Received: 23 June 2021 | Accepted: 10 March 2022

DOI: 10.1111/acer.14807

ALCOHOLISM CLINICAL & EXPERIMENTAL RESEARCH

RESEARCH ARTICLE

Using Facebook language to predict and describe excessive alcohol use

Rupa Jose[1] | Matthew Matero[2] | Garrick Sherman[1] | Brenda Curtis[3] | Salvatore Giorgi[3,4] | Hansen Andrew Schwartz[2] | Lyle H. Ungar[4,5]

[1]Positive Psychology Center, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[2]Department of Computer Science, Stony Brook University, Stony Brook, New York, USA

[3]Technology and Translational Research Unit, National Institute on Drug Abuse, Baltimore, Maryland, USA

[4]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[5]Department of Psychology, Positive Psychology Center, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Correspondence
Rupa Jose, Positive Psychology Center, University of Pennsylvania, 3600 Market Street Suite 560, Philadelphia, PA 19104, USA.
Email: rupajose@seas.upenn.edu

Funding information
National Institute on Alcohol Abuse and Alcoholism, Grant/Award Number: 1R01AA028032-01

Abstract

Background: Assessing risk for excessive alcohol use is important for applications ranging from recruitment into research studies to targeted public health messaging. Social media language provides an ecologically embedded source of information for assessing individuals who may be at risk for harmful drinking.

Methods: Using data collected on 3664 respondents from the general population, we examine how accurately language used on social media classifies individuals as at-risk for alcohol problems based on Alcohol Use Disorder Identification Test-Consumption score benchmarks.

Results: We find that social media language is moderately accurate (area under the curve = 0.75) at identifying individuals at risk for alcohol problems (i.e., hazardous drinking/alcohol use disorders) when used with models based on contextual word embeddings. High-risk alcohol use was predicted by individuals' usage of words related to alcohol, partying, informal expressions, swearing, and anger. Low-risk alcohol use was predicted by individuals' usage of social, affiliative, and faith-based words.

Conclusions: The use of social media data to study drinking behavior in the general public is promising and could eventually support primary and secondary prevention efforts among Americans whose at-risk drinking may have otherwise gone "under the radar."

KEYWORDS
excessive alcohol use, natural language processing, social media, subclinical drinking

INTRODUCTION

In the United States, excessive alcohol consumption is a serious public health problem. Estimates from the National Institute on Alcohol Abuse and Alcoholism (NIAAA, 2021) indicate that as many as 14.1 million adults have an alcohol use disorder (AUD) and around 26% of adults report engaging in binge drinking in the past month. Alcohol—a known addictive substance and carcinogen—can result in human morbidity and mortality (Baan et al., 2007). Alcohol abuse, dependence, and/or excessive consumption have specifically been linked to one's likelihood to develop cancer or diabetes, suffer from diseases of the liver, pancreas, or heart (Burns & Teesson, 2002), be diagnosed with mental disorders (i.e., drug use, mood, and anxiety disorders; Lai et al., 2015; Rehm, 2011), and die early.

National data from 2011 to 2015 implicate excessive alcohol consumption as the cause of over 95,000 preventable American deaths; shortening the lifespans of those who died by on average 29 years (Esser et al., 2020). Due to the health and public cost (CDC, 2019a) of excessive drinking, federal and international agencies have increased their investment in research/practices that help identify and assist populations at risk for harmful alcohol consumption (Community Preventive Services Task Force, 2016; WHO, 2010).

Social media language offers a cost-effective way to potentially provide outreach to and study populations at-risk for AUDs or excessive alcohol consumption.

Transformers—a new technique for the automatic processing of language—have resulted in the unprecedented improvement of core Artificial Intelligence (AI) applications including language translation, web search, and automatic speech recognition (e.g., SIRI or Alexa; Devlin et al., 2019; Rogers et al., 2020). Transformers have been found effective in processing social media language (Liu et al., 2019), but have rarely been used to understand health outcomes (Ganesan et al., 2021), especially in terms of excessive alcohol consumption. Here, we evaluate a modern transformer-based approach (which can not only identify words of interest but also uniquely captures word sense) referred to as "contextual embeddings," along with other more traditional language assessments based on word-counting methods (e.g., Linguistic Inquiry and Word Count [LIWC]; Tausczik & Pennebaker, 2010), to predict alcohol consumption from social media language.

Clinical approaches to identify people at-risk

Methods to identify individuals at-risk for AUDs or excessive drinking typically involve the use of self-report scales or clinical interviews (e.g., AUDADIS-5; Grant et al., 2015). Among the myriad of self-report scales used in research (T-ACE, TWEAK, MAST, etc.) two are widely used: the CAGE and the Alcohol Use Disorders Identification Test (AUDIT) (Connors & Volk, 2003). The CAGE questionnaire was first introduced by Mayfield et al. (1974) as a brief "sensitive alcoholism detector." It has since been used to successfully screen for harmful alcohol use among general primary care, elderly, and high-risk (e.g., HIV infected) populations (Maisto et al., 1995; Samet et al., 2004). The CAGE is however less useful at identifying "gray drinkers" (Dawson & Grant, 2011) or those whose drinking falls below the diagnostic threshold for severe alcohol use. Due to the desire for a cross-culturally validated screener capable of effectively identifying both clinical *and* subclinical drinkers, the AUDIT was developed by the World Health Organization (Saunders et al., 1993).

The AUDIT has been widely administered in primary care and general population settings inside and outside the United States to identify subclinical at-risk drinkers as well as those suffering from AUDs (Moehring et al., 2019). To create a short screen comparable to the 4-item CAGE (Ewing, 1984), the first 3 (consumption) items from the original 10-item AUDIT were used to construct the AUDIT-C (Bush et al., 1998). Compared to the full AUDIT, the AUDIT-C is better able to identify heavy drinkers that may benefit from alcohol interventions and has been found to be equally effective in detecting active AUDs among men (Aertgeerts et al., 2001; Bush et al., 1998).

Of course, even administering a short scale like the AUDIT-C to a large group requires population access and funds. This reality has resulted in most published studies on alcohol use relying on either general population data from national surveys (BRFSS, NHIS, NESARC, NHANES, etc.; CDC, 2019b) or survey data collected on select sub-groups (adolescents, college students, veterans, patients, etc.) from public or private sites like schools, colleges, hospitals, or substance use treatment facilities. We propose an alternative to such traditional methods; use social media data to help identify individuals at-risk for excessive drinking/AUDs.

Advantages of using social media

In the last few years, social media platforms have not only been used for public health surveillance (Gittelman et al., 2015; Gruebner et al., 2017) but also been recommended as the "next generation" tool to recruit, screen, and provide interventions to health vulnerable individuals (Community Preventive Services Task Force, 2016; Pedersen & Kurz, 2016). For U.S. adults, Facebook (69%) is the second most widely used social media platform, with 74% of adults who use Facebook logging in at least once a day (Gramlich, 2019). Facebook has also been found to be a preferred platform for women, young adults, low-income teenagers/adults (Gramlich, 2019; Hargittai, 2020), and mobile phone users (Blank & Lutz, 2017). Twitter, though less popular among adults (only 22% report ever using it; Gramlich, 2019), is one of the few social media platforms with consistently more African American users (Hargittai, 2020). For these reasons, researchers interested in a cost-effective way to study health in the general public have turned to social media platforms such as Facebook and Twitter.

Social media language and "likes" have been used to predict various health phenomena including depression, hospital visits, low birth weight, obesity, and life expectancy (Eichstaedt et al., 2018; Gittelman et al., 2015; Guntuku et al., 2020). Facebook and Twitter language have also been connected to alcohol use behaviors, especially in relation to college students. Results indicate that college students who use more negative emotion words, swear words, and refer to alcohol in posts are more likely to have drunk alcohol recently, report negative consequences as a result of drinking, score higher on the AUDIT, and experience alcohol cravings (Moreno et al., 2016; van Swol et al., 2020; Westgate et al., 2014). A follow-up study of AUD sufferers who completed a mobile intervention (Addiction-Comprehensive Health Enhancement Support System or A-CHESS) similarly found that swear, negative emotion, and inhibition words (like "stop" or "block") used in forum posts predicted relapse at follow-up (Kornfield et al., 2018). One study using the abbreviated AUDIT-C measure likewise found swear words and references to nightlife/sporting events positively associated with AUDIT-C total scores and expressions of friendliness, family, school, and love negatively associated with AUDIT-C total scores (Marengo et al., 2019). However, this study relied only on a small number of participants ($n = 296$) and focused only on AUDIT-C total scores as opposed to the diagnostically preferred binary measure of AUDIT-C (where "positive" is a score of 3+ for women or 4+ for men; Bush et al., 1998).

Benefits of using social media data to examine alcohol risk in the general population versus simply administering the AUDIT-C include (1) data collection ease, (2) access to pre-event data, and (3) improved accuracy when gathering sensitive health data. In terms of data acquisition, having participants share their social media data is not only quick but also allows researchers access to, in many cases, *years of data* across the United States (Curtis et al., 2018) without any administrative overhead (i.e., no need to hire/train staff to collect or input survey data). Barring the use of publicly accessible social media data (e.g., Twitter data), participants must consent to share their media content—which can be done in seconds. The historic and generic nature of social media language data also provides one with pre-study/event alcohol use data in addition to data on a variety of stigmatized health problems comorbid with excessive drinking (e.g., depression; Eichstaedt et al., 2018) without adding self-report instruments that may promote participant burnout. Not relying on one's self-reported substance use, but instead their everyday language patterns, reduces the error introduced by participants responding in socially desirable ways (Johnson & Fendrich, 2005). Even for longitudinal studies that use traditional survey instruments to assess hazardous drinking, social media data can be a valuable supplement to understanding the drinking risk levels of those who drop out of survey-based assessments.

Current study

An extension to early social media studies looking at the linguistic correlates associated with alcohol use behaviors, here we use binary AUDIT-C values (as opposed to the non-diagnostic, non-clinically relevant AUDIT-C total values used in prior research), the newest cutting-edge transformer methods (i.e., contextual embeddings), and Facebook data collected on a large sample (n = 3664) of respondents. Guided by a desire to support online identification and treatment efforts to reduce harmful drinking in the general population, we first focus on understanding if language can be used to predict excessive drinking and how different methods (traditional vs. new transformer-based approaches) used to process social media language can differentially predict alcohol risk. Based on previous research, we anticipate that language will be a fairly accurate predictor of excessive drinking with an area under the curve (AUC) value of around 0.6 to 0.7 (Eichstaedt et al., 2018; Guntuku et al., 2020). However, given that different methods of processing social media data treat language differently, we also expect to find differences in AUC estimates depending upon the technique used. To supplement these findings, for our best-fitting language model we further test how the quantity and timing of social media post data effect prediction. This is both a novel and practically relevant addition to the current study as it speaks to the data needs required for accurate predictions. Finally, to provide the linguistic insight we focus on the specific words or phrases that are associated with excessive drinking in our large sample of respondents. We expect to find positive associations between excessive drinking and negative emotion words, swear words, alcohol words, and/or references to nightlife/sporting events (Marengo et al., 2019; van Swol et al., 2020).

MATERIALS AND METHODS

Data

This study was reviewed and approved by the Institutional Review Board at the University of Pennsylvania. Data for the current study come from Qualtrics surveys and Facebook. Participants (18 years or older) were recruited through the Qualtrics panel[1] where they provided consent, completed an online survey, and were asked if they would be willing to share their Facebook data (i.e., "posts" or status updates originating from their own account) for research purposes. As part of a larger study, participants had the opportunity to take multiple surveys and were paid varying amounts depending upon the number of surveys they filled out. On average, participants were compensated $15/hour for questionnaire completion. Both datasets (Qualtrics and Facebook) are described in turn.

Qualtrics surveys completed online in 2017 (April 15, 2017 to June 28, 2017) and 2018 (March 14, 2018 to May 21, 2018) provided information on respondent demographics and health-related behaviors. Respondents who failed attention checks were dropped from the survey, resulting in an initial combined sample of 4739 people (2017 n = 3569; 2018 n = 1170). Given our focus on excessive drinking among men and women, only respondents with non-missing AUDIT-C scores and gender designations of "male" or "female" were retained (2017 n = 2768; 2018 n = 1166). A total of 65 respondents took both the 2017 and 2018 surveys; for these individuals, only their 2018 survey responses were used. This resulted in a combined survey sample N of 3869 (2017 n = 2703; 2018 n = 1166).

Qualtrics survey data were matched to the Facebook data shared by our survey respondents to create our language-based measures. Access to participants' Facebook timeline data—inclusive of all public posts *and* posts only visible to friends—was made possible using Facebook's Application Program Interface or API. Specifically, Qualtrics participants were asked to log into their Facebook accounts using their personal credentials. Once in, they were asked whether they wanted to allow our research application access to their data. For those who chose to "allow" (vs. "deny") access, our application then used Facebook's API to query all user-initiated timeline posts. To ensure adequate linguistic samples for language-based assessments (Kern et al., 2016), participants were retained only if their downloaded English posts totaled at least 500 words. This resulted in a final combined sample N of 3664 individuals who had usable social media language and self-reported survey data (2017 n = 2498; 2018 n = 1166).

Measures

Drinking behavior

Alcohol use behaviors are assessed using the 3-item AUDIT-C screen (Bush et al., 1998). Items ask one how often they drank an alcoholic beverage, how many standard drinks with alcohol they have in a typical day, and how often they have had six or more drinks on a single occasion. AUDIT-C total scores are generated by summing responses; score total values range from "0" to "12." For clinical purposes, AUDIT-C total scores can also be converted to binary scores which reflect one's likelihood of engaging in harmful drinking or suffering from an AUD (1 = high-risk drinkers; 0 = low-risk drinkers). Threshold values differ by gender (i.e., high-risk drinkers = 4+ for men; 3+ for women). All results are based on binary AUDIT-C scores.

Demographics

Participants provided information on their age (in years), gender (male or female; other was also an option but "other gender" individuals were excluded from analyses), race/ethnicity (White, Black or African American, Hispanic or Latinx, Native American or American Indian, Other race), household income (annual estimates provided in U.S. dollars), and zip code (converted to U.S. regions). Demographics were used to characterize the study sample (see Table 1) and create the train/test datasets needed for classification estimation (see Section 2.3 for details).

TABLE 1 Sample demographics ($N = 3664$)

	M (*SD*)	Percent
Age (in years)	42.85 (15.57)	
18 to 24 years		12.96%
25 to 34 years		23.94%
35 to 44 years		18.70%
45 to 64 years		33.62%
65 years or older		10.78%
Gender (Female = 1)		46.40%
Race/ethnicity		
White		79.53%
Black or African American		9.88%
Hispanic or Latinx		5.08%
Asian/Pacific Islander		2.57%
Native American or American Indian		1.28%
Other race		1.66%
Annual household income (in dollars)	55,184.24 (124,367.4)	
U.S. region[a]		
Northeast		17.04%
Midwest		23.66%
South		40.32%
West		18.98%

Note: Our sample of 3664 people includes only individuals with non-missing AUDIT-C values who identified as either "male" or "female" with usable Facebook data.

[a]Percents exclude the few participants ($n = 8$) with invalid/non-U.S. state zip codes.

Stress

For a comparison receiver operating characteristic (ROC) line, participant self-reported perceived stress is also included to showcase the difference in using Facebook language versus a survey scale (perceived stress) to predict individual excessive alcohol use. Perceived stress is measured using Cohen's 10-item perceived stress scale (Cohen, 1988; Cohen et al., 1983). Items ask participants, for example, how often they have been upset by unexpected occurrences, how often they were angered because things were outside of their control, and how often they felt they were on top of things. All items refer to the "last month" and are traditionally scored using a 5-point rating scale ranging from "Never" to "Very Often." In our survey, participants responded using a slider scale ranging from "0" (Never) to "6" (Very Often). To realign responses with Cohen's response scale, initial values were first converted to a fraction (divided by 6) and then multiplied by 4. Rounded integer values were then used to create our stress score measure. To generate our stress score measure, positively worded items were reverse coded before all 10 items were summed. Stress scores ranged from 0 to 40 (Sample: $M = 18.60$; $SD = 8.68$) with higher values indicating greater perceived stress (Cohen, 1988; Cohen et al., 1983). The decision to use Cohen's perceived stress measure for our comparison ROC curve was based on its strong psychometric properties, popularity, and the established link between stress and heavy alcohol use (Cohen, 1988; Cohen et al., 1983).

Language features

We used the Differential Language Analysis Toolkit (DLATK; Schwartz et al., 2017) to extract a set of language features (i.e., quantitative measures of language usage patterns) that have been found effective in past language-based assessments of human traits (Park et al., 2015) or states (Matero et al., 2019). These included: (a) words and phrases: counts the number of times words and phrases (i.e., two or three-word phrases like "take care" and "today is cold") appear in posts or measures the relative frequency of words and short phrases[2]; (b) Latent Dirichlet Allocation (LDA) topics: language use is grouped into 2000 different clusters based on an empirically derived or data-driven modeling approach (Blei et al., 2003; Schwartz et al., 2013); (c) LIWC categories: groups words into 73 categories (21 linguistic, 41 psychological, 6 personal concern, and 5 informal language marker categories; for additional details see Pennebaker et al., 2015) which have been found to hold relevance for social and behavioral health (Tausczik & Pennebaker, 2010); (d) contextual embeddings: mathematical vector representations of language use patterns (1024 dimensions in size) using a state-of-the-art machine learning model called RoBERTa (Liu et al., 2019).

Compared to the more traditional language-based measures (a to c), contextual embeddings have gained popularity among computational psycholinguists over the past few years given their ability to differentiate word meaning using surrounding context words.

TABLE 2 Estimated accuracy of demographics and social media language in predicting excessive alcohol use based on AUDIT-C binary values

Model	Area under the curve (AUC)
Age and gender	0.63
Words and phrases	0.67
LIWC categories	0.74
LDA topics	0.64
Three language features (words phrases, LIWC, LDA)	0.67
Contextual embeddings	0.75

Note: "Words and Phrases" model includes the count of words, bigrams, or trigrams as predictors (e.g., individual words along with groupings of two or three words that are used in a daily language like "happy birthday" or "see you later"). The words and phrases model applies a frequency occurrence filter of 0.05, specifying that only words/phrases used at least 5% of the time are to be included in analyses. LIWC and LDA stand for "Linguistic Inquiry and Word Count" and "Latent Dirichlet Allocation," respectively. For the "Contextual Embeddings" model we extracted features from layer 23 of RoBERTa-large. To allow for model comparisons, separate train and test datasets were used. This table shows the AUC generated on our test set sample ($n = 400$).

For example, contextual embeddings are capable of representing the different uses of the word "bank" in the following sentences by attending to context words (e.g., "river" and "check"): (1) I went walking to the river bank and (2) I am depositing a check at the bank. Previous approaches were not easily able to pick up on word sense (Wang et al., 2020). Additionally, the use of contextual embeddings has improved prediction performance for word-level (e.g., a word's part of speech), document-level (e.g., the sentiment/tone within a document), and human-level (e.g., demographics or mental health) outcomes (Akbik et al., 2019; Ganesan et al., 2021; Matero et al., 2019; Truşcă et al., 2020).

Prediction models

Before running our prediction models, we first had to split our data into mutually exclusive train and test sets. Doing so decreases the likelihood that our results will be artificially inflated because the observations that the model is trained on are separate from the data we test on. To be eligible for either the train or test set individuals needed to have non-missing demographic and health-relevant data.[3] The use of a test set in which observations are fixed enables one to make direct comparisons across all model variants (i.e., language features). The test set was selected so that there are roughly an equal proportion of males and females, similar distributions of sampled tercile age groups (i.e., 18 to 33, 34 to 51, and 52 to 89), and that participants' language data included at least 1000 words (this latter criterion was also applied to the train set during estimation). This resulted in a train set of 2991 people and a test set of 400 people.

To predict AUDIT-C binary drinking status, we used a series of logistic regression models that evaluate how successful our different language variants (features) are at defining a cutoff point that can distinguish between high-risk and low-risk drinkers. We also estimated drinking status using only basic demographics (age and gender) for comparison purposes. In line with prior research, we evaluate the effectiveness of our models using the AUC statistic. The ROC curve for our best–fitting language model is plotted along with our stress measure to show how varying the cutoff point/threshold value impacts the model's ability to detect a true positive (i.e., a high-risk drinker who would be classified as such by the AUDIT-C and the model) and false positive (i.e., a low-risk drinker according to the AUDIT-C whom the model identified as a high-risk drinker). In addition, for our best-fitting language model, we calculated the AUC as a function of social media language data using participant posts from the past 30, 180, and 365 days. The decision to use data from multiple time periods was to ascertain whether more recent and/or larger volumes of language data improved model predictions. Analogous to many survey scales which ask about health behaviors, we selected the past 30 days to represent the "past month" and 365 days to represent the "past year"; the past 180 days (about 6 months) were also included as a mid-point value.

Assessing language correlates

Interested in the language associated with high-risk drinkers and low-risk drinkers several word clouds were generated and reviewed for content. Language features—specifically words and phrases, LDA topics, and LIWC categories—were correlated with our binary AUDIT-C variable. Correlation coefficient (*r*) ranges were provided for all clouds. For LDA topic and LIWC category clouds, top word groupings (with Benjamini-Hochberg correction applied for multiple comparisons) are shown along with magnitude and significance statistics (i.e., *r* and *p*-values).

RESULTS

As noted in Table 1, our study participants were on average 43 years old, primarily White, and slightly more likely to be male (see table for details). Additionally, most participants indicated they had some education/professional training beyond high school (74%) and around half were currently employed at the time of the survey (51%).

In terms of our binary AUDIT-C measure, around 27% of our sample fell into our "high-risk drinkers" group while the rest were classified as "low-risk drinkers" (73%). Comparing AUDIT-C status to two other drinking items included in our surveys for a measure validity check, we found that high-risk drinkers reported more workdays (high risk: $M = 1.82$, $SD = 1.49$; low risk: $M = 0.15$, $SD = 0.46$) and weekdays (high risk: $M = 2.07$, $SD = 1.37$; low risk: $M = 0.23$, $SD = 0.58$) in which they drank alcohol compared to low-risk drinkers (both weekday and workday drinking were rated using a response scale of 0 to 4 days).

Using social media language to predict high-risk drinkers

The results of our prediction models utilizing different language features are summarized in Table 2. Looking at the AUC between our logistic regression models, we find that using the contextual embeddings extracted from the RoBERTa-large model (layer 23) performs the best (Liu et al., 2019). These embeddings are 1024-dimensional representations of aggregated language use from each person (average of all their word use). Compared to our other language features (AUC range: 0.64 to 0.74), our contextual embeddings model has the highest AUC at 0.75.[4] This means that our model has a 75% chance of correctly differentiating between a true positive and false positive. The ROC curve for our contextual embeddings model is plotted in Figure 1 to highlight the trade-off between model sensitivity and specificity. For comparison, we also include the ROC curve for the 10-item Cohen perceived stress measure (Cohen, 1988; Cohen et al., 1983). As shown, our language-based contextual embeddings model does a better job at predicting at-risk/high-risk drinking than simply using a participant's self-reported stress to predict drinking vulnerability.

For our contextual embeddings model, we also conducted some additional analyses (see Table 3) to understand whether the amount of social media data used made a meaningful difference in AUC estimates. Instead of using the entirety of one's shared post data (as done in Table 2), this time we restricted post data to include only user-initiated posts from the past 30, 180, and 365 days. As expected, we see an improvement in AUC each time we include more language history, yielding an AUC of 0.55 to 0.64 depending on the amount of language history leveraged to represent each person. Still, even when using 1 year of historical data, the AUC performance of our contextual embeddings model only matches our worst language model in Table 2 (LDA) which is reliant on the complete posting history of respondents.

What language tells us about high-risk and low-risk drinkers

Social media language associated with high-risk drinker status includes partying/going out references (e.g., "bar," "party," and "night"), informal words (e.g., "dude," "wanna," and "gonna"), swear

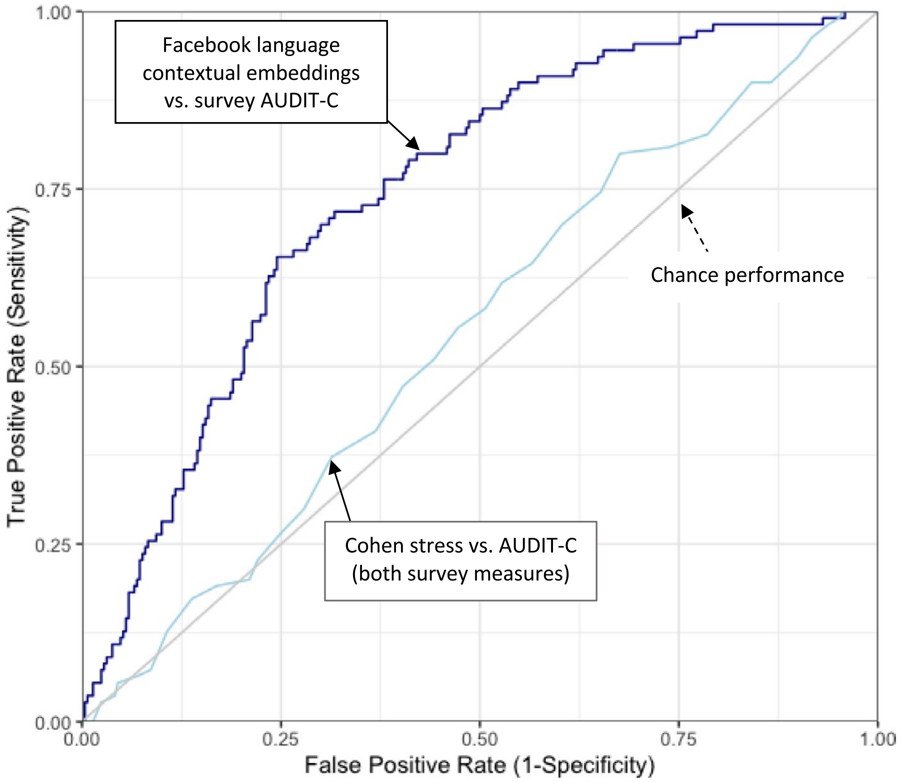

FIGURE 1 ROC curve for the best-preforming AUDIT-C classification model—the contextual embeddings model—in dark blue. For comparison, the ROC curve using the 10-item Cohen's perceived stress measure is shown in light blue. A standard reference line is shown by the solid vertical gray line

words (e.g., "fuck," "shit," and "ass"), and alcohol or drinking words (e.g., "beer," "drinking," and "wine"). By contrast, low-risk drinking status is associated with religious language (e.g., "prayers," "jesus," and "god"), relational references (e.g., "families," "them," and "those who"), future-oriented verbs (e.g., "will"), and conjunctions (e.g., "and"). These words and phrases shown in Figure 2 provide a general overview of the linguistic differences based on drinking status and are supplemented with insights provided by LIWC categories and LDA topics.

Figure 3 shows the top LIWC categories and LDA topics significantly (all $p < 0.0001$) correlated with high-risk drinking status and low-risk drinking status based on AUDIT-C threshold benchmarks.

TABLE 3 Predicting excessive alcohol use based on AUDIT-C binary values using truncated language data history

Data amount	Area under the curve (AUC)
30 days	0.55
180 days	0.62
365 days	0.64

Note: Demonstrates the performance of the contextual embeddings model when not using all data (as in Table 2). Uses only language or post data from the past 30, 180, and 365 days.

The top LIWC categories associated with high-risk drinking include informal, swear, sexual, anger, and netspeak categories (see Figure 3 for examples of LIWC words within each category; r range = 0.147 to 0.116). Though not shown, LIWC's negative emotions category (e.g., "miss," "bad," "hate," "lost," and "hell"; $r = 0.095$; $p < 0.0001$) was also associated with high-risk drinking status. The top LDA topics associated with high-risk drinking include drinking (e.g., "beer," "drinking," "cold," and "pong"), swear (e.g., "shit," "fuck," "ass," and "bullshit"), informal (e.g., "tryin," "ain't," "lookin," and "runnin"), partying (e.g., "tonight," "party," "partying," and "rockin"), and slang ("chick," "dude," and "lmao") topics (r range = 0.195 to 0.124).

For low-risk drinking status, the top LIWC categories (determined by correlation value) were religion, they, social, conjunction, and affiliation (see Figure 3 for examples of LIWC words within each category; r range = 0.113 to 0.092). LIWC's future focus category (e.g., "will," "going," "then," "hope," and "may"; $r = 0.049$; $p = 0.0108$) though not shown in Figure 3 was also significantly associated with low-risk drinking status. Finally, the top LDA topics related to low-risk drinking status all focused on religion or faith (r range = 0.134 to 0.108). There were multiple mentions of "god," "jesus," or "christ" across these topics along with words such as "prayer" and "blessed"; one topic also included mentions of the family (e.g., "daughter" and "family"). Our dominant LIWC categories and LDA topics remained

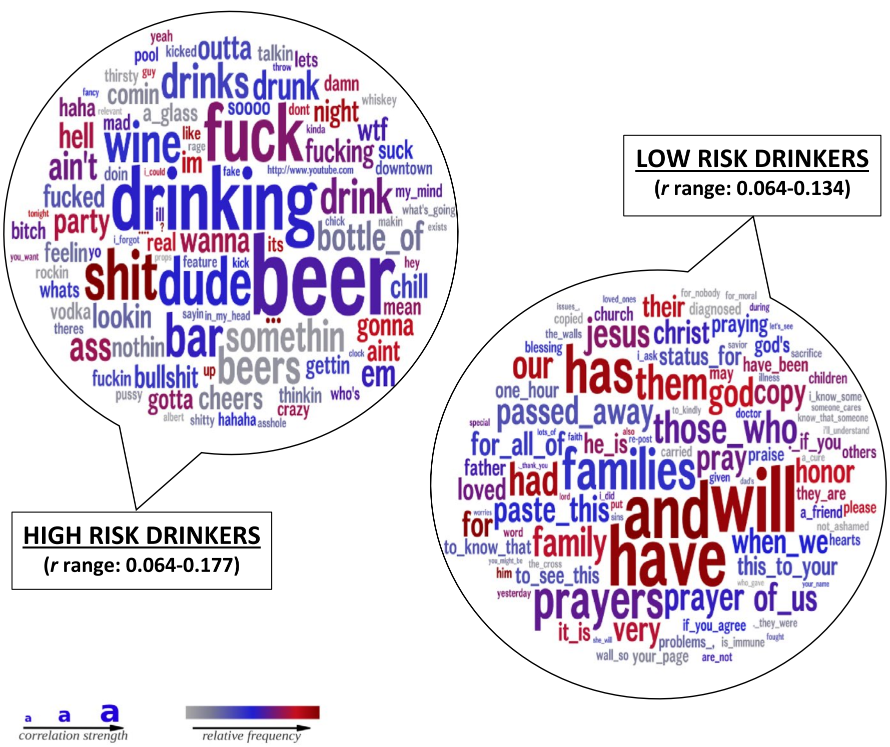

FIGURE 2 Word clouds showing the most correlated and frequent words and phrases used by individuals who are at risk for AUDs/hazardous drinking (i.e., high risk drinkers) and those who engage in low-risk drinking. Font size is indicative of correlation strength (i.e., larger words are more correlated with our drinking outcome), whereas font color symbolizes word frequency in which high-frequency words are in red, moderate-frequency words are in blue, and low-frequency words are in gray. Word clouds were generated using a frequency occurrence filter set at 0.1 (only uses the words/phrases which occur at least 10% of the time), a pmi, or pointwise mutual information of 3.0 (filters phrases/multigram features based on how commonly they appear), and by only selecting individuals whose posts have used at least 1000 words. This yielded the above word clouds, which were based on 10,904 language features collected from a total of 3392 individuals

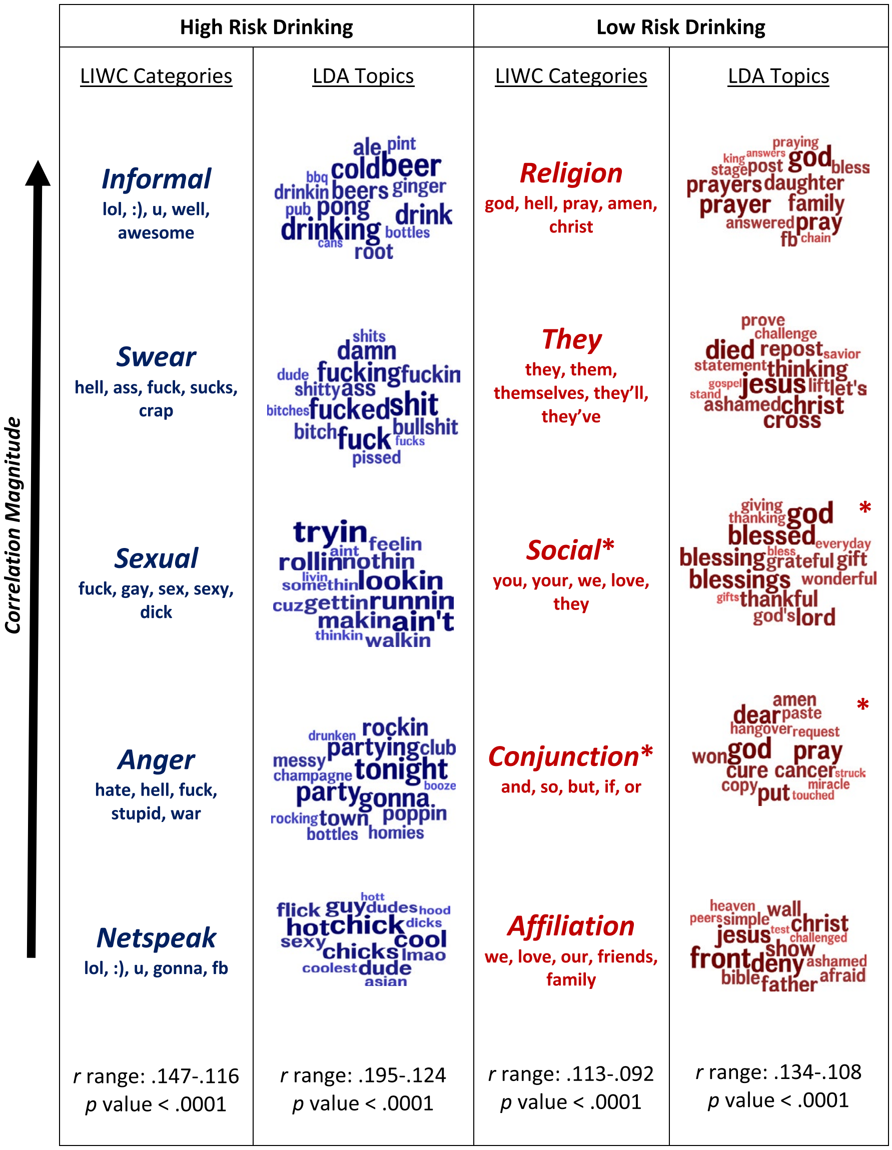

FIGURE 3 Top five LIWC categories and LDA topics associated with at-risk AUDs/hazardous drinking (i.e., high-risk drinking; blue clouds) or low-risk drinking behaviors (red clouds). Font size indicates the relative prevalence of the word within the category or topic. Categories and topics are presented in descending order (i.e., strongest correlations first). At the bottom of each column, the correlation range and *p*-values for the presented categories/topics are noted. If the correlation magnitude was identical between two stacked categories or topics, they were marked with an asterisk (*). LIWC categories were based on 73 features, LDA topics based on 2000 features (all estimated on the message data of 3392 individuals). Controlling for age and gender did not result in substantive changes in any of the LDA topics and only resulted in one LIWC category shift for "low-risk drinking." That is, with the additional controls, LIWC's function category (including words such as "the," "to," "I," "and," "a," and "you") nudged out affiliation as a top five category although affiliation remained statistically significant ($p = 0.0021$)

substantively similar even after controlling for age and gender (see descriptive text for Figure 3).

DISCUSSION

Social media data serves as a readily available, rich, and under-tapped resource to understand important public health problems, including excessive alcohol use. This study focused on determining whether language can be used to predict high-risk drinking (i.e., harmful drinking/likelihood of AUDs) according to AUDIT-C benchmarks and the specific language markers (words/phrases) that are associated with both high- and low-risk drinkers. Study findings support the use of Facebook language to help identify probable alcohol vulnerable populations in need of follow-up assessments or interventions, and note multiple language markers (informal words, religion words, etc.) that describe individuals in high/low alcohol risk groups.

When testing multiple language-based models that used participants' full post history, we found that our contextual embeddings model was most accurate or best able to differentiate high- and low-risk drinkers classified using the AUDIT-C. The contextual embeddings model outperformed models trained on traditional language features of words and phrases, LIWC categories, LDA topics, or all three feature sets combined—in addition to our age and gender and Cohen stress models. This is likely due to the sophisticated nature of a contextual embeddings model which considers not only the words but also their meaning in context (Wang et al., 2020). The fact that such models are indeed newer to the field of psycholinguistics also explains why our AUC of 0.75 exceeded the AUC values anticipated (0.6 to 0.7) based on prior literature (Eichstaedt et al., 2018; Guntuku et al., 2020). Limiting the amount of Facebook language data to only user-initiated posts in the past 30, 180, and 365 days did however correspond to a drop in predictive accuracy for our contextual embeddings model (AUC range: 0.54 to 0.64). This indicates that, when possible, using all available language data is better than only using more recent language data.

Consistent with prior studies (Kornfield et al., 2018; Marengo et al., 2019; Moreno et al., 2016; van Swol et al., 2020), we also found that high-risk drinkers not only talk more about drinking alcohol or partying/going out but they also use more informal and swear language. Our general word clouds, along with our top LIWC categories and LDA topics, echo this finding. Importantly, informal and swear language are represented differently in LIWC versus LDA groupings (see Figure 3). For example, high-risk drinkers using informal language relied on more internet/netspeak abbreviations (e.g., "lol," "fb," and ":)") when looking at top LIWC categories but more slang (e.g., "chick" or "lmao") or colloquial words (e.g., "cuz" or "ain't") when looking at top LDA topics. For high-risk drinkers using swear language, we saw multiple top LIWC categories (swear, sexual, and anger) including the word "fuck" while the top LDA swear cloud grouped that word and its variants (e.g., "fucking" and "fucked") along with other swear words like "ass," "bitch," and "bullshit" into one cloud. These differences highlight the utility of presenting both LIWC categories and LDA topics. They also suggest that when using LIWC categories, one must look beyond the category name to understand how the words subsumed within the category are being used (e.g., "sexual" category words being used to swear).

The language correlates of high-risk drinking also map onto studies examining risk factors for excessive alcohol consumption. For example, considering "behavioral under control" (Slutske et al., 2002), one could view swear words (including LIWC's sexual or anger category words) and partying/going out words as potential linguistic markers of an impulsive or sensation-seeking individual. Likewise, at the neighborhood level, we know that a greater concentration of alcohol outlets (Duncan et al., 2002; Slutske et al., 2016) is tied to greater alcohol consumption. To the extent that alcohol place-based references (e.g., "bar") are indicative of greater community alcohol access, having more spaces to drink does seem to be a key factor to consider and surveil (WHO, 2010) when trying to understand harmful or clinical alcohol use. In the absence of geo-located post data, we can only speculate on the significance of neighborhood drinking establishments. Future work however may consider collecting such data to better understand hazardous drinking within neighborhoods.

As with other studies (Marengo et al., 2019; Meyers et al., 2017) aimed at identifying attributes that protect against or promote low-risk drinking, we see that language related to religion and social ties are especially salient among low-risk drinkers. All top LDA topics centered on religion/faith-based words and the most correlated LIWC category was that of religion. Given that active engagement in religious and spiritual communities can provide members with social support, access to non-drinking activities (e.g., volunteering), and suggest behavioral norms of drinking abstinence or moderation while simultaneously discouraging other health-compromising behaviors (e.g., smoking, drug use, and risky sex; Seybold & Hill, 2001), it is not entirely surprising that low-risk drinkers appear to use more religious language over social media. The choice of using religious language in posts may also encourage healthier drinking habits by placing a spotlight on how well an individual's behavior matches their beliefs. Similar to advertising oneself as a "party animal," adopting a devout persona on social media may lead one to monitor their behavior (in and away from religious spaces) with greater scrutiny than if they instead choose to keep their spiritual beliefs off Facebook. Additional linguistic themes noted among low-risk drinkers in the general word cloud and LIWC categories were that of social/affiliative/relational language ("family," "friends," "a_friend," "we," "loved," "they," etc.), conjunctions, and a future focus.

Social ties to family or friends have long been an important risk/protective factor for adolescent and adult alcohol use (Marshall, 2014; Wang et al., 2016). Close ties to family and friends may ensure one access to a network of people invested in their health and well-being; discouraging any emerging unhealthy drinking habits before they become problematic. As social/affiliative/relational language is associated with low-risk drinkers, it is assumed that the social ties referenced in posts are not heavy drinkers themselves (an established risk factor; Nolen-Hoeksema, 2004; Wang et al., 2016; Warner et al., 2007) but agents of influence that use alcohol with appropriate restraint. Our results also showed conjunctions as a top LIWC category which appears to correspond to differences in linguistic complexity based on average Flesch–Kincaid scores (approximates reading level or difficulty). Ad-hoc analyses indicated that low-risk drinkers on average have higher Flesch–Kincaid scores than high-risk drinkers (low-risk drinkers: $M = 5.27$, $SD = 1.92$; high-risk drinkers: $M = 5.02$, $SD = 1.81$). This difference in reading level may in part explain why connector words found in more elaborate sentences such as "and" or "but" are used more by low-risk drinkers. Additionally, the word "will" which was largely used to reference future happenings consistent with the LIWC future focus category was frequently used by low-risk drinkers. This aligns well with the literature on adolescents (Klanjšek & Tement, 2019) and adults (Sansone et al., 2013), which states that holding a future orientation may lead one to be more cognizant of the consequences their actions hold and/or may serve to motivator to keep one focused on future plans or goals; ultimately corresponding to reduced substance use.

Our results suggest that clinicians and researchers interested in studying subclinical/clinical alcohol use populations should consider asking individuals for their social media language data and using contextual embeddings-based approaches to process these data. For those that consent, Facebook social media language can be used to identify, monitor, or garner pre-enrollment alcohol risk data. Researchers and clinicians can use social media data to understand current risk levels, as well as risk levels *before and after* an individual is diagnosed or treated for a substance use disorder; even for participants who have no baseline survey or clinical data available. In addition to excessive alcohol use, prior work has validated that these same data can also be used to create measures of comorbid health conditions (e.g., depression) without any extra effort by the participant. Social media language analyses can thus provide one with a non-self-report multi-marker signal of alcohol vulnerability. These data can also be used to capture changes in individual health experiences over time. Though here we used a one-time data pull, in future work researchers may choose to ask participants to consent to a long-term data share plan eliminating the need to re-contact participants for updated alcohol assessments. For highly transient or vulnerable populations where sustained participant engagement is rare and drop-out common, initially asking participants to consent to a social media language data share may improve the chance of getting future alcohol risk data while helping better understand those who drop out. Furthermore, language correlates associated with low/high-risk drinkers ("bar," "beer," "party," "god," etc.) in the present study may be used for public health campaigns and outreach. One way this could occur is using an individual's social media language to determine their alcohol risk level in real-time based on the frequency of words or phrases associated with high- or low-risk drinkers. An individual's alcohol risk level, along with other profile elements (gender, age, etc.), might then be used to determine the volume and content of preventative public health messaging. How best to deliver confidential and equitable messages or treatment over social media platforms or popular intervention apps remains an open area of study.

A few limitations exist in the present study. First, our participants were recruited from Qualtrics and are not a nationally representative sample. Compared to the 2018 U.S. adult population, our participants reported a lower annual income and were less diverse albeit similar in terms of age and sex (Census, 2019). Also, as a Qualtrics panel was used, we do not have any information on participants who declined our study. Second, we did not have clinically validated alcohol disorder data on our participants. Such data would have served as the ideal comparison to our contextual embeddings ROC curve but, in its absence, we used survey-reported Cohen stress scores. As mentioned, the response scale for the Cohen stress measure was slightly different from the standard response scale and thus additional adjustments were made to recode all values along a 5-point response scale (see Materials and Methods section). The need for such an adjustment may have introduced some imprecision into the Cohen stress ROC curve (Figure 1). However, given the vast difference in performance between both lines, the overall conclusion of the contextual embeddings model being superior remains immutable. Third, our analyses only included adult participants who agreed to share their Facebook language data,[5] posted in English, and had a sufficient word count (1000 words). As such, our findings do not generalize to all Facebook users, non-English populations, or individuals with sparse posting histories. Moreover, as most of our participants are White and middle-aged, our findings may also not generalize to younger or more diverse populations at-risk for hazardous drinking. Finally, all results are specific to the AUDIT-C. Given the popularity of the CAGE (Connors & Volk, 2003; Ewing, 1984; Mayfield et al., 1974), future studies may also consider examining the predictive accuracy and language correlates associated with the CAGE instrument.

In the United States, adults use social media regularly to connect and communicate with others (Gramlich, 2019). Serving in many ways as a digital "diary" people may share their thoughts, feelings, interests, and beliefs in their posts. For researchers and clinicians interested in gaining insight into the drinking behavior of a diverse array of Americans, the language data stored within social media sites like Facebook serve as a low-cost treasure trove. In the present study, using shared Facebook post data, we are able to predict with a relatively high degree of accuracy an individual's self-reported AUDIT-C classification. Future studies can thus consider using social media language as a pre-screening tool to recruit diverse or hard-to-reach participants (Pedersen & Kurz, 2016) into studies in which more comprehensive alcohol assessments or interventions are made available. A strength of the present study was our use of the AUDIT-C, a measure meant to flag potential clinical *and* subclinical drinkers. More attention should be paid to subclinical drinkers as they remain a vulnerable and understudied subgroup (Dawson & Grant, 2011). What we write and the words we choose are not random but serve as signals of wellbeing (Pennebaker et al., 2003; Tausczik & Pennebaker, 2010). Social media language, therefore, offers great promise in addressing the historic problem of alcohol morbidity and mortality in the United States (Baan et al., 2007; NIAAA, 2021).

ACKNOWLEDGMENTS

This research was supported by the National Institute on Alcohol Abuse and Alcoholism grant #1R01AA028032-01 and in part by the Intramural Research Program of the NIH, NIDA #ZIA-DA000628.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

ORCID

Rupa Jose https://orcid.org/0000-0002-4712-7119
Brenda Curtis https://orcid.org/0000-0002-2511-3322

ENDNOTES

[1] Potential participants are recruited into Qualtrics panels most often from different market research panels. To avoid self-selection bias, emailed survey invitations do not mention the survey content (Qualtrics, 2014).

[2] Due to the vast number of features included in our "words and phrases" measure, a filter was applied such that only those words and/or phrases used by at least a 5% of participants were included. This adjustment both reduces the noise (error) in this relative language measure and improves the computational efficiency of our analyses.

After filtering we are left with 37,977 distinct language features across all our participants.

[3] In addition to standard demographics (age, race, gender, educational attainment, income, and employment status) participants needed complete data on several alcohol use measures (days of alcohol use during the weekday/workday and number of unhealthy drinking days in the past year) and other health measures. Health measures included multiple items from the Behavioral Risk Factor Surveillance System or BRFSS survey (frequency of days where physical/mental health was "not good," frequency of days in which usual activities were missed, and frequency of social support; CDC, 2020) along with items capturing the frequency one missed work/an activity for health reasons, incidence of high blood pressure, and self-reported general health. A few people were excluded from the train sample as they were missing responses to the unhealthy drinking days ($n = 79$) item.

[4] Contextual embeddings are significantly more accurate ($p < 0.01$) than LIWC which has the second highest AUC value of 0.74.

[5] Most people (71%) who use Facebook regularly are willing to share their Facebook data for research purposes (Padrez et al., 2016).

REFERENCES

Aertgeerts, B., Buntinx, F., Ansoms, S. & Fevery, J. (2001) Screening properties of questionnaires and laboratory tests for the detection of alcohol abuse or dependence in a general practice population. *British Journal of General Practice*, 51, 206–217.

Akbik, A., Bergmann, T. & Vollgraf, R. (2019) Pooled contextualized embeddings for named entity recognition. In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, pp. 724–728.

Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Bouvard, V. et al. (2007) Carcinogenicity of alcoholic beverages. *The Lancet Oncology*, 8, 292–293.

Blank, G. & Lutz, C. (2017) Representativeness of social media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, 61, 741–756.

Blei, D.M., Ng, A.Y. & Jordan, M.I. (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Burns, L. & Teesson, M. (2002) Alcohol use disorders comorbid with anxiety, depression and drug use disorders. *Drug and Alcohol Dependence*, 68, 299–307.

Bush, K., Kivlahan, D.R., McDonell, M.B., Fihn, S.D. & Bradley, K.A. (1998) The AUDIT Alcohol Consumption Questions (AUDIT-C): an effective brief screening test for problem drinking. *Archives of Internal Medicine*, 158, 1789–1795.

CDC (2019a) *Excessive drinking is draining the U.S. economy [CDC Web site]*. Available at: https://www.cdc.gov/alcohol/features/excessive-drinking.html [Accessed 1st April 2021].

CDC (2019b) *Surveillance on alcohol [CDC Web site]*. Available at: https://www.cdc.gov/alcohol/surveillance.htm [Accessed 1st April 2021].

CDC (2020) *Behavioral risk factor surveillance system [CDC Web site]*. Available at: https://www.cdc.gov/brfss/index.html [Accessed 1st April 2021].

Census (2019) *American community survey 5-year estimates - public use microdata sample 2018*. Available at: data.census.gov [Accessed 16 October 2021].

Cohen, S. (1988) Perceived stress in a probability sample of the United States. In: Spacapan, S. & Oskamp, S. (Eds.) *The social psychology of health. The claremont symposium on applied social psychology*. Thousand Oaks, CA: Sage Publications, Inc, pp. 31–67.

Cohen, S., Kamarck, T. & Mermelstein, R. (1983) A Global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385–396.

Community Preventive Services Task Force (2016) Alcohol electronic screening and brief intervention: recommendation of the community preventive services task force. *American Journal of Preventive Medicine*, 51, 812–813.

Connors, G.J. & Volk, R.J. (2003) Self-report screening for alcohol problems among adults. In: *Assessing alcohol problems: a guide for clinicians and researchers (NIH Pub. No. 03–3745)*. 2nd ed., Washington, DC: Dept. of Health and Human Services, Public Health Service, pp.21–35.

Curtis, B., Giorgi, S., Buffone, A.E.K., Ungar, L.H., Ashford, R.D., Hemmons, J. et al. (2018) Can Twitter be used to predict county excessive alcohol consumption rates? *PLoS One*, 13, 1–16.

Dawson, D.A. & Grant, B.F. (2011) The " Gray area " of consumption between moderate and risk drinking. *Journal of Studies on Alcohol and Drugs*, 72, 453–458.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pp. 4171–4186.

Duncan, S.C., Duncan, T.E. & Strycker, L.A. (2002) A multilevel analysis of neighborhood context and youth alcohol and drug problems. *Prevention Science*, 3, 125–133.

Eichstaedt, J.C., Smith, R.J., Merchant, R.M., Ungar, L.H., Crutchley, P., Preoţiuc-Pietro, D. et al. (2018) Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 11203–11208.

Esser, M.B., Sherk, A., Liu, Y., Naimi, T.S., Stockwell, T., Stahre, M. et al. (2020) Deaths and years of potential life lost from excessive alcohol use — United States, 2011–2015. *Morbidity and Mortality Weekly Report*, 69, 1428–1433.

Ewing, J.A. (1984) Detecting alcoholism: the CAGE questionnaire. *JAMA*, 252, 1905–1907.

Ganesan, A.V., Matero, M., Ravula, A.R., Vu, H. & Schwartz, H.A. (2021) Empirical evaluation of pre-trained transformers for human-level NLP: the role of sample size and dimensionality. In: *Proceedings of the Association for Computational Linguistics*.

Gittelman, S., Lange, V., Crawford, C.A.G., Okoro, C.A., Lieb, E., Dhingra, S.S. et al. (2015) A new source of data for public health surveillance: facebook likes. *Journal of Medical Internet Research*, 17, e98.

Gramlich, J. (2019) *10 facts about Americans and Facebook [PEW Web site]*. Available at: https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook/ [Accessed 1st April 2021].

Grant, B.F., Goldstein, R.B., Smith, S.M., Jung, J., Zhang, H., Chou, S.P. et al. (2015) The alcohol use disorder and associated disabilities interview schedule-5 (AUDADIS-5): reliability of substance use and psychiatric disorder modules in a general population sample. *Drug and Alcohol Dependence*, 148, 27–33.

Gruebner, O., Lowe, S.R., Sykora, M., Shankardass, K., Subramanian, S.V. & Galea, S. (2017) A novel surveillance approach for disaster mental health. *PLoS One*, 12, 1–15.

Guntuku, S.C., Schwartz, H.A., Kashyap, A., Gaulton, J.S., Stokes, D.C., Asch, D.A. et al. (2020) Variability in language used on social media prior to hospital visits. *Scientific Reports*, 10, 1–9.

Hargittai, E. (2020) Potential biases in big data: omitted voices on social media. *Social Science Computer Review*, 38, 10–24.

Johnson, T. & Fendrich, M. (2005) Modeling sources of self-report bias in a survey of drug use epidemiology. *Annals of Epidemiology*, 15, 381–389.

Kern, M.L., Park, G., Eichstaedt, J.C., Schwartz, H.A., Sap, M., Smith, L.K. et al. (2016) Gaining insights from social media language: methodologies and challenges. *Psychological Methods*, 21, 507–525.

Klanjšek, R. & Tement, S. (2019) Alcohol and drug use among youth-the interplay of future orientation, parenting, and housing conditions. *Substance Use and Misuse*, 54, 1956–1969.

Kornfield, R., Toma, C.L., Shah, D.V., Moon, T.J. & Gustafson, D.H. (2018) 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 Communication*, 33, 1184–1193.

Lai, H.M.X., Cleary, M., Sitharthan, T. & Hunt, G.E. (2015) Prevalence of comorbid substance use, anxiety and mood disorders in epidemiological surveys, 1990–2014: a systematic review and meta-analysis. *Drug and Alcohol Dependence*, 154, 1–13.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D. et al. (2019) RoBERTa: a robustly optimized BERT pretraining approach. *arXiv*, Prepr 1907.11692.

Maisto, S.A., Connors, G.J. & Allen, J.P. (1995) Contrasting self-report screens for alcohol problems: a review. *Alcoholism, Clinical and Experimental Research*, 19, 1510–1516.

Marengo, D., Azucar, D., Giannotta, F., Basile, V. & Settanni, M. (2019) Exploring the association between problem drinking and language use on Facebook in young adults. *Heliyon*, 5, e02523.

Marshall, E.J. (2014) Adolescent alcohol use: risks and consequences. *Alcohol and Alcoholism*, 49, 160–164.

Matero, M., Idnani, A., Son, Y., Giorgi, S., Vu, H., Zamani, M. et al. (2019) Suicide risk assessment with multi-level dual-context language and BERT. In: *Proceedings of the sixth workshop on computational linguistics and clinical psychology*, pp. 39–44.

Mayfield, D., McLeod, G. & Hall, P. (1974) The CAGE questionnaire: validation of a new alcoholism screening instrument. *American Journal of Psychiatry*, 131, 1121–1123.

Meyers, J.L., Brown, Q., Grant, B.F. & Hasin, D. (2017) Religiosity, race/ethnicity, and alcohol use behaviors in the United States. *Psychological Medicine*, 47, 103–114.

Moehring, A., Rumpf, H.J., Hapke, U., Bischof, G., John, U. & Meyer, C. (2019) Diagnostic performance of the Alcohol Use Disorders Identification Test (AUDIT) in detecting DSM-5 alcohol use disorders in the general population. *Drug and Alcohol Dependence*, 204, 107530.

Moreno, M.A., Arseniev-Koehler, A., Litt, D. & Christakis, D. (2016) Evaluating college students' displayed alcohol references on Facebook and Twitter. *Journal of Adolescent Health*, 58, 527–532.

NIAAA (2021) *Alcohol facts and statistics [NIAAA Web site]*. Available at: https://www.niaaa.nih.gov/sites/default/files/publications/NIAAA_Alcohol_FactsandStats_102020_0.pdf [Accessed 1st April 2021].

Nolen-Hoeksema, S. (2004) Gender differences in risk factors and consequences for alcohol use and problems. *Clinical Psychology Review*, 24, 981–1010.

Padrez, K.A., Ungar, L., Schwartz, H.A., Smith, R.J., Hill, S., Antanavicius, T. et al. (2016) Linking social media and medical record data: a study of adults presenting to an academic, urban emergency department. *BMJ Quality & Safety*, 25, 414–423.

Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J. et al. (2015) Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108, 934–952.

Pedersen, E.R. & Kurz, J. (2016) Using Facebook for health-related research study recruitment and program delivery. *Current Opinion in Psychology*, 9, 38–43.

Pennebaker, J.W., Boyd, R.L., Jordan, K. & Blackburn, K. (2015) *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J.W., Mehl, M.R. & Niederhoffer, K.G. (2003) Psychological aspects of natural language use: our words, our selves. *Annual Review of Psychology*, 54, 547–577.

Qualtrics (2014) *ESOMAR 28: 28 questions to help research buyers of online samples*. Available at: https://success.qualtrics.com/rs/qualtrics/images/ESOMAR%2028%202014.pdf [Accessed 27th August 2021].

Rehm, J. (2011) The risks associated with alcohol use and alcoholism. *Alcohol Research and Health*, 34, 135–143.

Rogers, A., Kovaleva, O. & Rumshisky, A. (2020) A primer in BERTology: what we know about how BERT works. *Transactions of the Association for Computational Linguistics*, 8, 842–866.

Samet, J.H., Phillips, S.J., Horton, N.J., Traphagen, E.T. & Freedberg, K.A. (2004) Detecting alcohol problems in HIV-infected patients: use of the CAGE questionnaire. *AIDS Research and Human Retroviruses*, 20, 151–155.

Sansone, G., Fong, G.T., Hall, P.A., Guignard, R., Beck, F., Mons, U. et al. (2013) Time perspective as a predictor of smoking status: findings from the International Tobacco Control (ITC) surveys in Scotland, France, Germany, China, and Malaysia. *BMC Public Health*, 13, 1–8.

Saunders, J.B., Aasland, O.G., Babor, T.F., De La Fuente, J.R. & Grant, M. (1993) Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*, 88, 791–804.

Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M. et al. (2013) Personality, gender, and age in the language of social media: the open-vocabulary approach. *PLoS One*, 8, e73791.

Schwartz, H.A., Giorgi, S., Sap, M., Crutchley, P., Eichstaedt, J.C. & Ungar, L. (2017) DLATK: differential language analysis toolkit. In *Proceedings of the 2017 EMNLP system demonstrations*, pp. 55–60.

Seybold, K.S. & Hill, P.C. (2001) The role of religion and spirituality in mental and physical health. *Current Directions in Psychological Science*, 10, 21–24.

Slutske, W.S., Deutsch, A.R. & Piasecki, T.M. (2016) Neighborhood contextual factors, alcohol use, and alcohol problems in the United States: evidence from a nationally representative study of young adults. *Alcoholism, Clinical and Experimental Research*, 40, 1010–1019.

Slutske, W.S., Heath, A.C., Madden, P.A.F., Bucholz, K.K., Statham, D.J. & Martin, N.G. (2002) Personality and the genetic risk for alcohol dependence. *Journal of Abnormal Psychology*, 111, 124–133.

van Swol, L.M., Chang, C.T., Kerr, B. & Moreno, M. (2020) Linguistic predictors of problematic drinking in alcohol-related Facebook posts. *Journal of Health Communication*, 25, 214–222.

Tausczik, Y.R. & Pennebaker, J.W. (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24–54.

Truşcă, M.M., Wassenberg, D., Frasincar, F. & Dekker, R. (2020) A hybrid approach for aspect-based sentiment analysis using deep contextual word embeddings and hierarchical attention. In: *International conference on web engineering*. Springer, pp. 365–380.

Wang, C., Hipp, J.R., Butts, C.T., Jose, R. & Lakon, C.M. (2016) Coevolution of adolescent friendship networks and smoking and drinking behaviors with consideration of parental influence. *Psychology of Addictive Behaviors*, 30, 312–324.

Wang, S., Zhou, W. & Jiang, C. (2020) A survey of word embeddings based on deep learning. *Computing*, 102, 717–740.

Warner, L.A., White, H.R. & Johnson, V. (2007) Alcohol initiation experiences and family history of alcoholism as predictors of problem-drinking trajectories. *Journal of Studies on Alcohol and Drugs*, 68, 56–65.

Westgate, E.C., Neighbors, C., Heppner, H., Jahn, S. & Lindgren, K.P. (2014) "I will take a shot for every 'like' I get on this status": posting alcohol-related Facebook content is linked to drinking outcomes. *Journal of Studies on Alcohol and Drugs*, 75, 390–398.

WHO (2010) *Global strategy to reduce the harmful use of alcohol*. [WHO. Web site]. Available at: https://www.who.int/substance_abuse/activities/gsrhua/en/ [Accessed 1st April 2021].

How to cite this article: Jose, R., Matero, M., Sherman, G., Curtis, B., Giorgi, S., Schwartz, H.A., et al (2022) Using Facebook language to predict and describe excessive alcohol use. *Alcoholism: Clinical and Experimental Research*, 00, 1–12. Available from: https://doi.org/10.1111/acer.14807