with others. Alternatively, the perceptions of similarity may be increased by matching group participants based on similarities or by displaying information about other participants that highlights shared experience [37].

Our findings also show that the expression types linked to positive recovery outcomes (ie, emotional support giving and

universality expressions) increased over time as a proportion of messages sent, suggesting the potential importance of sustaining participation and motivating more participants to interact. Often, as in the case of any online discussion forum, peer-to-peer forums face a *cold start problem*, wherein early contributions to a forum generally receive little engagement; yet, without engagement, other participants are unlikely to participate [95]. This has led to strategies such as seeding a group with highly motivated participants to create a sense of community, given that participation is based on anticipated future interaction from group members [35,96]. Studies also suggest tailoring the engagement tactics based on a forum's life cycle [97], including investing time early to engage core members [98].

Real-time understanding of communication dynamics may also facilitate just-in-time intervention and system adaptation; for example, expressing a high proportion of messages that convey negative affect may be an important signal that reflects when participants are struggling and need additional support. This support might come from peers (eg, by prompting peers to provide support to a specific participant), professionals (eg, system *moderators* may reach out privately or respond to threads involving this participant), or from automated system features (eg, a chatbot may launch a dialogue, or the system might recommend or launch modules to cope with negative emotions). Similarly, individuals who participate without expressing emotional social support could potentially benefit from tailored guidance in delivering this type of support, whereas such guidance may not be indicated for others.

Finally, this study does not develop or test models that predict real-time relapse, but the communication types examined here could be considered as potential features to integrate into such models, allowing for identifying specific moments when individuals may need support and resources. A growing body of work shows that peer-to-peer communication dynamics (eg, communication on social media) are associated with individuals' health behaviors [90]. Integrating algorithmically captured content categories can potentially improve upon past approaches because we focused on communication types that are prevalent in peer-to-peer communication for SUDs and where the literature suggests potential relationships to an individual's recovery course. The rates of producing content falling into these categories can be computed automatically, essentially in real time. Although most of the communication types examined were not related to substance use at 6 months, future research can look at both communication and health behavior at a more granular timescale. Past work suggests that models that incorporate heterogeneous features often perform best in real-time prediction [10], and thus peer-to-peer communication dynamics may be considered alongside other features such as app use, location, movement, and ecological momentary assessment.

This study has limitations. First, we examined only the effects of giving support and not receiving it. Receiving support of various forms is also associated with individuals' recovery outcomes [31], and the types of support received could potentially be automatically assessed using similar methods. A second limitation is that our measurement of substance use did

not encompass recovery efforts other than abstinence, such as by capturing reduction in use. According to a recent recommendation [99], recovery outcomes should reflect 3 components: biopsychosocial functioning, remission, and cessation. Moreover, our sample may not represent those with SUDs as a whole because the participants were drawn from among those in contact with the primary care system. These participants were also willing to consent to the study, and they engaged with the app by voluntarily communicating via the support forum. It is unclear whether the same patterns of communication and associations with recovery outcomes would emerge in other populations with SUDs or if tested using social networking platforms (eg, Facebook) or messaging tools (eg, texting or SMS text messaging). We are also unable to use these methods for those participants who never post, and the utility may be reduced for those who post infrequently. There may be expression types that are more strongly associated with recovery and well-being outcomes for those who post more versus those who post less within the forum. It is also likely that associations between expression and health and well-being outcomes are more robust among individuals who post more often. As with any manual content analysis, it is possible that our codebook may have been biased toward emphasizing certain patterns of communication and overlooking other important dynamics. As with any study of supervised ML, it is also possible that the automated coding process may be biased to systematically overlook or misclassify some patterns of language. Future work may apply innovations in ML such as neural networks to improve the classifiers [100]. In addition, more robust classification approaches would include multifold cross-validation [101] and a holdout validation set to assess

classifier performance in new data, reduce bias in the classification models, and facilitate transfer learning [102]. Furthermore, messages from the same participants may have been in both the training and test sets, whereas a more conservative approach would be to have mutually exclusive sets of participants in the training and test sets.

A final limitation is that the association between communication styles and recovery outcomes is correlational. We cannot say with confidence whether the relationships observed are causal (ie, whether communicating in a particular way produces changes in the recovery course). However, to increase the strength of the claims we can make, we attempted to control for potential confounds. Furthermore, the data overwhelmingly come from the first several months on the study, with these expressions used to predict outcomes months later such that temporal ordering is consistent with a potentially causal relationship.

Conclusions

With the prevalence of the internet, online peer-to-peer support forums represent a growing support venue for those in recovery. This study extracted 7 types of messages exchanged on a smartphone-based forum and applied supervised ML to perform large-scale quantitative content analysis over 6 months. Analyses leveraging the machine-coded data suggest forms of peer-to-peer communication that distinguish individuals' likely recovery course, notably emotional support, universality, and negative affect expressions. Attending to these forms of expression may help to develop interventions that better respond to participants' recovery needs.

Acknowledgments

This work was supported by the National Institute on Drug Abuse under grant R01-DA034279. All interpretations of these data should be attributed to the authors, who thank the Center for Health Enhancement Systems Studies at UW-Madison for making these data available. YL was supported by Humanities and Social Science Foundation of Ministry of Education of China (19YJC860029) and Shanghai Pujiang Program (2020PJC056). RK was supported by the National Institute of Mental Health (P50MH119029 and K01MH125172).

Data Availability

We are not able to make the discussion post data from this study publicly available owing to the sensitive nature of the peer-to-peer communication data in recovery support forums and the fact that the participants did not consent to broad sharing of the research data. Nonetheless, numerical data underlying the regression analyses may be provided upon reasonable request to the corresponding author.

Conflicts of Interest

DG has an ownership stake in CHESS Health, a digital health company. This relationship is managed by the University of Wisconsin–Madison. All other authors declare no other conflicts of interest.

Multimedia Appendix 1

Extra regression models on the Positive and Negative Affect Schedule–Short Form.

[\[DOCX File, 27 KB-Multimedia Appendix 1\]](#)

References

1. Key substance use and mental health indicators in the United States: results from the 2018 national survey on drug use and health. Samhsa.gov. 2019. URL: <https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf> [accessed 2023-06-06]
2. Chih MY, Patton T, McTavish FM, Isham AJ, Judkins-Fisher CL, Atwood AK, et al. Predictive modeling of addiction lapses in a mobile health application. *J Subst Abuse Treat* 2014 Jan;46(1):29-35 [FREE Full text] [doi: [10.1016/j.jsat.2013.08.004](https://doi.org/10.1016/j.jsat.2013.08.004)] [Medline: [24035143](https://pubmed.ncbi.nlm.nih.gov/24035143/)]
3. Paliwal P, Hyman SM, Sinha R. Craving predicts time to cocaine relapse: further validation of the now and brief versions of the cocaine craving questionnaire. *Drug Alcohol Depend* 2008 Mar 01;93(3):252-259 [FREE Full text] [doi: [10.1016/j.drugalcdep.2007.10.002](https://doi.org/10.1016/j.drugalcdep.2007.10.002)] [Medline: [18063320](https://pubmed.ncbi.nlm.nih.gov/18063320/)]
4. Farvolden P, Cunningham J, Selby P. Using e-health programs to overcome barriers to the effective treatment of mental health and addiction problems. *J Technol Hum Serv* 2009 Feb 03;27(1):5-22 [FREE Full text] [doi: [10.1080/15228830802458889](https://doi.org/10.1080/15228830802458889)]
5. Green-Hamann S, Campbell EK, Sherblom J. An exploration of why people participate in second life social support groups. *J Comput Mediat Commun* 2011 Jul 05;16(4):465-491 [FREE Full text] [doi: [10.1111/j.1083-6101.2011.01543.x](https://doi.org/10.1111/j.1083-6101.2011.01543.x)]
6. Gustafson DH, McTavish FM, Chih MY, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. *JAMA Psychiatry* 2014 May;71(5):566-572 [FREE Full text] [doi: [10.1001/jamapsychiatry.2013.4642](https://doi.org/10.1001/jamapsychiatry.2013.4642)] [Medline: [24671165](https://pubmed.ncbi.nlm.nih.gov/24671165/)]
7. Yang F, Shah DV, Tahk A, Vjorn O, Dietz S, Pe-Romashko K, et al. mHealth and social mediation: Mobile support among stigmatized people living with HIV and substance use disorder. *New Media & Society* 2023 Apr 17;25(4):702-731 [doi: [10.1177/14614448231158653](https://doi.org/10.1177/14614448231158653)]
8. Aharonovich E, Amrhein PC, Bisaga A, Nunes EV, Hasin DS. Cognition, commitment language, and behavioral change among cocaine-dependent patients. *Psychol Addict Behav* 2008 Dec;22(4):557-562 [FREE Full text] [doi: [10.1037/a0012971](https://doi.org/10.1037/a0012971)] [Medline: [19071981](https://pubmed.ncbi.nlm.nih.gov/19071981/)]
9. Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. *J Med Internet Res* 2021 May 04;23(5):e15708 [FREE Full text] [doi: [10.2196/15708](https://doi.org/10.2196/15708)] [Medline: [33944788](https://pubmed.ncbi.nlm.nih.gov/33944788/)]
10. Adler DA, Wang F, Mohr DC, Choudhury T. Machine learning for passive mental health symptom prediction: generalization across different longitudinal mobile sensing studies. *PLoS One* 2022 Apr 27;17(4):e0266516 [FREE Full text] [doi: [10.1371/journal.pone.0266516](https://doi.org/10.1371/journal.pone.0266516)] [Medline: [35476787](https://pubmed.ncbi.nlm.nih.gov/35476787/)]
11. Elbattah M, Arnaud E, Gignon M, Dequen G. The role of text analytics in healthcare: a review of recent developments and applications. In: *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies*. 2021 Presented at: BIOSTEC '21; January 19-21, 2018; Madeira, Portugal p. 825-832 URL: <https://www.scitepress.org/Papers/2021/104145/104145.pdf> [doi: [10.5220/0010414508250832](https://doi.org/10.5220/0010414508250832)]
12. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoticiu-Pietro D, et al. Facebook language predicts depression in medical records. *Proc Natl Acad Sci U S A* 2018 Oct 30;115(44):11203-11208 [FREE Full text] [doi: [10.1073/pnas.1802331115](https://doi.org/10.1073/pnas.1802331115)] [Medline: [30322910](https://pubmed.ncbi.nlm.nih.gov/30322910/)]
13. Chancellor S, Nitzburg G, Hu A, Zampieri F, De Choudhury M. Discovering alternative treatments for opioid use recovery using social media. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, Scotland p. 1-15 URL: <https://dl.acm.org/doi/10.1145/3290605.3300354> [doi: [10.1145/3290605.3300354](https://doi.org/10.1145/3290605.3300354)]
14. Kalyanam J, Katsuki T, R G Lanckriet G, Mackey TK. Exploring trends of nonmedical use of prescription drugs and polydrug abuse in the Twittersphere using unsupervised machine learning. *Addict Behav* 2017 Feb;65:289-295 [doi: [10.1016/j.addbeh.2016.08.019](https://doi.org/10.1016/j.addbeh.2016.08.019)] [Medline: [27568339](https://pubmed.ncbi.nlm.nih.gov/27568339/)]
15. Tamersoy A, De Choudhury M, Chau DH. Characterizing smoking and drinking abstinence from social media. *HT ACM Conf Hypertext Soc Media* 2015 Sep;2015:139-148 [FREE Full text] [doi: [10.1145/2700171.2791247](https://doi.org/10.1145/2700171.2791247)] [Medline: [26640831](https://pubmed.ncbi.nlm.nih.gov/26640831/)]
16. Murnane EL, Counts S. Unraveling abstinence and relapse: smoking cessation reflected in social media. In: *Proceedings of the 2014 SIGCHI Conference on Human Factors in Computing Systems*. 2014 Presented at: CHI '14; April 26-May 1, 2014; Toronto, Canada p. 1345-1354 URL: <https://dl.acm.org/doi/10.1145/2556288.2557145> [doi: [10.1145/2556288.2557145](https://doi.org/10.1145/2556288.2557145)]
17. Ernala SK, Kashiparekh KH, Bolous A, Ali A, John M Kane, Birnbaum ML, et al. A social media study on mental health status transitions surrounding psychiatric hospitalizations. *Proc ACM Hum Comput Interact* 2021 Apr;5(CSCW1):155 [FREE Full text] [doi: [10.1145/3449229](https://doi.org/10.1145/3449229)] [Medline: [36267476](https://pubmed.ncbi.nlm.nih.gov/36267476/)]
18. Liu Y, Kornfield R, Shaw BR, Shah DV, McTavish F, Gustafson DH. When support is needed: social support solicitation and provision in an online alcohol use disorder forum. *Digit Health* 2017 May;3:2055207617704274 [FREE Full text] [doi: [10.1177/2055207617704274](https://doi.org/10.1177/2055207617704274)] [Medline: [29942595](https://pubmed.ncbi.nlm.nih.gov/29942595/)]
19. Shim M, Cappella JN, Han JY. How does insightful and emotional disclosure bring potential health benefits?: Study based on online support groups for women with breast cancer. *J Commun* 2011 Jun;61(3):432-464 [FREE Full text] [doi: [10.1111/j.1460-2466.2011.01555.x](https://doi.org/10.1111/j.1460-2466.2011.01555.x)] [Medline: [25568496](https://pubmed.ncbi.nlm.nih.gov/25568496/)]

20. Moon TJ, Mathias CW, Mullen J, Karns-Wright TE, Hill-Kapturczak N, Roache JD, et al. Social support and the rehabilitation of alcohol-impaired drivers: drinking motives as moderators. *Health Commun* 2021 May;36(5):540-550 [FREE Full text] [doi: [10.1080/10410236.2020.1731914](https://doi.org/10.1080/10410236.2020.1731914)] [Medline: [32091242](https://pubmed.ncbi.nlm.nih.gov/32091242/)]
21. Cutrona CE, Russell DW. Type of social support and specific stress: toward a theory of optimal matching. In: Sarason BR, Sarason IG, Pierce GR, editors. *Social Support: An Interactional View*. Hoboken, NJ: John Wiley & Sons; 1990:319-366
22. Mo PK, Coulson NS. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group. *Cyberpsychol Behav* 2008 Jun;11(3):371-374 [doi: [10.1089/cpb.2007.0118](https://doi.org/10.1089/cpb.2007.0118)] [Medline: [18537512](https://pubmed.ncbi.nlm.nih.gov/18537512/)]
23. Wright K. Motives for communication within on - line support groups and antecedents for interpersonal use. *Commun Res Rep* 2002;19(1):89-98 [FREE Full text] [doi: [10.1080/08824090209384835](https://doi.org/10.1080/08824090209384835)]
24. Namkoong K, McLaughlin B, Yoo W, Hull SJ, Shah DV, Kim SC, et al. The effects of expression: how providing emotional support online improves cancer patients' coping strategies. *J Natl Cancer Inst Monogr* 2013 Dec;2013(47):169-174 [FREE Full text] [doi: [10.1093/jncimonographs/igt033](https://doi.org/10.1093/jncimonographs/igt033)] [Medline: [24395987](https://pubmed.ncbi.nlm.nih.gov/24395987/)]
25. Yoo W, Namkoong K, Choi M, Shah DV, Tsang S, Hong Y, et al. Giving and receiving emotional support online: communication competence as a moderator of psychosocial benefits for women with breast cancer. *Comput Human Behav* 2014 Jan;30:13-22 [FREE Full text] [doi: [10.1016/j.chb.2013.07.024](https://doi.org/10.1016/j.chb.2013.07.024)] [Medline: [24058261](https://pubmed.ncbi.nlm.nih.gov/24058261/)]
26. Kim E, Han JY, Moon TJ, Shaw B, Shah DV, McTavish FM, et al. The process and effect of supportive message expression and reception in online breast cancer support groups. *Psychooncology* 2012 May;21(5):531-540 [FREE Full text] [doi: [10.1002/pon.1942](https://doi.org/10.1002/pon.1942)] [Medline: [21416553](https://pubmed.ncbi.nlm.nih.gov/21416553/)]
27. Tracy EM, Munson MR, Peterson LT, Floersch JE. Social support: a mixed blessing for women in substance abuse treatment. *J Soc Work Pract Addict* 2010 Jul;10(3):257-282 [FREE Full text] [doi: [10.1080/1533256X.2010.500970](https://doi.org/10.1080/1533256X.2010.500970)] [Medline: [20953326](https://pubmed.ncbi.nlm.nih.gov/20953326/)]
28. Chuang KY, Yang CC. Interaction patterns of nurturant support exchanged in online health social networking. *J Med Internet Res* 2012 May 03;14(3):e54 [FREE Full text] [doi: [10.2196/jmir.1824](https://doi.org/10.2196/jmir.1824)] [Medline: [22555303](https://pubmed.ncbi.nlm.nih.gov/22555303/)]
29. Coursaris CK, Liu M. An analysis of social support exchanges in online HIV/AIDS self-help groups. *Comput Human Behav* 2009 Jul;25(4):911-918 [FREE Full text] [doi: [10.1016/j.chb.2009.03.006](https://doi.org/10.1016/j.chb.2009.03.006)]
30. Pantridge CE, Charles VA, DeHart DD, Iachini AL, Seay KD, Clone S, et al. A qualitative study of the role of peer support specialists in substance use disorder treatment: examining the types of support provided. *Alcohol Treat Q* 2016 Jun 21;34(3):337-353 [FREE Full text] [doi: [10.1080/07347324.2016.1182815](https://doi.org/10.1080/07347324.2016.1182815)]
31. Liu Y, Kornfield R, Shaw BR, Shah DV, McTavish F, Gustafson DH. Giving and receiving social support in online substance use disorder forums: how self-efficacy moderates effects on relapse. *Patient Educ Couns* 2020 Jun;103(6):1125-1133 [FREE Full text] [doi: [10.1016/j.pec.2019.12.015](https://doi.org/10.1016/j.pec.2019.12.015)] [Medline: [31901364](https://pubmed.ncbi.nlm.nih.gov/31901364/)]
32. Jourard S. *The Transparent Self*. 2nd edition. Hoboken, NJ: John Wiley & Sons; 1971.
33. Bottrill S, Pistrang N, Barker C, Worrell M. The use of therapist self-disclosure: clinical psychology trainees' experiences. *Psychother Res* 2010;20(2):165-180 [FREE Full text] [doi: [10.1080/10503300903170947](https://doi.org/10.1080/10503300903170947)]
34. Derlega VJ, Grzelak J. Appropriateness of self disclosure. In: Chelune G, editor. *Self-disclosure: Origins, Patterns, and Implications of Openness in Interpersonal Relationships*. San Francisco, CA: Jossey-Bass; 1999.
35. Bazarova NN, Choi YH. Self-disclosure in social media: extending the functional approach to disclosure motivations and characteristics on social network sites. *J Commun* 2014 Jun 19;64(4):635-657 [FREE Full text] [doi: [10.1111/jcom.12106](https://doi.org/10.1111/jcom.12106)]
36. Rains SA, Wright KB. Social support and computer-mediated communication: a state-of-the-art review and agenda for future research. *Ann Int Commun Assoc* 2016 May 23;40(1):175-211 [FREE Full text] [doi: [10.1080/23808985.2015.11735260](https://doi.org/10.1080/23808985.2015.11735260)]
37. Kornfield R, Toma CL. When do online audiences amplify benefits of self-disclosure? The role of shared experience and anticipated interactivity. *J Broadcast Electron Media* 2020;64(2):277-297 [FREE Full text] [doi: [10.1080/08838151.2020.1757366](https://doi.org/10.1080/08838151.2020.1757366)] [Medline: [33584002](https://pubmed.ncbi.nlm.nih.gov/33584002/)]
38. Wang YC, Kraut RE, Levine JM. Eliciting and receiving online support: using computer-aided content analysis to examine the dynamics of online social support. *J Med Internet Res* 2015 Apr 20;17(4):e99 [FREE Full text] [doi: [10.2196/jmir.3558](https://doi.org/10.2196/jmir.3558)] [Medline: [25896033](https://pubmed.ncbi.nlm.nih.gov/25896033/)]
39. Kiecolt-Glaser JK, McGuire L, Robles TF, Glaser R. Emotions, morbidity, and mortality: new perspectives from psychoneuroimmunology. *Annu Rev Psychol* 2002;53:83-107 [doi: [10.1146/annurev.psych.53.100901.135217](https://doi.org/10.1146/annurev.psych.53.100901.135217)] [Medline: [11752480](https://pubmed.ncbi.nlm.nih.gov/11752480/)]
40. Donnelly DA, Murray EJ. Cognitive and emotional changes in written essays and therapy interviews. *J Soc Clin Psychol* 1991 Sep;10(3):334-350 [FREE Full text] [doi: [10.1521/jscp.1991.10.3.334](https://doi.org/10.1521/jscp.1991.10.3.334)]
41. Pennebaker JW. Writing about emotional experiences as a therapeutic process. *Psychol Sci* 1997 May 01;8(3):162-166 [FREE Full text] [doi: [10.1111/j.1467-9280.1997.tb00403.x](https://doi.org/10.1111/j.1467-9280.1997.tb00403.x)]
42. Lepore SJ, Buzaglo JS, Lieberman MA, Golant M, Davey A. Standard versus prosocial online support groups for distressed breast cancer survivors: a randomized controlled trial. *BMC Cancer* 2011 Aug 25;11:379 [FREE Full text] [doi: [10.1186/1471-2407-11-379](https://doi.org/10.1186/1471-2407-11-379)] [Medline: [21867502](https://pubmed.ncbi.nlm.nih.gov/21867502/)]

43. Lewallen AC, Owen JE, Bantum EO, Stanton AL. How language affects peer responsiveness in an online cancer support group: implications for treatment design and facilitation. *Psychooncology* 2014 Jul;23(7):766-772 [[FREE Full text](#)] [doi: [10.1002/pon.3477](https://doi.org/10.1002/pon.3477)] [Medline: [24519856](#)]
44. Street RL, Gordon HS. Companion participation in cancer consultations. *Psychooncology* 2008 Mar;17(3):244-251 [doi: [10.1002/pon.1225](https://doi.org/10.1002/pon.1225)] [Medline: [17534881](#)]
45. Roos C, Bowen S, Witkiewitz K. Approach coping and substance use outcomes following mindfulness-based relapse prevention among individuals with negative affect symptomatology. *Mindfulness (N Y)* 2020 Oct;11(10):2397-2410 [[FREE Full text](#)] [doi: [10.1007/s12671-020-01456-w](https://doi.org/10.1007/s12671-020-01456-w)] [Medline: [33510820](#)]
46. Carliner H, Gary D, McLaughlin KA, Keyes KM. Trauma exposure and externalizing disorders in adolescents: results from the national comorbidity survey adolescent supplement. *J Am Acad Child Adolesc Psychiatry* 2017 Sep;56(9):755-64.e3 [[FREE Full text](#)] [doi: [10.1016/j.jaac.2017.06.006](https://doi.org/10.1016/j.jaac.2017.06.006)] [Medline: [28838580](#)]
47. Rollnick S, Miller WR. *Motivational Interviewing: Helping People Change*. 3rd edition. New York, NY: Guilford Publications; 2012.
48. Miller WR, Benefield RG, Tonigan JS. Enhancing motivation for change in problem drinking: a controlled comparison of two therapist styles. In: Hill CE, editor. *Helping Skills: The Empirical Foundation*. New York, NY: American Psychological Association; 2001:243-255
49. Amrhein PC, Miller WR, Yahne CE, Palmer M, Fulcher L. Client commitment language during motivational interviewing predicts drug use outcomes. *J Consult Clin Psychol* 2003 Oct;71(5):862-878 [doi: [10.1037/0022-006X.71.5.862](https://doi.org/10.1037/0022-006X.71.5.862)] [Medline: [14516235](#)]
50. Sarpavaara H, Koski-Jännes A. Change as a journey—clients' metaphoric change talk as an outcome predictor in initial motivational sessions with probationers. *Qual Res Psychol* 2013;10(1):86-101 [[FREE Full text](#)] [doi: [10.1080/14780887.2011.586256](https://doi.org/10.1080/14780887.2011.586256)]
51. D'Amico EJ, Houck JM, Hunter SB, Miles JN, Osilla KC, Ewing BA. Group motivational interviewing for adolescents: change talk and alcohol and marijuana outcomes. *J Consult Clin Psychol* 2015 Feb;83(1):68-80 [[FREE Full text](#)] [doi: [10.1037/a0038155](https://doi.org/10.1037/a0038155)] [Medline: [25365779](#)]
52. Magill M, Janssen T, Mastroleo N, Hoadley A, Walthers J, Barnett N, et al. Motivational interviewing technical process and moderated relational process with underage young adult heavy drinkers. *Psychol Addict Behav* 2019 Mar;33(2):128-138 [[FREE Full text](#)] [doi: [10.1037/adb0000440](https://doi.org/10.1037/adb0000440)] [Medline: [30640505](#)]
53. Shaw BR, Hawkins R, McTavish F, Pingree S, Gustafson DH. Effects of insightful disclosure within computer mediated support groups on women with breast cancer. *Health Commun* 2006;19(2):133-142 [doi: [10.1207/s15327027hc1902_5](https://doi.org/10.1207/s15327027hc1902_5)] [Medline: [16548704](#)]
54. Lyons EJ, Mehl MR, Pennebaker JW. Pro-anorexics and recovering anorexics differ in their linguistic Internet self-presentation. *J Psychosom Res* 2006 Mar;60(3):253-256 [doi: [10.1016/j.jpsychores.2005.07.017](https://doi.org/10.1016/j.jpsychores.2005.07.017)] [Medline: [16516656](#)]
55. DeSteno D, Bartlett MY, Baumann J, Williams LA, Dickens L. Gratitude as moral sentiment: emotion-guided cooperation in economic exchange. *Emotion* 2010 Apr;10(2):289-293 [doi: [10.1037/a0017883](https://doi.org/10.1037/a0017883)] [Medline: [20364907](#)]
56. Ghandeharioun A, Azaria A, Taylor S, Picard RW. "Kind and grateful": a context-sensitive smartphone app utilizing inspirational content to promote gratitude. *Psychol Well Being* 2016;6:9 [[FREE Full text](#)] [doi: [10.1186/s13612-016-0046-2](https://doi.org/10.1186/s13612-016-0046-2)] [Medline: [27441170](#)]
57. D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Escobar Lee K, Cavazos-Rehg PA. Social networking online to recover from opioid use disorder: a study of community interactions. *Drug Alcohol Depend* 2017 Dec 01;181:5-10 [[FREE Full text](#)] [doi: [10.1016/j.drugalcdep.2017.09.010](https://doi.org/10.1016/j.drugalcdep.2017.09.010)] [Medline: [29024875](#)]
58. Setoyama Y, Yamazaki Y, Namayama K. Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. *J Med Internet Res* 2011 Dec 29;13(4):e122 [[FREE Full text](#)] [doi: [10.2196/jmir.1696](https://doi.org/10.2196/jmir.1696)] [Medline: [22204869](#)]
59. Finn J. An exploration of helping processes in an online self-help group focusing on issues of disability. *Health Soc Work* 1999 Aug;24(3):220-231 [doi: [10.1093/hsw/24.3.220](https://doi.org/10.1093/hsw/24.3.220)] [Medline: [10505283](#)]
60. Malik SH, Coulson NS. Coping with infertility online: an examination of self-help mechanisms in an online infertility support group. *Patient Educ Couns* 2010 Nov;81(2):315-318 [doi: [10.1016/j.pec.2010.01.007](https://doi.org/10.1016/j.pec.2010.01.007)] [Medline: [20144521](#)]
61. Braithwaite DO, Waldron VR, Finn J. Communication of social support in computer-mediated groups for people with disabilities. *Health Commun* 1999;11(2):123-151 [[FREE Full text](#)] [doi: [10.1207/s15327027hc1102_2](https://doi.org/10.1207/s15327027hc1102_2)] [Medline: [16370973](#)]
62. Buck B, Hester N. "Just like someone without mental illness, only more so": normalizing beliefs and their buffering effects on psychiatric symptoms. *J Soc Clin Psychol* 2018 Jan;37(1):69-82 [[FREE Full text](#)] [doi: [10.1521/jscp.2018.37.1.69](https://doi.org/10.1521/jscp.2018.37.1.69)]
63. Dowla R, Sinmaz H, Mavros Y, Murnion B, Cayanan E, Rooney K. The effectiveness of exercise as an adjunct intervention to improve quality of life and mood in substance use disorder: a systematic review. *Subst Use Misuse* 2022;57(6):911-928 [doi: [10.1080/10826084.2022.2052098](https://doi.org/10.1080/10826084.2022.2052098)] [Medline: [35354366](#)]
64. Mi RZ, Kornfield R, Shah DV, Maus A, Gustafson DH. Intraindividual, dyadic, and network communication in a digital health intervention: distinguishing message exposure from message production. *Health Commun* 2022 Apr;37(4):397-408 [[FREE Full text](#)] [doi: [10.1080/10410236.2020.1846273](https://doi.org/10.1080/10410236.2020.1846273)] [Medline: [33238732](#)]

65. Quanbeck A, Gustafson DH, Marsch LA, Chih MY, Kornfield R, McTavish F, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. *J Med Internet Res* 2018 Jan 30;20(1):e37 [FREE Full text] [doi: [10.2196/jmir.8928](https://doi.org/10.2196/jmir.8928)] [Medline: [29382624](https://pubmed.ncbi.nlm.nih.gov/29382624/)]
66. Quanbeck AR, Gustafson DH, Marsch LA, McTavish F, Brown RT, Mares ML, et al. Integrating addiction treatment into primary care using mobile health technology: protocol for an implementation research study. *Implement Sci* 2014 May 29;9:65 [FREE Full text] [doi: [10.1186/1748-5908-9-65](https://doi.org/10.1186/1748-5908-9-65)] [Medline: [24884976](https://pubmed.ncbi.nlm.nih.gov/24884976/)]
67. O'Leary K, Schueller SM, Wobbrock JO, Pratt WM. "Suddenly, we got to become therapists for each other": designing peer support chats for mental health. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, Canada p. 1-14 URL: <https://dl.acm.org/doi/10.1145/3173574.3173905> [doi: [10.1145/3173574.3173905](https://doi.org/10.1145/3173574.3173905)]
68. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. *Psychol Bull* 1985 Sep;98(2):310-357 [Medline: [3901065](https://pubmed.ncbi.nlm.nih.gov/3901065/)]
69. Lacy S, Riffe D. Sampling error and selecting intercoder reliability samples for nominal content categories. *Journal Mass Commun Q* 2016 Dec;73(4):963-973 [FREE Full text] [doi: [10.1177/107769909607300414](https://doi.org/10.1177/107769909607300414)]
70. Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. *J Biomed Inform* 2013 Dec;46(6):998-1005 [FREE Full text] [doi: [10.1016/j.jbi.2013.08.011](https://doi.org/10.1016/j.jbi.2013.08.011)] [Medline: [24025513](https://pubmed.ncbi.nlm.nih.gov/24025513/)]
71. Sajeev S, Champion S, Beleigoli A, Chew D, Reed RL, Magliano DJ, et al. Predicting Australian adults at high risk of cardiovascular disease mortality using standard risk factors and machine learning. *Int J Environ Res Public Health* 2021 Mar 19;18(6):3187 [FREE Full text] [doi: [10.3390/ijerph18063187](https://doi.org/10.3390/ijerph18063187)] [Medline: [33808743](https://pubmed.ncbi.nlm.nih.gov/33808743/)]
72. Cohan A, Young S, Yates A, Goharian N. Triaging content severity in online mental health forums. *J Assoc Inf Sci Technol* 2017 Sep 25;68(11):2675-2689 [FREE Full text] [doi: [10.1002/asi.23865](https://doi.org/10.1002/asi.23865)]
73. Kornfield R, Sarma PK, Shah DV, McTavish F, Landucci G, Pe-Romashko K, et al. Detecting recovery problems just in time: application of automated linguistic analysis and supervised machine learning to an online substance abuse forum. *J Med Internet Res* 2018 Jun 12;20(6):e10136 [FREE Full text] [doi: [10.2196/10136](https://doi.org/10.2196/10136)] [Medline: [29895517](https://pubmed.ncbi.nlm.nih.gov/29895517/)]
74. El-Hasnony IM, Elzeiki OM, Alshehri A, Salem H. Multi-label active learning-based machine learning model for heart disease prediction. *Sensors (Basel)* 2022 Feb 04;22(3):1184 [FREE Full text] [doi: [10.3390/s22031184](https://doi.org/10.3390/s22031184)] [Medline: [35161928](https://pubmed.ncbi.nlm.nih.gov/35161928/)]
75. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. University of Texas at Austin. 2015. URL: https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf?sequence=3&isAllowed=y [accessed 2023-06-05]
76. Pedregosa F, Varoquaux G, Gramfort A. Scikit-learn: machine learning in Python. arXiv. Preprint posted online January 2, 2012 [FREE Full text] [doi: [10.1145/2786984.2786995](https://doi.org/10.1145/2786984.2786995)]
77. Wang X, Zhao K, Street N. Analyzing and predicting user participations in online health communities: a social support perspective. *J Med Internet Res* 2017 Apr 24;19(4):e130 [doi: [10.2196/jmir.6834](https://doi.org/10.2196/jmir.6834)] [Medline: [28438725](https://pubmed.ncbi.nlm.nih.gov/28438725/)]
78. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002 Jun 01;16:321-357 [FREE Full text] [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]
79. McKinney W. pandas: a foundational Python library for data analysis and statistics. Deutsches Zentrum für Luft-und Raumfahrt. URL: https://www.dlr.de/sc/en/Portaldat/15/Resourcen/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf [accessed 2023-06-05]
80. Oliphant TE. *Guide to Numpy*. 2nd edition. New York, NY: Trelgol Publishing; 1999.
81. pickle — Python object serialization. Python. URL: <https://docs.python.org/3/library/pickle.html> [accessed 2023-06-05]
82. Hardeniya N, Perkins J, Chopra D, Joshi N, Mathur I. *Natural Language Processing: Python and NLTK*. Birmingham, UK: Packt Publishing; 2016.
83. Thompson ER. Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *J Cross Cult Psychol* 2007 Mar 01;38(2):227-242 [FREE Full text] [doi: [10.1177/0022022106297301](https://doi.org/10.1177/0022022106297301)]
84. Kim J, Hatfield E. Love types and subjective well-being: a cross cultural study. *Soc Behav Pers* 2004 Jan 01;32(2):173-182 [FREE Full text] [doi: [10.2224/sbp.2004.32.2.173](https://doi.org/10.2224/sbp.2004.32.2.173)]
85. Furlong MJ, You S, Renshaw TL, Smith DC, O'Malley MD. Preliminary development and validation of the social and emotional health survey for secondary school students. *Soc Indic Res* 2013 Jun 27;117(3):1011-1032 [FREE Full text] [doi: [10.1007/s11205-013-0373-0](https://doi.org/10.1007/s11205-013-0373-0)]
86. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res* 2009 Sep;18(7):873-880 [FREE Full text] [doi: [10.1007/s11136-009-9496-9](https://doi.org/10.1007/s11136-009-9496-9)] [Medline: [19543809](https://pubmed.ncbi.nlm.nih.gov/19543809/)]
87. Senaviratna NA, Cooray TM. Diagnosing multicollinearity of logistic regression model. *Asian J Probab Stat* 2019 Oct 01;5(2):1-9 [FREE Full text] [doi: [10.9734/ajpas/2019/v5i230132](https://doi.org/10.9734/ajpas/2019/v5i230132)]
88. Bergman BG, Wu W, Marsch LA, Crosier BS, DeLise TC, Hassanpour S. Associations between substance use and Instagram participation to inform social network-based screening models: multimodal cross-sectional study. *J Med Internet Res* 2020 Sep 16;22(9):e21916 [FREE Full text] [doi: [10.2196/21916](https://doi.org/10.2196/21916)] [Medline: [32936081](https://pubmed.ncbi.nlm.nih.gov/32936081/)]
89. Jones-Sanpei HA, Nance RJ. Financial capability in addiction research and clinical practice. *Subst Use Misuse* 2021;56(2):214-223 [doi: [10.1080/10826084.2020.1853776](https://doi.org/10.1080/10826084.2020.1853776)] [Medline: [33300413](https://pubmed.ncbi.nlm.nih.gov/33300413/)]

90. de Choudhury M, Counts S, Horvitz EJ. Social media as a measurement tool of depression in populations. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: WebSci '13; May 2-4, 2013; Paris, France p. 47-56 URL: <https://dl.acm.org/doi/10.1145/2464464.2464480> [doi: [10.1145/2464464.2464480](https://doi.org/10.1145/2464464.2464480)]
91. Liang OS, Chen Y, Bennett DS, Yang CC. Identifying self-management support needs for pregnant women with opioid misuse in online health communities: mixed methods analysis of web posts. *J Med Internet Res* 2021 Feb 04;23(2):e18296 [FREE Full text] [doi: [10.2196/18296](https://doi.org/10.2196/18296)] [Medline: [33538695](https://pubmed.ncbi.nlm.nih.gov/33538695/)]
92. Maslach C, Jackson SE. The measurement of experienced burnout. *J Organ Behav* 1981 Apr;2(2):99-113 [FREE Full text] [doi: [10.1002/job.4030020205](https://doi.org/10.1002/job.4030020205)]
93. Kornfield R, Toma CL, Shah DV, Moon TJ, Gustafson DH. What do you say before you relapse? How language use in a peer-to-peer online discussion forum predicts risky drinking among those in recovery. *Health Commun* 2018 Sep;33(9):1184-1193 [FREE Full text] [doi: [10.1080/10410236.2017.1350906](https://doi.org/10.1080/10410236.2017.1350906)] [Medline: [28792228](https://pubmed.ncbi.nlm.nih.gov/28792228/)]
94. Greidanus E, Everall RD. Helper therapy in an online suicide prevention community. *Br J Guid Couns* 2010 May;38(2):191-204 [FREE Full text] [doi: [10.1080/03069881003600991](https://doi.org/10.1080/03069881003600991)]
95. Risch J, Krestel R. Top comment or flop comment? Predicting and explaining user engagement in online news discussions. *Proc Int AAAI Conf Web Soc Media* 2020 May 26;14:579-589 [FREE Full text] [doi: [10.1609/icwsm.v14i1.7325](https://doi.org/10.1609/icwsm.v14i1.7325)]
96. Zhang TC, Kandampully J, Bilgihan A. Motivations for customer engagement in online co-innovation communities (OCCs): a conceptual framework. *J Hosp Tour Technol* 2015 Oct;6(3):311-328 [FREE Full text] [doi: [10.1108/jhtt-10-2014-0062](https://doi.org/10.1108/jhtt-10-2014-0062)]
97. Young C. Community management that works: how to build and sustain a thriving online health community. *J Med Internet Res* 2013 Jun 11;15(6):e119 [FREE Full text] [doi: [10.2196/jmir.2501](https://doi.org/10.2196/jmir.2501)] [Medline: [23759312](https://pubmed.ncbi.nlm.nih.gov/23759312/)]
98. Blanchard AL. Developing a sense of virtual community measure. *Cyberpsychol Behav* 2007 Dec;10(6):827-830 [doi: [10.1089/cpb.2007.9946](https://doi.org/10.1089/cpb.2007.9946)] [Medline: [18085972](https://pubmed.ncbi.nlm.nih.gov/18085972/)]
99. Hagman BT, Falk D, Litten R, Koob GF. Defining recovery from alcohol use disorder: development of an NIAAA research definition. *Am J Psychiatry* 2022 Nov 01;179(11):807-813 [doi: [10.1176/appi.ajp.21090963](https://doi.org/10.1176/appi.ajp.21090963)] [Medline: [35410494](https://pubmed.ncbi.nlm.nih.gov/35410494/)]
100. Biggiogera J, Boateng G, Hilpert P, Vowels MJ, Bodenmann G, Neysari M, et al. BERT meets LIWC: exploring state-of-the-art language models for predicting communication behavior in couples' conflict interactions. In: Proceedings of the 2021 International Conference on Multimodal Interaction. 2021 Presented at: ICMI '21; October 18-22, 2021; Montreal, Canada p. 385-389 URL: <https://dl.acm.org/doi/10.1145/3461615.3485423> [doi: [10.1145/3461615.3485423](https://doi.org/10.1145/3461615.3485423)]
101. Ramezan CA, Warner TA, Maxwell AE. Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sens* 2019 Jan 18;11(2):185 [FREE Full text] [doi: [10.3390/rs11020185](https://doi.org/10.3390/rs11020185)]
102. Allen A, Iqbal Z, Green-Saxena A, Hurtado M, Hoffman J, Mao Q, et al. Prediction of diabetic kidney disease with machine learning algorithms, upon the initial diagnosis of type 2 diabetes mellitus. *BMJ Open Diabetes Res Care* 2022 Jan;10(1):e002560 [FREE Full text] [doi: [10.1136/bmjdr-2021-002560](https://doi.org/10.1136/bmjdr-2021-002560)] [Medline: [35046014](https://pubmed.ncbi.nlm.nih.gov/35046014/)]

Abbreviations

- FQHC:** federally qualified health center
- IRR:** interrater reliability
- LIWC:** Linguistic Inquiry and Word Count
- ML:** machine learning
- OLS:** ordinary least squares
- PANAS-SF:** Positive and Negative Affect Schedule–Short Form
- PROMIS:** Patient-Reported Outcomes Measurement Information System
- SUD:** substance use disorder

Edited by T de Azevedo Cardoso; submitted 09.01.23; peer-reviewed by S Goodday, M Elbattah, B Senst, J Colditz, A Russell; comments to author 17.04.23; revised version received 06.06.23; accepted 04.07.23; published 22.08.23

Please cite as:

Yang EF, Kornfield R, Liu Y, Chih MY, Sarma P, Gustafson D, Curtin J, Shah D
Using Machine Learning of Online Expression to Explain Recovery Trajectories: Content Analytic Approach to Studying a Substance Use Disorder Forum
J Med Internet Res 2023;25:e45589
URL: <https://www.jmir.org/2023/1/e45589>
doi: [10.2196/45589](https://doi.org/10.2196/45589)
PMID:

©Ellie Fan Yang, Rachel Kornfield, Yan Liu, Ming-Yuan Chih, Prathusha Sarma, David Gustafson, John Curtin, Dhavan Shah. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 22.08.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.