Language of ADHD in Adults on Social Media

Sharath Chandra Guntuku[1], J. Russell Ramsay[1], Raina M. Merchant[1], and Lyle H. Ungar[1]

Abstract

Objective: We computationally analyze the language of social media users diagnosed with ADHD to understand what they talk about, and how their language is correlated with users' characteristics such as personality and temporal orientation. **Method:** We analyzed approximately 1.3 million tweets written by 1,399 Twitter users with self-reported diagnoses of ADHD, comparing their posts with those used by a control set matched by age, gender, and period of activity. **Results:** Users with ADHD are found to be less agreeable, more open, to post more often, and to use more negations, hedging, and swear words. Posts are suggestive of themes of emotional dysregulation, self-criticism, substance abuse, and exhaustion. A machine learning model can predict which of these Twitter users has ADHD with an out-of-sample AUC of .836. **Conclusion:** Based on this emerging technology, conjectures of future uses of social media by researchers and clinicians to better understand the naturalistic manifestations and sequelae of ADHD. *(J. of Att. Dis. XXXX; XX(X) XX-XX)*

Keywords

Adult ADHD, language assessment, computational analysis, social media

Introduction

ADHD is a condition marked by a combination of impulsiveness, hyperactivity, and inattention. About 8 million adults in the United States are known to be diagnosed with ADHD, and only one-quarter of the adults who were diagnosed when they were children are reported as having received treatment for their symptoms in the past one year (Kessler et al., 2006). Less than half of the persons with ADHD symptoms are believed to have ever received a clinical diagnosis (Chamberlain et al., 2017). Similar to other psychiatric conditions, ADHD is severely underdiagnosed in adults (Ali, Teich, Woodward, & Han, 2016). Moreover, despite the significant personal and societal burden associated with the diagnosis of ADHD, there are many contextual factors, in both the professional and community domains that contribute to the consistent underdiagnosis of ADHD (Asherson et al., 2012). In addition to educating health care professionals about appropriate screening and diagnostic evaluation tools and strategies, there is the ongoing issue of gathering corroborative data with which to document the symptoms and related impairments to the level that meet diagnostic criteria (Ramsay, 2017). Various technologies have allowed for the ecological momentary assessment (EMA) of some aspects of clinically relevant behaviors associated with ADHD that provide novel and clinically relevant insights, such as the antecedents and consequences of cigarette smoking in adults with ADHD (Mitchell et al., 2014). Although there are no such technologies that are being used in conventional diagnostic evaluations for ADHD, it is likely that such behavioral assessment technologies may someday at least be considered for use in identifying ADHD.

Social media represents an emerging technology that may provide useful signals for identifying ADHD in adolescents and adults and/or providing insights about the condition not available through traditional modes of inquiry. At the very least, it provides a new, relatively untapped avenue through which to gain naturalistic data on the day-to-day experiences and public expressions of individuals living with ADHD. With the widespread availability of internet and mobile technologies, an increasing number of teens and adults use social media every day. A recent study (Lenhart, 2015) showed that about 89% of 1,060 teens between the ages 13 and 17 years old who were interviewed used social media, with 71% of them having accounts on more than one platform. Studies have shown that several characteristics such as valence and arousal (Preoţiuc-Pietro et al., 2016); temporal orientation (Park et al., 2016); users' personality

[1]University of Pennsylvania, Philadelphia, USA

Corresponding Author:
Sharath Chandra Guntuku, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.
Email: sharathg@sas.upenn.edu

(Guntuku et al., 2017a; Schwartz et al., 2013); perception of traits such as age, gender, and personality (Guntuku, Qiu, Roy, Lin, & Jakhetiya, 2015; Preoţiuc-Pietro, Guntuku, & Ungar, 2017); and even mental health (Guntuku, Yaden, Kern, Ungar, & Eichstaedt, 2017b) can be inferred based on users' social media footprint. Moreover, social media is being increasingly used for communicating about mental health (Schein, Wilson, & Keelan, 2010). This makes social media a promising platform for researchers to study the behavior of users with ADHD, gaining potentially useful insights for researchers and clinicians not otherwise available to them, and a possible avenue for feedback to patients with ADHD about how they are doing rather than relying solely on traditional office visits.

Twitter is a widely used, free social media platform which is used by about 21% of U.S. adults. Users get to post 140 characters messages called "tweets," which are public and able to be viewed by others Twitter users by default (unless marked as private by the user). On an average, 500 million tweets are posted per day. Tweets convey the views of individuals on a variety of topics and, Twitter being a social networking platform, their reach has a great impact through social multiplier effects.

In this work, we used Twitter to analyze the language of users diagnosed with ADHD, and consequently, as a means to gain different insights about the behavioral symptoms associated with ADHD, as they are manifested on social media. Our fundamental hypothesis was that the language of people who self-identify as having ADHD would be significantly different from matched controls, that this language would reveal differences in characteristics such as personality and temporal orientation between ADHD and controls, and that the language usage patterns would both confirm existing understanding of ADHD and give new insights into the daily lives of those with ADHD. Towards this goal, we took an inductive approach of computationally analyzing the large volumes of social media data with the aim of better understanding the varying manifestations of ADHD and generating new hypotheses. This data-driven approach is not an alternative to hypothesis-driven studies; the two are complementary and iterative partners in knowledge discovery (Kell & Oliver, 2004). Data-driven approaches such as used here suggest hypotheses for confirmation (or refutation). Our expectation, based on previous work (Guntuku et al., 2017b; Park et al., 2015), is that social media language can be used to help find previously unrecognized symptoms and comorbidities.

Related Work

A number of studies have analyzed social media content on mental health–related topics including depression, suicidal ideation, and schizophrenia. A majority of these studies used a closed-vocabulary approach to analyze the language

of users (Coppersmith et al., 2015a). In this method, words are grouped into a set of categories (such as pronouns, positive words, food words, etc.) and their relative frequency within the text are obtained. Linguistic Inquiry and Word Count (LIWC; Pennebaker, Boyd, Jordan, & Blackburn, 2015) is one of the most popular implementations, and has over 60 psychologically relevant categories, such as social processes (words referring to family, friends, and humans), affective processes (positive and negative emotions, anger, sadness), perceptual processes (seeing, hearing, and feeling), and so on. However, the closed-vocabulary approach relies on scientists making conjectures in advance about which words are likely to be relevant, and often, particularly in social media, misses many words predictive of the outcome of interest (Schwartz et al., 2013).

Researchers are increasingly turning to open-vocabulary methods for analyzing social media (Schwartz et al., 2013). The Computational Linguistics and Clinical Psychology (CLPsych) workshop was started in 2014 to foster cooperation between clinical psychologists and computer scientists. Data sets were made available and "shared tasks" designed to explore and evaluate different solutions to a shared problem. In the 2015 workshop, participants were asked to predict if a user had PTSD or depression based on self-declared diagnoses (PTSD = 246, depression = 327, with the same number of age- and gender-matched controls; Coppersmith, Dredze, Harman, Hollingshead, & Mitchell, 2015b). On this data set, and other similar data sets consisting of multiple mental illness conditions, researchers have used open-vocabulary methods to predict conditions such as anxiety, depression, and PTSD based on self-declared diagnoses in English. One such work (Coppersmith, Dredze, Harman, & Hollingshead, 2015a) examined the differences in 10 comorbid mental illness conditions including ADHD. However, their analysis was based on a sample of 100 users with the condition of ADHD. While findings showed that some of the conditions show differences in generic categories of word usage such as auxiliary verbs, functional words, and those related to cognitive mechanisms, death, and so on between the users with condition and control groups, it is not clear what exactly the users with ADHD talk about, and if these differences in linguistic topics can be exploited to extract signal for ADHD to inform researchers and clinicians working in the domain. While Coppersmith and colleagues (2015a) studied comorbidity of the 10 conditions they examine, we attempt to provide specific insights into the language of ADHD.

The following are our methods for studying the language of users with ADHD on social media:

a. We take an open-vocabulary-based differential language analysis for understanding the topics that ADHD users talk about on a large-scale data set, an order of magnitude larger than previous work;
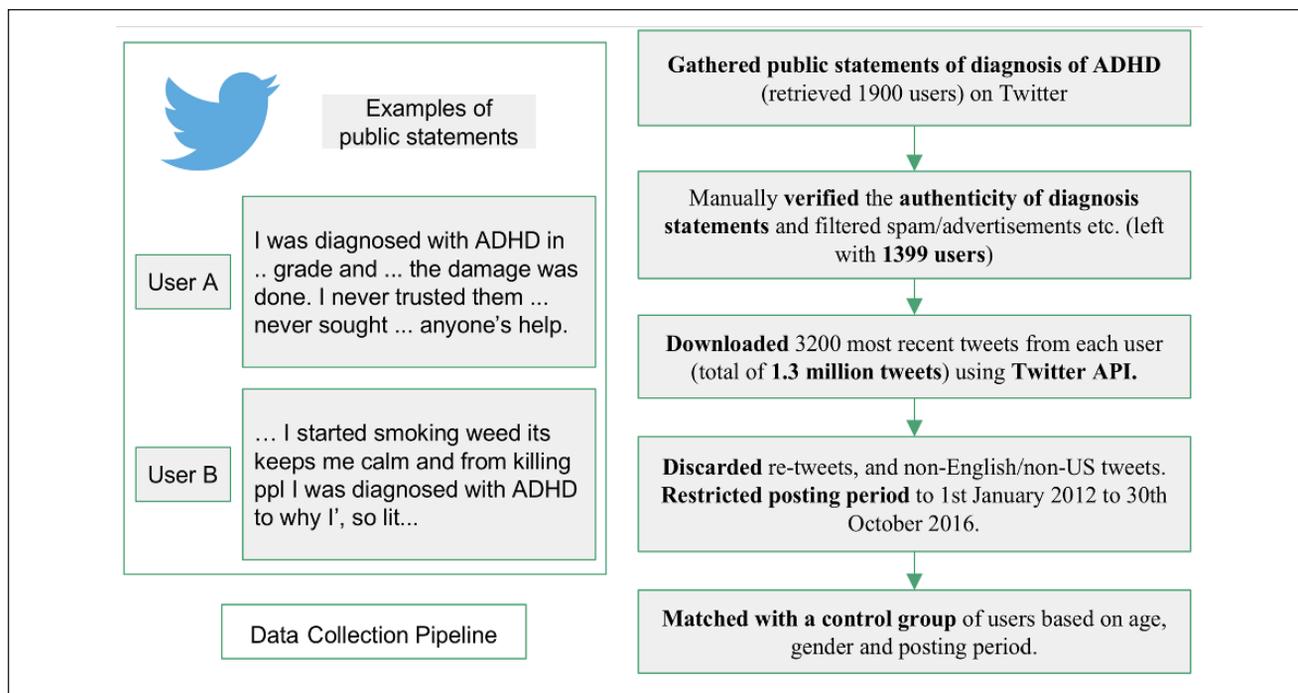
Figure 1. Overview of the data collection pipeline.

b. We study the relationship between ADHD and other traits such as users' personality, temporal orientation, and word usage;
c. We build models to predict if a user has ADHD or not based on the language they use on social media.

Materials and Method

Data

All research procedures for this study were approved by the University's Institutional Review Board. Our materials are public messages posted on Twitter. Many users self-identify their health condition on Twitter. Looking for statements such as "I am/was diagnosed with ADHD/ADD" returned about 1,900 users. Of these, 1,399 users were filtered based on verification of the tweets—removing ads, spam, messages, retweets, talking about others having ADHD, and so on. While we cannot be certain that the users were actually diagnosed with ADHD, we can only check if their statement of diagnosis appears to be genuine based on self-descriptions of a diagnosis. Previous studies indicate a good inter-rater reliability for this form of data collection (κ = .77; Coppersmith, Dredze, & Harman, 2014). We collected 1.3 million public posts between January 1, 2012 to October 30, 2016, of all users with the Twitter API. We do not have access to private messages. As we are focusing on the language of users with ADHD themselves, we excluded retweets from users. We excluded non-English tweets and

non-U.S. users to avoid cross-cultural artefacts associated with health-related discussion, and also removed the tweets that were used to assess the diagnosis. In our data set, we found 11 users who also reported a diagnosis of depression, five users who reported a diagnosis of anxiety, and four who reported a diagnosis of bipolar disorder, in addition to ADHD. We did not exclude them from our analysis as understanding the sequelae of the comorbidities might be important in identifying persons with the actual condition.

As the prevalence of mental health conditions varies based on age and gender (Dos Reis & Culotta, 2015), we formed control group of users by matching each user in our ADHD database to another user by age, gender, and period of activity. We obtained the age and gender estimates by using lexica developed by (Sap et al., 2014). Overview of the data pipeline is shown in Figure 1. Then, we selected users with a minimum of 500 words across all their posts so that we have sufficient language for analysis. This retained 1,032 users with ADHD and 1,029 users in the control group. The average age in the cohorts was 23 years, with 637 females in the ADHD group and 639 in the control group.

Linguistic Analysis

We automatically extracted the relative frequency of single words, phrases (consisting of two or three consecutive words), topics, and LIWC features (Pennebaker et al., 2015) across all tweets of all users (see Schwartz et al., 2013, for details of the method). First, all words used by less than 1%

of users were removed from analysis so as to remove uncommonly used words (outliers). Additionally, all messages with the phrase used to identify diagnosis ("I was diagnosed with ADHD") were removed prior to further analysis.

Two-hundred topics were generated using tweets across all users in the data set of ADHD and non-ADHD users using the Mallet implementation of Latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003). The topic distribution of each user aggregated across all the messages was then calculated.

We isolated the patterns in users' language using these features by correlating the words/phrases with the condition of ADHD being present or not. We used Simes *p*-correction to control the false discovery rate and use *p* < .001 for indicating meaningful correlations.

User Trait Prediction

We used automatic text-regression methods to assign to each user scores on the Big Five personality traits (Gosling, Rentfrow, & Swann, 2003) and temporal orientation scores for users. This personality model was trained on a sample of over 70,000 Facebook users, using tokens and topics extracted from status updates as features, achieving a validation predictive performance of *r* = ~.35 on average for all five traits (Schwartz et al., 2013), which is considered a high correlation in psychology, especially when measuring internal states (Meyer et al., 2001). We used a temporal orientation model (Park et al., 2015), which was trained on 6,000 Facebook and Twitter statuses with a predictive accuracy of 71.8%, for classifying users into being in the past, present, and future temporal orientation categories. We also calculated a few statistics such as number of posts that the user has, the number of followers and friends that the user has, and also the number of characters used to describe the user's profile on Twitter.

Results

In this section, we first present the results obtained from closed-vocabulary analysis (LIWC) followed by the insights derived from differential language analysis based on open-vocabulary approach (including the correlations between different user traits and the condition of ADHD), and then the prediction performance obtained by different features.

LIWC Categories

For each user, we measure the proportion of word tokens that fall into a given LIWC category. Then, we compare it against the word tokens from the control data using an empirical distribution of the proportion of language attributable to each LIWC category. The correlations between word tokens in different LIWC categories with the condition of ADHD are shown in Table 1.

Table 1. Correlations Between LIWC Categories and ADHD.

LIWC category	Words	Pearson *r*
Cognitive processes		
Insight	think, know, consider	.566
Tentativeness	maybe, perhaps, guess	.457
Differentiation	hasn't, but, else	.427
Cause	because, effect, hence	.303
Discrepancies	should, would, could	.291
Affective processes		
Anger	hate, kill, pissed	.524
Negative emotion	hate, worthless, enemy	.495
Anxiety	nervous, afraid, tense	.308
Sadness	grief, cry, sad	.113
Informal language		
Swear words	***	.502
Fillers	blah, you know, I mean	.297
Netspeak	btw, lol, thx	.165
Function words		
3rd person plural	they, their, they'd	.558
Impersonal pronouns	it, it's, those	.408
Auxiliary verbs	am, will, have	.392
Conjunctions	and, but, whereas	.357
3rd person singular	she, her, him	.321
Negations	no, not, never	.278
Prepositions	to, with, above	.244
Personal pronouns	I, them, her	.211
Articles	a, an, the	.199
1st person singular (I)	I, me, mine	.123
Other grammar		
Quantifiers	few, many, much	.323
Comparisons	greater, best, after	.293
Interrogatives	how, when, what	.26

Note. All correlations are significant at *p* < .001 and Simes corrected. LIWC = linguistic inquiry and word count.

Cognitive Processes

It is interesting to see that words indicating uncertainty and a hedging behavior are highly correlated among users with ADHD. The presence of conditional statements, along with words such as "think," "know," and so on, may be suggestive of misgivings or a sense of unpredictability among these users. For instance, they might know that they have to do something, and that doing that at that point of time is the right thing to do, but might end up not doing it, which is consistent with the executive dysfunction and motivational deficits characteristic of ADHD (Faraone et al., 2015; Ramsay & Rostain, 2016).

Affective Processes

Anxiety is a common comorbid condition with ADHD (Nigg, 2013), and words indicating the condition are highly correlated with ADHD on social media. Day-to-day aspects
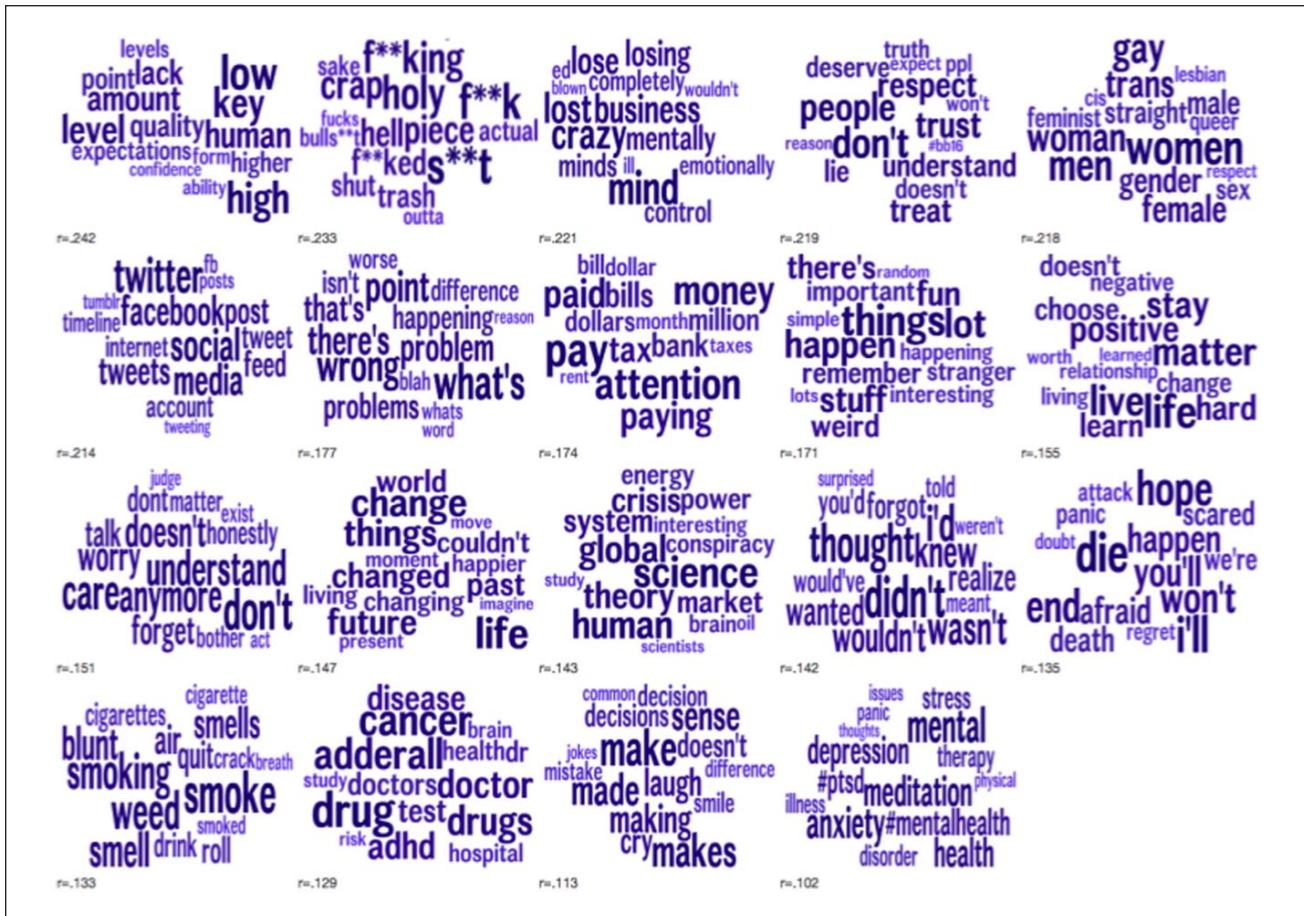
Figure 2. Words/phrases more likely to be posted by Twitter users with self-reported diagnoses of ADHD compared with the control group.
Note. Word size indicates strength of the correlation and word color indicates relative word frequency. (*p* < .01, Simes *p*-corrected).

of living, such as paying bills, or arriving at a place on time and many other such tasks are more difficult for ADHD adults to manage than for non-ADHD adults. Also, words representing anger are used more frequently, which is a known feature of emotional impulsivity for many people with ADHD (Barkley, 2010; Faraone et al., 2015; Surman et al., 2011).

Informal Language

Informal language such as fillers might be associated with insecurities of people with ADHD and/or difficulties adequately organizing and expressing ideas within the parameters of Twitter, and swearing might be a result of emotional impulsivity.

Open-Vocabulary Approach

As closed-vocabulary approaches like LIWC cover only a small subset of the entire language used on social media, we use an open-vocabulary approach to improve the coverage and find topics that people with ADHD talk about. Figure 2 shows some of the most prominent words and phrases in the messages posted by self-identified users with ADHD on Twitter. Users with ADHD swear significantly more than the control group, and also talk about others ("they," "people," "they want"), about

issues ("control," "problem_with"). Interestingly, they also talk about pokemon go.

To understand the themes underlying the language of ADHD, we then use the 200 topics created using LDA, to obtain the Pearson correlation between the topic distribution and ADHD (shown in Figure 3).

Most of the LDA topics have messages relating to lack of focus and self-regulation (Barkley, 2010), emotional dysregulation (Faraone et al., 2015; Surman et al., 2011), intention and failure, negation, self-criticism, expressions of mental, physical, and emotional exhaustion.

User Traits

We predicted users' personality traits and temporal orientation associated with their messages, and also calculated a few statistics, such as number of posts that the user has made, the number of followers and friend that the user has, and the number of characters used to describe the user's profile on Twitter. These are shown in Table 2.

Personality

We find that users with ADHD are more open (*r* = .22; Gomez & Corr, 2014; Van Dijk et al., 2017). They are also less agreeable (*r* = −.27), which corroborates our previous findings on high correlation with the LIWC category of swearing.

Figure 3. Highly correlated topics with ADHD.
Note. r below each topic represents the Pearson correlation with Simes *p*-correction and *p* < .001.

Posting Characteristics

Users with ADHD tend to post much more frequently (*r* = .53), and tend to have more followers (*r* = .4). Also, when compared with users who do not have ADHD, they tend to post significantly higher number of tweets during the night (12 a.m.–6.a.m.) as seen in Figure 4. These characteristics in posting may reflect self-regulatory difficulties in terms of engaging in a more immediately rewarding (but low priority) activity and doing so during hours typically devoted to sleep.

Temporal Orientation

Social media posts associated with users diagnosed with ADHD tend to be focused in the past with very little future-oriented content.

Use of Drug Words

We also looked the distribution of words pertaining to drugs in the posts of users with ADHD, and found that they significantly talk more about dope (*r* = .102), wax, lean, smoke, adderall, and weed which are known to be consumed by those diagnosed with ADHD (Mitchell, Sweitzer, Tunno, Kollins, & McClernon, 2016; Wilens, 2004), running the gamut from prescribed medications for treatment, recreational drug use, and recreational drugs that are viewed as self-medication for ADHD symptoms.

Predicting ADHD

We then looked at the feasibility of predicting whether a user has ADHD or not based on their social media language. Automated analysis of social media is accomplished by building predictive models, which use "features," or variables that have been extracted from social media data. In this work, we used LIWC and Topics as features. Features are then treated as independent variables in an algorithm to predict the dependent variable of an outcome of interest (e.g., users' having ADHD or not). Predictive models are trained, using an algorithm, on part of the data (the training set) and then are evaluated on the other part (the test set) to avoid overfitting—a process called cross-validation. We

Table 2. Correlations Between User Traits and ADHD.

Big Five personality	Pearson *r*
Openness	.22
Conscientiousness	−.07
Extraversion	−.04
Agreeableness	−.27
Neuroticism	−.07

Twitter user activity	Pearson *r*
Number of posts	.53
Followers count	.4
Friends count	.23
Profile description length	.18

Temporal orientation	Pearson *r*
Present	−.05
Past	.2
Future	−.16
Dope	.102

Drug words	Pearson *r*
Wax	.098
lean	.091
Smoke	.087
Adderall	.085
Weed	.077

Note. All correlations are significant at *p* < .001, Simes *p*-corrected.
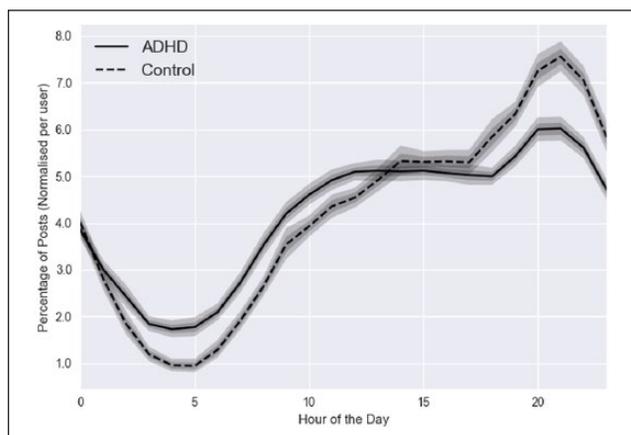
Figure 4. Proportion of posts across all users in the data set at different times (by hour) in the day.
Note. ADHD users post more during the night, when compared with the control group.

tried several algorithms such as Logistic Regression, Random Forests, Support Vector Classification, and found that Support Vector Classification showed marginally superior performance over the others. The prediction performances are then reported as one of several possible metrics

Table 3. Performance of Different Features at Predicting ADHD, Reported on an Out-of-Sample Fivefold Cross-Validation Setting.

Feature	Accuracy	F1 score	Precision	Recall	AUC
LIWC 2015	71.8	0.74	0.722	0.76	0.782
Topics	76.32	0.774	0.779	0.77	0.836
LIWC + Topics	76.1	0.774	0.775	0.773	0.833

Note. AUC = area under the curve; LIWC = linguistic inquiry and word count.
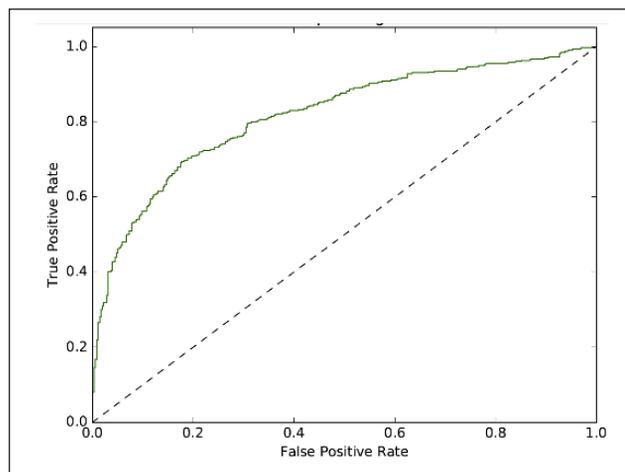

Figure 5. ROC curve of the classifier trained on LIWC + Topics.
Note. X-axis represents False Positive Rate (1 − Specificity) and Y-axis represents True Positive Rate (Sensitivity). LIWC = linguistic inquiry and word count; ROC = Receiver Operating Characteristic.

(Table 3) on an out-of-sample fivefold cross-validation setting. As there was not a huge difference in the performance, we report results only using Support Vector Classification. Topics outperform LIWC by ~5% giving the best accuracy of ~76% and an Area Under the Curve (AUC) of 0.836, which is base-rate independent. The sensitivity and specificity of the trained algorithm is shown in Figure 5.

Discussion

In this work, we studied the behavior of adults with self-reported diagnoses of ADHD as described by them on social media posts using Twitter. We compared their language with a control group using closed-vocabulary instrument (LIWC) and open-vocabulary approach (words/phrases and topics), and found significant differences in descriptions of self-efficacy, emotional dysregulation, negation, self-criticism, substance use, and expressions of mental, physical, and emotional exhaustion based on a review of linguistic themes of posts. We also found that users with ADHD are more open and less agreeable (Gomez & Corr, 2014; Van

Dijk et al., 2017), and have a significantly different posting behavior than controls in terms of number of posts and the timing of posts, and tend to have low future-orientation, posting more about the past than controls. While some of these online behaviors can be attributed to known symptoms of ADHD, we also discovered certain novel findings as highlighted in the previous section. These insights could be used by researchers and clinicians for informing hypotheses generation to understand the varying manifestations of ADHD, and social media could potentially be used as a complementary feedback tool to give patients suffering with ADHD personal insights.

One of the features of social media posts is that it provides a naturalistic expression of a user's particular mindset at that time. Hence, the linguistic analyses utilized in this study provide a novel adjunct to recent research examining the cognitions of adults with ADHD. Existing studies have reported that adults with ADHD endorse significantly more negative or maladaptive thoughts than control groups, both in cases of ADHD with and without comorbid diagnoses (Knouse, Zvorsky, & Safren, 2013; Mitchell, Benson, Knouse, Kimbrel, & Anastopoulous, 2013; Strohmeier, Rosenfield, DiTomasso, & Ramsay, 2016). A recent empirically based adult ADHD cognition scale is comprised of maladaptive positive thoughts about tasks that function as rationalizations for deferring a task, consistent with the observation of hedging themes in the user language in the current study (Knouse, Mitchell, Kimbrel, & Anastopoulos, 2017).

The more frequent use of words reflecting themes of uncertainty and hedging, as well as various descriptions of exhaustion, self-criticism, and negation (i.e., unfulfilled endeavors) may be consistent with preliminary prospective studies of the maladaptive thoughts and beliefs of adults with ADHD, with failure and related themes being the most commonly endorsed, which likely reflect the effects of the core features of ADHD on the self-efficacy of those with the diagnosis (Miklósi, Máté, Somogyi, & Szabó, 2016; Philipsen et al., 2017). Lastly, descriptors of mental, physical, and emotional exhaustion reflect the toll exacted by living with ADHD on the overall wherewithal of those affected, consistent with recent findings of elevated fatigue and lower self-efficacy in adults with ADHD compared with healthy controls (Rogers, Dittner, Rimes, & Chalder, 2017).

The posting behavior of adults with self-described ADHD is consistent with the features of the disorder. That ADHD adults posted more frequently on a platform that provides immediate and quick feedback makes sense; such a platform should be desirable to individuals who, as a group, have difficulties delaying gratification. Going beyond the data of this study, we might conjecture that social media provides a compelling source of distraction from other higher priority tasks. One of the tasks from which adults with ADHD may be prone to distraction is that of sleep, with this group being more likely to post during hours typically devoted to sleep. The temporal orientation of posts was much more likely to include content focused on the past than the present or future. An aspect of ADHD has been described as temporal myopia or difficulties planning, organizing, and efficiently working toward future-focused goals. This may be reflected in the emphasis on posting content about the past; users with ADHD may be more likely to look back on past endeavors with after-the-fact regret and frustration due to the effects of ADHD on following through on plan. Lastly, even the finding of ADHD adults using more filler words, netspeak, and other informal language may reflect difficulties related to self-expression in terms of developing, organizing, and expressing one's ideas. In sum, many of the findings herein can be viewed as being consistent with ADHD behaviors seen in other settings as well as providing novel insights into the experience and mindset of adults with ADHD in a forum not available in traditional research or clinical settings, though these are hypotheses to be tested in future studies.

While the motivation of this work is to understand the manifestation of ADHD symptoms in online spaces, it is only a preliminary step toward considering the possible role of online platforms in the detection of and intervention with social media users with ADHD. We found that social media language is predictive of ADHD with an out-of-sample accuracy of 78.5% (.836 AUC). The use of a balanced data set for our analysis (an equal number of ADHD users and control users) allowed us to investigate the emergent linguistic markers of ADHD, but is inadequate for screening detection of a condition estimated to affect less than 10% of the U.S. population. Future studies should examine the integration of both social media data and clinician judgment for ADHD screening and/or diagnosis, such as with a semistructured interview delivered by a clinician (*Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; American Psychiatric Association, 2013).

There are several other limitations of this study. While identifying individuals with ADHD who described having been diagnosed with the condition, there is no means to verify whether they were formally diagnosed by a mental health professional and, even if so, the quality of the evaluation, including identifying comorbid conditions. The nature of the Twitter platform limits users to 140 character posts, which puts a limit on expression that thereby may limit the content and manner of posts and, thus, conclusions that can be drawn, such as the necessity to use netspeak to stay within the character limit. Moreover, while using established protocols for gathering data and models of ADHD for examining the results, interpretation of the meaning and intent of word clusters used in posts in this sort of research requires a degree of empirically- and clinically informed conjecture.

The feasibility of social-media-based assessment of ADHD also raises ethical questions. Employers and insurance companies, for example, may be motivated to assess people using their social media. As ADHD may be viewed as a "mental illness" and carry social stigma and may engender discrimination, data protection and ownership frameworks are needed to make sure the data are not used against the users' interest (McKee, 2013). Few users realize the amount of mental-health-related information that can be gleaned from their digital traces, so transparency about which indicators are derived by whom for what purpose should be part of ethical and policy discourse.

The current study reflects the use of a novel source of information about the experience of adults with ADHD. It involved the examination of social media posts by adults self-described as having been diagnosed with ADHD compared with those of non-ADHD controls. These results have provided preliminary insights into the thoughts and beliefs and emotional experiences of adults living with ADHD based on the language-based analysis of content they chose to post on a public forum. These results were considered in light of their correspondence with existing extant research on adult ADHD, including that of common thoughts and beliefs endorsed by ADHD adults in light of the linguistic categories and themes from online posts in this study. It is hoped that this line of research, in general, and the findings reported in this study, specifically, offer a new and distinct avenue for developing a better understanding of adults with ADHD and potential for using social media for ongoing research as well as potential screening and clinical uses yet to be developed.

Declaration of Conflicting Interests

Funding

References

Ali, M. M., Teich, J., Woodward, A., & Han, B. (2016). The implications of the Affordable Care Act for behavioral health services utilization. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(1), 11-22.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Asherson, P., Akehurst, R., Kooij, J. S., Huss, M., Beusterien, K., Sasané, R., . . . Hodgkins, P. (2012). Under diagnosis of adult ADHD: Cultural influences and societal burden. *Journal of Attention Disorders*, *16*(Suppl. 5), 20S-38S.

Barkley, R. A. (2010). Deficient emotional self-regulation: A core component of attention-deficit/hyperactivity disorder. *Journal of ADHD & Related Disorders*, *1*, 5-37.

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

Chamberlain, S. R., Ioannidis, K., Leppink, E. W., Niaz, F., Redden, S. A., & Grant, J. E. (2017). ADHD symptoms in non-treatment seeking young adults: Relationship with other forms of impulsivity, *CNS Spectrums*, *22*, 22-30.

Coppersmith, G., Dredze, M., & Harman, C. (2014). *Quantifying mental health signals in Twitter*. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality* (pp. 51-60).

Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015a, June). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. In M. Mitchell, G. Coppersmith, & K. Hollingshead (Eds.), *CLPsych@ HLT-NAACL* (pp. 1-10). Denver, Colorado, USA: Association for Computational Linguistics.

Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., & Mitchell, M. (2015b, June). CLPsych 2015 shared task: Depression and PTSD on Twitter. In M. Mitchell, G. Coppersmith, & K. Hollingshead (Eds.), *CLPsych@ HLT-NAACL* (pp. 31-39). Denver, Colorado, USA: Association for Computational Linguistics.

Dos Reis, V. L., & Culotta, A. (2015, January). Using matched samples to estimate the effects of exercise on mental health from Twitter. In S. Zilberstein, B. Bonet, & S. Koenig (Eds.), *Proceedings of the twenty-ninth AAAI conference on artificial intelligence* (pp. 182-188). Austin, Texas, USA: Association for the Advancement of Artificial Intelligence.

Faraone, S. V., Asherson, P., Banaschewski, T., Biederman, J., Buitelaar, J. K., Ramos-Quiroga, J. A., . . . Franke, B. (2015). Attention-deficit/hyperactivity disorder. *Nature Reviews: Disease Primers*, *1*, Article 15020. doi:10.1038/nrdp.2015.20

Gomez, R., & Corr, P. J. (2014). ADHD and personality: A meta-analytic review. *Clinical Psychology Review*, *34*, 376-388. doi:10.1016/j.cpr.2014.05.002

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*, 504-528.

Guntuku, S. C., Lin, W., Carpenter, J., Ng, W. K., Ungar, L. H., & Preoțiuc-Pietro, D. (2017a, June). Studying personality through the content of posted and liked images on Twitter. In Paolo Boldi & Katharina Kinder-Kurlanda (Eds.), *Proceedings of the 2017 ACM on web science conference* (pp. 223-227). Association for Computing Machinery: Troy, New York, USA.

Guntuku, S. C., Qiu, L., Roy, S., Lin, W., & Jakhetiya, V. (2015, October). Do others perceive you as you want them to? Modeling personality based on selfies. In Mohammad Soleymani, Yi-Husan Yang, Yu-Gang Jiang, Shih-Fu Chang (Eds.), *Proceedings of the 1st international workshop on affect & sentiment in multimedia* (pp. 21-26). Brisbane, Australia: Association for Computing Machinery.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017b). Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43-49.

Lenhart, A. (2015). *Mobile access shifts social media use and other online activities*. Retrieved from http://www.pewinternet.org/2015/04/09/mobile-access-shifts-social-media-use-and-other-online-activities/

Kell, D. B., & Oliver, S. G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *BioEssays*, *26*(1), 99-105.

Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., . . . Spencer, T. (2006). The prevalence and correlates of adult ADHD in the United States: Results from the National Comorbidity Survey replication. *American Journal of psychiatry*, *163*, 716-723.

Knouse, L. E., Mitchell, J. T., Kimbrel, N. A., & Anastopoulos, A. D. (2017). Development and Evaluation of the ADHD Cognitions Scale for Adults. *Journal of Attention Disorders*, 1087054717707580.

Knouse, L. E., Zvorsky, I., & Safren, S. A. (2013). Depression in adults with attention-deficit/hyperactivity disorder (ADHD): The mediating role of cognitive-behavioral factors. *Cognitive Therapy and Research*, *37*, 1220-1232. doi:10.1007/s10608-013-9569-5

McKee, R. (2013). Ethical issues in using social media for health and health care research. *Health Policy*, *110*, 298-301.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., . . . Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist*, *56*(2), 128.

Miklósi, M., Máté, O., Somogyi, K., & Szabó, M. (2016). Adult attention deficit hyperactivity disorder symptoms, perceived stress, and well-being. *The Journal of Nervous and Mental Disease*, *204*, 364-369. doi:10.1097/NMD.0000000000000472

Mitchell, J. T., Benson, J. W., Knouse, L. E., Kimbrel, N. A., & Anastopoulous, A. D. (2013). Are negative automatic thoughts associated with ADHD in adulthood? *Cognitive Therapy and Research*, *37*, 851-859. doi:10.1007/s10608-013-9525-4

Mitchell, J. T., Dennis, M. F., English, J. S., Dennis, P. A., Birghtwood, A., Beckham, J. C., . . . Kollins, S. H. (2014). Ecological momentary assessment of antecedents and consequences of smoking in adults with Attention-Deficit/Hyperactivity Disorder. *Substance Use & Misuse*, *49*, 1446-1456. doi:10.3109/10826084.2014.912229

Mitchell, J. T., Sweitzer, M. M., Tunno, A. M., Kollins, S. H., & McClernon, F. J. (2016). "I use weed for my ADHD": A qualitative analysis of online forum discussions on cannabis use and ADHD. *PLoS ONE*, *11*(5), e0156614.

Nigg, J. T. (2013). Attention-deficit/hyperactivity disorder and adverse health outcomes. *Clinical Psychology Review*, *33*, 215-228. doi:10.1016/j.cpr.2012.11.005

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

Park, G., Schwartz, H. A., Sap, M., Kern, M. L., Weingarten, E., Eichstaedt, J. C., . . . Seligman, M. E. (2016). Living in the past, present, and future: Measuring temporal orientation with language. Journal of Personality. *Advance online publication*. doi:10.1111/jopy.12239

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

Philipsen, A., Lam, A. P., Breit, S., Lücke, C., Müller, H. H., & Matthies, S. (2017). Early maladaptive schemas in adult patients with attention deficit hyperactivity disorder. *Attention Deficit Hyperactivity Disorder*, *9*, 101-111. doi:10.1007/s12402-016-0211-8

Preoţiuc-Pietro, D., Guntuku, S. C., & Ungar, L. (2017). Controlling human perception of basic user traits. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2325-2331). Copenhagen, Denmark: Association for Computational Linguistics.

Preoţiuc-Pietro, D., Schwartz, H. A., Park, G., Eichstaedt, J. C., Kern, M., Ungar, L., . . . Shulman, E. P. (2016). Modelling valence and arousal in Facebook posts. In *Proceedings of NAACL-HLT* (pp. 9-15). Association for Computational Linguistics: San Diego, CA, USA.

Ramsay, J. R. (2017). Assessment and monitoring of treatment response in adult ADHD patients: Current perspectives. *Neuropsychiatric Disease and Treatment*, *13*, 221-232. doi:10.2147/NDT.S104706

Ramsay, J. R., & Rostain, A. L. (2016). Adult ADHD as an implementation problem: Clinical significance, underlying mechanisms, and psychosocial treatment. *Practice Innovations*, *1*, 36-52.

Rogers, D. C., Dittner, A. J., Rimes, K. A., & Chalder, T. (2017). Fatigue in an adult attention deficit hyperactivity disorder population: A trans-diagnostic approach. *British Journal of Clinical Psychology*, *56*, 33-52.doi:10.1111/bjc.12119

Sap, M., Park, G., Eichstaedt, J., Kern, M., Stillwell, D., Kosinski, M., . . . Schwartz, H. A. (2014). *Developing age and gender predictive lexica over social media*. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1146-1151). Association for Computational Linguistics: Doha, Qatar.

Schein, R., Wilson, K., & Keelan, J. E. (2010). *Literature review on effectiveness of the use of social media: A report for Peel Public Health* (Region of Peel). Peel Public Health. Internet in medicine.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., . . . Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, *8*(9), e73791.

Surman, C. B. H., Biederman, J., Spencer, T., Yorks, D., Miller, C. A., Petty, C., . . . Faraone, S. V. (2011). Deficient emotional self-regulation and adult attention deficit hyperactivity disorder: A family risk analysis. *American Journal of Psychiatry*, *168*, 617-623. doi:10.1176/appi.ajp.2010.10081172

Van Dijk, F. E., Mostert, J., Glennon, J., Onnink, M., Dammers, J., Vaquez, A. A., . . . Buitelaar, J. K. (2017). Five factor model personality traits relate to adult attention-deficit disorder but not to their distinct neurocognitive profiles. Psychiatry Research. Advance online publication. doi:10.1016/j.psychres.2017.08.037

Wilens, T. E. (2004). Impact of ADHD and its treatment on substance abuse in adults. *Journal of Clinical Psychiatry*, *65*, 38-45.

Author Biographies

Sharath Chandra Guntuku, PhD, is a postdoctoral fellow in Department of Computer and Information Science and the Perelmen School of Medicine at the University of Pennsylvania. His research aims to leverage large-scale social media image and text data to model social health outcomes and psychological traits.

J. Russell Ramsay, PhD, is a licensed psychologist and an associate professor of clinical psychology in psychiatry at the University of Pennsylvania Perelman School of Medicine, where he serves as co-director of Penn's Adult ADHD Treatment and Research Program. His primary clinical and research interests are in improving diagnostic assessment, psychosocial treatments, and other support services for adults with ADHD

Raina M. Merchant, MD, serves as the director of the Penn Medicine Center for Digital Health, an associate professor in the Department of Emergency Medicine at the University of Pennsylvania and a senior fellow at the Leonard Davis Institute of Health Economics. Raina's research focuses on innovation diffusion, social media, and resuscitation science.

Lyle H. Ungar, PhD, is a professor of computer and information science at the University of Pennsylvania, where he also holds appointments in multiple departments, including psychology. He has published over 200 articles, supervised two dozen PhD students, and is coinventor on 10 patents. His current research focuses on developing scalable machine learning methods for data mining and text mining and on the analysis of social media to better understand the drivers of physical and mental well-being.