

Original Paper

Using Twitter to Detect Psychological Characteristics of Self-Identified Persons With Autism Spectrum Disorder: A Feasibility Study

Yulin Hswen^{1,2*}, MPH; Anuraag Gopaluni^{3*}, BA; John S Brownstein^{2,4,5}, PhD; Jared B Hawkins^{2,4}, MMSc, PhD

¹Department of Social and Behavioral Sciences, Harvard TH Chan School of Public Health, Boston, MA, United States

²Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States

³Department of Mathematics and Statistics, Boston University, Boston, MA, United States

⁴Department of Pediatrics, Harvard Medical School, Boston, MA, United States

⁵Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

*these authors contributed equally

Corresponding Author:

Yulin Hswen, MPH

Department of Social and Behavioral Sciences

Harvard TH Chan School of Public Health

677 Huntington Ave

Boston, MA,

United States

Phone: 1 617 775 1889

Fax: 1 617 730 0267

Email: yuh958@mail.harvard.edu

Abstract

Background: More than 3.5 million Americans live with autism spectrum disorder (ASD). Major challenges persist in diagnosing ASD as no medical test exists to diagnose this disorder. Digital phenotyping holds promise to guide in the clinical diagnoses and screening of ASD.

Objective: This study aims to explore the feasibility of using the Web-based social media platform Twitter to detect psychological and behavioral characteristics of self-identified persons with ASD.

Methods: Data from Twitter were retrieved from 152 self-identified users with ASD and 182 randomly selected control users from March 22, 2012 to July 20, 2017. We conducted a between-group comparative textual analysis of tweets about repetitive and obsessive-compulsive behavioral characteristics typically associated with ASD. In addition, common emotional characteristics of persons with ASD, such as fear, paranoia, and anxiety, were examined between groups through textual analysis. Furthermore, we compared the timing of tweets between users with ASD and control users to identify patterns in communication.

Results: Users with ASD posted a significantly higher frequency of tweets related to the specific repetitive behavior of counting compared with control users ($P < .001$). The textual analysis of obsessive-compulsive behavioral characteristics, such as fixate, excessive, and concern, were significantly higher among users with ASD compared with the control group ($P < .001$). In addition, emotional terms related to fear, paranoia, and anxiety were tweeted at a significantly higher rate among users with ASD compared with control users ($P < .001$). Users with ASD posted a smaller proportion of tweets during time intervals of 00:00-05:59 ($P < .001$), 06:00-11:59 ($P < .001$), and 18:00-23:59 ($P < .001$), as well as a greater proportion of tweets from 12:00 to 17:59 ($P < .001$) compared with control users.

Conclusions: Social media may be a valuable resource for observing unique psychological characteristics of self-identified persons with ASD. Collecting and analyzing data from these digital platforms may afford opportunities to identify the characteristics of ASD and assist in the diagnosis or verification of ASD. This study highlights the feasibility of leveraging digital data for gaining new insights into various health conditions.

(JMIR Mhealth Uhealth 2019;7(2):e12264) doi:[10.2196/12264](https://doi.org/10.2196/12264)

KEYWORDS

autism; digital data; emotion; mobile phone; obsessive-compulsive disorder; social media; textual analysis; tweets; Twitter; infodemiology

Introduction

Autism spectrum disorder (ASD) is a group of developmental disorders that affect >3.5 million Americans [1]. The increasing prevalence of ASD requires greater investigation and methods to diagnose this disorder [2-5]. Diagnosing ASD is difficult, as there is no medical test to diagnose it [6]. Inadequate screening practices [7] and a general lack of awareness of symptoms can lead to delays or misdiagnoses [8]. Symptoms of ASD can sometimes remain unnoticed in childhood as social and language impairments can be subtle and difficult to detect [9].

The Diagnostic and Statistical Manual of Mental Disorders (DSM-5), a widely used classification system, describes ASD as a disorder characterized by marked impairments in social interaction and communication accompanied by a pattern of restricted, repetitive behaviors and activities [10]. However, many of these behaviors may be mistaken for other disorders. For instance, anxiety symptoms may be present in response to social impairments and, as a result, may be considered social anxiety without recognizing the underlying behaviors or the presence of ASD [9]. Reportedly, anxiety is a source and consequence of ASD behaviors and may be a signal to aid in diagnoses [9,11]. In addition, research shows that persons with ASD exhibit higher states of fear and phobia compared with matched patients without ASD [12]. The presence and identification of these symptoms may be markers for persons with ASD coping with judgment faced by social situations [9,13].

Furthermore, common repetitive behaviors observed among persons with ASD may be confused with similarities of having obsessive-compulsive disorder (OCD) [7]. Time-consuming obsessions and compulsions frequently characterize OCD-related behaviors; however, patients with ASD can display repetitive behavior, such as repetitive checking, washing, and counting, which may appear to be compulsive obsessions [14]. Research has found that patients with ASD have significantly fewer compulsions than persons with OCD [15]. Other studies have shown that behaviors like counting, checking, cleaning, and ordering were also displayed less frequently in patients with ASD compared with OCD individuals [16].

Appropriate and timely diagnosis of ASD is necessary to provide proper treatment and symptom management. Numerous diagnostic tools exist today, yet many of these tools face limitations and serve slightly different purposes [17,18]. For example, the Autism Diagnostic Observation Schedule-Generic (ADOS-G) is a widely used diagnostic tool that allows for the inclusion of direct observational data that, for example, the Autism Diagnostic Interview-Revised (ADI-R) lacks [19,20]. However, the ADOS-G often tends to result in the overdiagnosis of ASD [17]. Observational methods and parental interviews have been deemed as helpful to further identify behaviors of ASD, yet these symptoms may often go unnoticed [21,22]. Although multidisciplinary approaches have been considered

most effective at producing an accurate diagnosis, these require a significant amount of time, resources, and information, which is not always available [21,23].

Moreover, the use of multidisciplinary approaches extends to different domains, as well as novel sources of information [24]. Recently, exploring digital data has emerged as a promising method for examining human behavior and identifying symptoms of various mental and behavioral conditions [25-27]. Dubbed as the “digital phenotype,” the collection of data from social media, Web-based searches, and use of smartphones and other digital devices can serve as a source of observational information that can yield insights into the lived experiences and everyday realities of persons living with various health conditions [27]. These digital information sources can be further harnessed to potentially detect symptomatic signals to aid in identifying and diagnosing ASD.

Twitter has emerged as a potentially valuable platform for examining a wide variety of behavioral and mental health conditions [25-26,28]. Exploring textual differences in Web-based communication on Twitter is a commonly used method of analyzing affective states and behavioral patterns [29,30]. Previous studies have documented that Web-based data may capture real-time projections of emotional states and health behaviors [26,31]. For example, drawing from analyses of conversations on Twitter, it may be possible to identify the onset of depressive symptoms and detect mood and affective states [32,33]. Given this promising evidence base, there may be opportunities to leverage data from Twitter for identifying the characteristics of ASD.

To date, few studies have used social media for studying ASD, though it may be possible to leverage these popular platforms and novel sources of Web-based data to contribute to a more in-depth understanding of these disorders. A prior study extracted data from Twitter to examine conversations about ASD-related information, thereby demonstrating that this is a topic of frequent discussion on this platform [28]. It may be especially important to consider the use of social media for studying ASD given that a recent report highlighted that roughly 80% of adults living with ASD use popular social media platforms [34]. In this study, our overarching objective is to expand on this prior work and further our understanding about whether publicly available social media data captured from Twitter can yield insights into the presence of digital diagnostic signals for ASD [27,35,36]. In addition, we aim to determine the feasibility of informing a digital phenotype for ASD using social media.

Twitter is a social networking service through which users can post tweets that can contain up to 140 characters (increased to 280 characters in November 2017). Twitter is a prominent social medium, as evidenced by its >336 million monthly active users [37]. Twitter is a source of observational data that can yield rich data, including real-time thoughts, feelings, and attitudes, which traditional surveys are typically unable to capture. Therefore,

this study aims to compare the textual patterns of communication of Twitter users who self-identify as having ASD to a general population of control users to identify textual signals related to fear, anxiety, and paranoia, which have often been recognized as a response to symptoms of ASD among individuals living with the disorder [12]. Investigating discussion surrounding these emotions for Twitter users with ASD can potentially enable the detection of Web-based emotional signals that can provide more information to better inform diagnoses of ASD. In addition, this study aims to explore common OCD behaviors using the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) to understand whether these symptoms are present among Twitter users with ASD and whether the presence or absence of digital obsessive-compulsive symptoms could assist in reducing misdiagnoses of ASD. Finally, we investigated the timing of social media communications to explore differences in patterns of social communication between Twitter users with ASD and control users. Information collected from Twitter as a novel digital data source could potentially guide and assist in the clinical diagnoses and treatment of ASD.

Methods

Data Collection

In this study, we retrieved data from the Twitter application programming interface and collected meta-data such as Twitter users' total number of tweets, number of followers, number of friends, and favorites. We collected data from Twitter between March 22, 2012 and July 20, 2017. The trajectories of ASD symptoms are not stable and can fluctuate over the course of a few months or even years, especially during developmental periods. A 5-year period was selected to ensure that symptoms would be detected even with fluctuations to capture these detectable differences between the 2 groups. This study was considered exempt from ethical review because only publicly available Web-based data collected from the Twitter platform were analyzed. We acknowledge the ethics of drawing in data from users who self-identify as ASD. To ensure the privacy of data from this group, data collected were aggregated and analyzed at the population level and were anonymous in this format. Furthermore, this study only used data from users who consented on Twitter to disclose their data publicly (ie, no privacy settings were selected by users) and are completely public.

Twitter Users

We identified the cohort of Twitter users with ASD by self-identification and explicit Twitter discussion of having ASD. If users described or labeled themselves as autistic or having ASD, then they were determined to be Twitter users with ASD. We searched for terms relating to ASD in users' tweets, and users were considered Twitter users with ASD if they tweeted or wrote a caption about their condition or diagnosis of ASD. We used a previously validated approach to recruit participants with behavioral disorders by Hswen et al [25]. We retrieved tweets that contained one or more keywords associated with ASD; these keywords were searched for in the content of users' tweets and information in their Twitter profile. The keyword list is as follows: autism, autistic, Spectrum

Disorders, #ASD, #ActuallyAutistic, #AutismSpeaks, #AutismSpeaks10, #AutismParent, #Autchat, #AutismAwareness, #LIUB, aspergers, and aspies. To validate if users did have ASD, 2 independent researchers coded users if they had ASD or not and an agreement from 2 researchers was necessary for a user to be labeled as a person with ASD. If a disagreement occurred, a third researcher was used to tie break the labeling of the user. To the best of our knowledge, this is the first study to identify ASD users through Twitter, and it is an exploratory study for the identification and labeling of ASD users by machine learning classification and manual curation. A random sample was selected that did not have a self-identified user caption or tweet mentioning ASD and served as the control group. Originally, 220 users with ASD and 220 control users were selected. After filtering out bots and spam accounts and inaccessible accounts (ie, accounts that were private, deleted, deactivated, or banned), we enrolled 152 users with ASD and 182 control users for analysis.

Textual Analysis of Tweets Related to Fear, Anxiety, and Paranoia

We examined the tweets related to the emotions of fear, anxiety, and paranoia through textual analysis. These emotions were selected on the basis of previous research studies that have identified these emotions as key differentiating emotions of persons with ASD compared with matched populations without an ASD diagnosis [12]. We aimed to collect all discussion about these 3 emotional states. The root keyword of the emotion was expanded to all tenses and pluralities to capture the breath of tweets related to that emotion (Table 1). Manual selection of these terms was used to ensure that users were explicitly referring to the emotion in their conversations, a procedure we have applied in prior studies [25]. In addition, the terms "scare" and "scared" were used in the category of fear as a qualitative exploration of discussions around fear to express fearful states. Tweets from users with ASD and control users were queried for all tweets that contained keywords for each category. The number of tweets that contained each keyword was summed to create the total number of tweets for each emotion category. A tweet that contained instances of multiple words, such as "anxious" or "anxiety," was counted as 1 tweet. Overlapping keyword categories, such as a tweet that contained "anxious" and "fear," were counted as 1 tweet per category.

Tweets Containing Obsessive-Compulsive Disorder-Related Keywords

To examine the presence of OCD-related tweets, we compiled a list of key terms from symptom categories within the Y-BOCS. First developed by Goodman et al, this widely used checklist comprehensively covers significant OCD-related behaviors [38]. From the Y-BOCS, 13 categories were selected on the basis of previous research that identified these specific behaviors to be confused with symptoms of OCD and can be mislabeled as an OCD diagnosis [14-16]. These OCD symptom categories included the following: obsess; fixate; repeat; routine; freak; clean; check; count; hoard; wash; worry; excess; and concern. Stem words were expanded with the same methodology as applied for emotion keywords to conduct textual analysis (Table 2). For each category, the number of tweets containing these

terms was collected and summed to generate the total number of tweets per OCD category.

Table 1. All terms related to fear, anxiety, and paranoia.

Root keyword	All forms of keyword
fear	fear, feared, fearful, fearing, fears, scare, and scared
anxiety	anxieties, anxiety, anxious, and anxiousness
paranoia	paranoia, paranoiac, paranoiacs, and paranoid

Table 2. All terms related to obsessive-compulsive disorder-related keywords.

Root keyword	All forms of keyword
obsess	obsessed, obsessing, obsession, obsessions, obsessive, obsessively
fixate	fixated, fixation, fixations
repeat	repeat, repeated, repeatedly, repeating, repeats, repetition, repetitions
routine	routinely, routines
freak	freaked, freaking, freakout, freakouts
clean	cleaned, cleaning
check	checking, checked, checks, recheck, rechecks, rechecked, rechecking
count	counted, counting, counts, recount, recounted, recounting, recounts
hoard	hoarded, hoarder, hoarders, hoards
wash	washed, washing, washes
worry	worried, worrying
excess	excessive, excessively, excessiveness
concern	concerned, concerning, concern

Mean of Proportions for Each Category

As a method to standardize the variation of the frequency of tweets for each user, a “mean of proportions” was calculated for each category. This measure considers the mean number of tweets by category and the proportion of users that tweeted the keywords in each category for emotions and OCD terms. For each user, the total number of tweets for each category was divided by the users’ total number of tweets during the full study period between March 22, 2012 and July 20, 2017 (a total of 1946 days). This measure aimed to control the volume of tweets by users to ensure that the mean number of tweets by category was not skewed by individual users’ frequency of tweets. The mean of each proportion for users with ASD and the control user group was calculated by dividing the proportion of tweets by category over the number of users within each group. Of note, we used these means of proportions values solely for comparison between the ASD user group and the control user group; we do not know or study the clinical significance of the values.

Timing of Tweets

To examine differences in social communication patterns between users with ASD and control users, we analyzed the timing of tweets between the 2 groups to identify whether tweets occurred at different times of the day between the 2 groups. Meta-data were collected on the local time of users’ tweets using the universal time code (UTC) offset data. Twitter users can choose to select whether to include their local time zone in their

account settings, and tweets with UTC offset data were included in this analysis. These tweets were further classified into the following time intervals: 00:00-05:59; 06:00-11:59; 12:00-17:59; and 18:00-23:59.

Statistical Analyses

We conducted 2-tailed Welch’s *t* tests to compare the “mean of proportions” for each category of emotion and OCD keywords between users with ASD and control users. A chi-square test was used to compare the proportion of users in each category of the timing of tweet data between the 2 groups. For all tests, a $P < .05$ was considered statistically significant. All analyses were performed with the Python Programming Language Version 3.6 (Python Software Foundation).

Results

Characteristics of Twitter Users

In total, 4,093,559 tweets were collected between March 22, 2012 and July 20, 2017. Twitter users with ASD ($n=152$) tweeted a total of 1,700,841 tweets, and control users ($n=182$) tweeted a total of 2,392,718 tweets. We observed no statistically significant differences between the group of users with ASD and the group of control users in the mean number of tweets per user or for meta-data, including the number of friends, followers, or favorites (Table 3).

Mean Proportion of Tweets for Fear, Anxiety, and Paranoia

Users with ASD posted a greater number of tweets compared with control users for all 3 emotions—fear ($3.11\text{e-}03$ vs $1.98\text{e-}03$; $t_{225}=4.410$, $P<.001$), anxiety ($1.91\text{e-}03$ vs $6.49\text{e-}04$; $t_{172}=3.529$, $P=.001$), and paranoia ($1.90\text{e-}03$ vs $8.12\text{e-}05$; $t_{211}=3.021$, $P=.003$; Table 4; Figure 1). Aggregating all 3 emotions together, Twitter users with ASD posted a higher number of tweets compared with control users ($5.15\text{e-}03$ vs $2.68\text{e-}03$; $t_{190}=4.981$, $P<.001$; Table 4; Figure 1).

Mean Proportion of Tweets for Obsessive-Compulsive Disorder-Related Terms

Users with ASD posted a higher number of tweets compared with control users for 4 OCD-related keyword categories—fixate ($3.64\text{e-}05$ vs $1.23\text{e-}05$; $t_{230}=2.356$, $P=.02$), count ($1.43\text{e-}03$ vs $9.69\text{e-}04$; $t_{219}=3.457$, $P<.001$), excessive ($1.15\text{e-}04$ vs $4.82\text{e-}05$; $t_{201}=2.807$, $P=.005$), and concern ($9.26\text{e-}04$ vs $4.55\text{e-}04$; $t_{219}=3.959$, $P<.001$; Table 5; Figure 2). No statistically significant differences between users with ASD and control users were observed for the remaining 8 OCD-related categories

(Table 5; Figure 2). In addition, upon aggregating all OCD-related terms, no statistically significant difference between groups was observed ($1.20\text{e-}02$ vs $1.44\text{e-}02$; $t_{193}=1.292$, $P=.20$; Table 5; Figure 2).

Timing of Tweets

The number of Twitter users who had UTC offset data necessary to retrieve the timing of tweets was 69.2% (231/334). We observed no statistically significant difference in the availability of time zone information between users with ASD (111/152, 73.0%) compared with control users (120/182, 65.9%; $\chi^2_1=2.0$, $P=.16$). A lower proportion of tweets from users with ASD compared with control users was observed during the time intervals of 00:00-05:59 (222,810/1,700,841, 13.10% vs 330,915/2,392,718, 13.83%; $P<.001$), 06:00-11:59 (294,245/1,700,841, 17.30% vs 468,972/2,392,718, 19.60%; $P<.001$), and 18:00-23:59 (612,302/1,700,841, 36.00% vs 890,091/2,392,718, 37.20%; $P<.001$; Table 6). A higher proportion of tweets from users with ASD compared with control users was observed during the time interval 12:00-17:59 (571,483/1,700,841, 33.60% vs 703,459/2,392,718, 29.40%; $P<.001$; Table 6).

Table 3. The characteristics of Twitter users with autism spectrum disorder (ASD) and control users.

Twitter user characteristics	Twitter users with ASD (n=152), mean (SD)	Control twitter users (n=182), mean (SD)	<i>t</i> (df) ^a	<i>P</i> value
Mean overall tweets per user	11,189 (23,019)	13,146 (20,159)	-0.818 (303)	.41
Mean retweets per user	2237 (4886)	4942 (9680)	-3.299 (277)	.001
Mean original tweets per user	8952 (20,663)	8204 (12,369)	0.391 (237)	.70
Number of friends	1460 (5422)	1193 (5559)	0.442 (324)	.66
Number of followers	1778 (5753)	1891 (7752)	0.153 (327)	.88
Number of favorites	15,556 (31,966)	12,515 (19,978)	1.018 (243)	.31

^a2-tailed Welch's *t* tests were used to compare the "mean of proportions" for each category.

Table 4. Tweets containing emotion-related keywords among Twitter users with autism spectrum disorder (ASD) and control users.

Emotion terms category	Twitter users with ASD (n=152), mean (SD)	Control twitter users (n=182), mean (SD)	<i>t</i> (df) ^a	<i>P</i> value
fear	$3.105\text{e-}03$ ($2.822\text{e-}03$)	$1.975\text{e-}03$ ($1.553\text{e-}03$)	4.410 (225)	<.001
paranoid	$1.902\text{e-}04$ ($4.057\text{e-}04$)	$8.120\text{e-}05$ ($1.993\text{e-}04$)	3.021 (211)	.003
anxious	$1.914\text{e-}03$ ($4.275\text{e-}03$)	$6.486\text{e-}04$ ($1.230\text{e-}03$)	3.529 (172)	.001
Tweets with any of the 3 emotional categories' terms	$5.154\text{e-}03$ ($5.760\text{e-}03$)	$2.679\text{e-}03$ ($2.282\text{e-}03$)	4.981 (190)	<.001

^a2-tailed Welch's *t* tests were used to compare the "mean of proportions" for each category.

Figure 1. Differences in emotions between autism spectrum disorder (ASD) and control users.

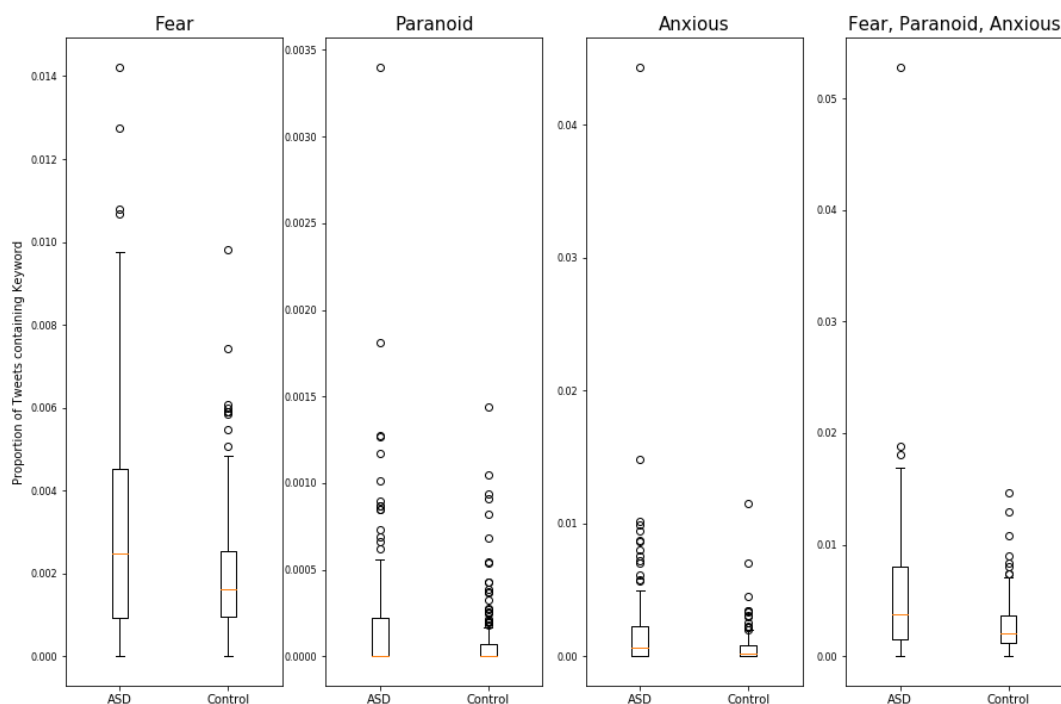
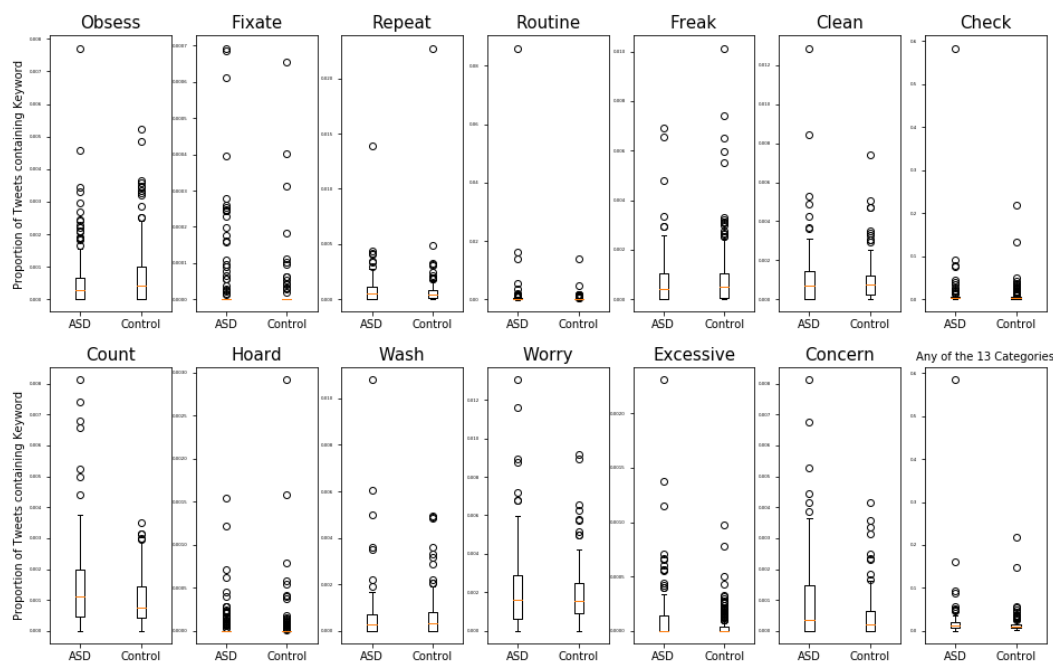


Table 5. Tweets containing obsessive-compulsive disorder (OCD)–related keywords among Twitter users with autism spectrum disorder (ASD) and control users.

OCD keyword category	Twitter users with ASD (n=152), mean (SD)	Control twitter users (n=182), mean (SD)	<i>t</i> (<i>df</i>) ^a	<i>P</i> value
obsess	5.826e-04 (9.807e-04)	7.086e-04 (9.377e-04)	-1.193 (316)	.23
fixate	3.640e-05 (1.120e-04)	1.226e-05 (6.404e-05)	2.356 (230)	.02
repeat	8.711e-04 (1.381e-03)	7.239e-04 (1.775e-03)	0.852 (330)	.40
routine	1.027e-03 (7.171e-03)	2.866e-04 (1.114e-03)	1.260 (157)	.21
freak	7.584e-04 (1.066e-03)	8.827e-04 (1.342e-03)	-0.943 (331)	.35
clean	1.068e-03 (1.534e-03)	9.481e-04 (1.026e-03)	0.821 (255)	.41
check	1.061e-02 (4.840e-02)	6.944e-03 (1.982e-02)	0.874 (193)	.38
count	1.427e-03 (1.471e-03)	9.693e-04 (7.744e-04)	3.457 (219)	.001
hoard	5.608e-05 (1.876e-04)	4.938e-05 (2.616e-04)	0.272 (325)	.79
wash	5.766e-04 (1.176e-03)	5.986e-04 (8.416e-04)	-0.193 (267)	.85
worry	2.071e-03 (2.167e-03)	1.878e-03 (1.454e-03)	0.934 (256)	.35
excessive	1.152e-04 (2.728e-04)	4.816e-05 (1.215e-04)	2.807 (201)	.005
concern	9.260e-04 (1.322e-03)	4.548e-04 (6.964e-04)	3.959 (219)	<.001
Tweets with any of the 13 categories' terms	1.997e-02 (4.940e-02)	1.445e-02 (2.023e-02)	1.292 (193)	.20

^a2-tailed Welch's *t* tests were used to compare the “mean of proportions” for each category.

Figure 2. Differences in obsessive-compulsive disorder-related discussion between autism spectrum disorder (ASD) and control users.**Table 6.** Timing of users' tweets between Twitter users with autism spectrum disorder (ASD) and control users.

Time interval	Proportion of tweets (%)		χ^2 (df)	P value
	Among twitter users with ASD	Among control twitter users		
00:00-05:59	13.1	13.8	351.9 (1)	<.001
06:00-11:59	17.3	19.6	2952.9 (1)	<.001
12:00-17:59	33.6	29.4	6713.7 (1)	<.001
18:00-23:59	36.0	37.2	474.3 (1)	<.001

Discussion

Principal Findings

The diagnosis of ASD is complex. Often a multidisciplinary approach is required to produce the most accurate diagnosis. This study investigated the feasibility of using Twitter as a novel digital platform for generating additional insights to assist in the diagnosis of ASD. The findings revealed that through the textual analysis of tweets, discussions about commonly heightened emotional states associated with ASD are expressed among Twitter users who self-identify as having ASD. Specifically, Twitter users with ASD tweeted more frequently about fear, anxiety, and paranoia compared with a randomly generated control group of Twitter users. In addition, these results suggest that Web-based emotional patterns expressed among Twitter users with ASD may parallel offline emotional symptoms known to be associated with the condition and, therefore, may be used as a signal for detecting or confirming a diagnosis of ASD.

As misdiagnoses of ASD for OCD are common because of many overlapping and similar symptoms [39,40], in this study, we also explored whether there may be differences in the expression of OCD-related behaviors among Twitter users with ASD compared with control users. Of 13 selected OCD-related categories, users with ASD discussed only 4 categories—fixate, count, excessive, and concern—significantly more than the control user group. Furthermore, when aggregating all 13 categories, we observed no statistically significant difference. These findings indicate that Twitter users with ASD may not display a wide number of symptoms of OCD and may highlight the most salient symptoms that could be explored further as a method to distinguish symptoms of ASD from individuals with OCD.

Although research on the presence of some of these behaviors may steer clinicians toward a diagnosis of OCD, misattribution to a behavior or misdiagnosis can lead to the inappropriate treatment for ASD [7,41]. Therefore, it is important to consider the spectrum of symptoms of OCD and determine whether Web-based discussions of these behaviors are consistent with offline representations of OCD. However, this highlights an

important area for future investigation because offline observational information of this variety of symptoms can be difficult to obtain. Ideally, there may be opportunities to leverage Web-based social media data that allow us to capture minute-by-minute thoughts and behaviors to provide clinicians a more objective illustration of patients' behaviors toward better understanding the daily occurrences of these symptoms and inform diagnosis and treatment recommendations.

Finally, we sought to determine differences in patterns of the timing of Twitter communication between users with ASD and control users. As ASD is characterized as a social disorder [41], there could be differences in patterns of the Web-based communication timing among individuals with this condition compared with other mental disorders. We found some differences in the proportion of tweets between groups at each time-point throughout the day; however, these differences were very small and are likely not indicative of any diagnostically relevant differences between groups. This is consistent with previous studies that have found no differences in the timing of tweets between a general population of Twitter users and users with mental disorders such as schizophrenia [25]. Additional research is needed to explore whether potential differences in the timing of social media use and posts between users with ASD and control users could inform the development of targeted clinical and public health interventions to reach this patient group online.

Limitations

Although Web-based social media offers an opportunity to observe a wide range of thoughts and behaviors, our data have some limitations. First, Twitter users who self-identify as having ASD may not be representative of the general population of individuals with ASD. The Twitter population, on average, is composed of a larger proportion of younger adults aged 18-29 years with higher levels of education compared with those who do not use Twitter [42]. Therefore, it is more likely that users with ASD in our study are younger adults and with higher education levels than persons with ASD who do not use social media. Second, identification of users with ASD was through self-identification on Twitter. Many individuals with ASD who are on Twitter may not openly self-identify as having ASD, thereby making the sample included in this study potentially more comfortable discussing their disorder over the Web with others. This might mean that our study population is different compared with Twitter users with ASD who choose not to self-identify as having ASD online. Although no formal clinical diagnosis was confirmed for Twitter users included in the analyses conducted in this study, the nature of the stigmatizing diagnosis of ASD makes it highly unlikely that individuals would be dishonest about publicly self-identifying as having ASD on Twitter. There is a possibility, however, that there are some who might have self-diagnosed themselves as having ASD without a clinical consultation, and we, therefore, cannot verify the clinical diagnosis of ASD in our study population. Of note, this was an exploratory study that used textual analyses to explore whether common emotions and OCD-related behaviors could be detected among Twitter users with ASD and whether these discussions may differ compared with control users. As a result, we can only report on publicly available discussions

captured from Twitter and, therefore, we are unable to confirm whether these are also occurring offline. In addition, we did not identify a cohort of OCD users and did not conduct a comparison between ASD and OCD users. Therefore, we are not able to confirm the differences in Web-based patterns between the 2 groups and further examination is necessary. Furthermore, we note that we developed an adaptive list for the emotion categories related to autism but did not expand this list for the OCD-related categories. We specifically used the terminology used in the Y-BOCS to mimic the survey and the terms used to diagnose in the clinical setting. However, we acknowledge that these OCD terms could have been adapted; therefore, we potentially missed OCD-related behaviors in our analysis of tweets. Finally, we acknowledge that small yet detectable differences were yielded from our analyses. Owing to the size and breadth of our dataset, although small, these differences can detect important differences in the characteristics between ASD users and the general population and are worth exploring further as to what is a meaningful change to influence a diagnosis.

The strength of this study is that this is the first known study to leverage Web-based social media data from Twitter to compare emotional and OCD-related discussions between users with ASD and a randomly select group of control users. A comprehensive picture of a digital footprint of emotions and behaviors can be gleaned from Twitter through the examination of Web-based social media data. This study highlights the feasibility of conducting this type of analysis through these emerging digital streams to further contribute to our understanding of a "digital phenotype" for ASD. This study yields additional insights that contribute to a growing body of evidence depicting digital phenotyping and the potential to draw from Web-based data sources to detect signals and symptoms for a range of disorders such as ASD [27,35,36]. The data captured in this way were in a naturalistic and unsolicited format, thereby avoiding the concerns of social desirability bias or recall bias that are present with traditional survey data collection methods. Therefore, the methods applied in this study highlight the potential of collecting novel data streams that could be used to supplement the existing data collection methods. Furthermore, it is important to note that we are dealing with individuals from vulnerable population groups, and they are contributing their conversations and thoughts for observational research studies such as ours; however, all of these publicly available data are always aggregated, and we purposefully did not identify the individuals.

Conclusions

Digital signals from Web-based social media may offer an additional resource to capture information not always available for direct observation. This Web-based data source is not meant to replace traditional forms of diagnosis but to supplement these existing approaches by helping to identify symptoms or confirm diagnoses of ASD. Multidisciplinary approaches are often required to gain a full breadth of information and appear to be the gold standard of diagnostic methods. The accuracy of diagnoses is necessary to ensure that persons with ASD receive proper treatment and care in a timely manner so that they can live a fulfilling and healthy life [21,43-45]. Findings from this

study can contribute to the development of multidisciplinary clinical assessments that draw from novel digital sources, as well as traditional surveys and clinical interviews, and takes steps toward illustrating the potential feasibility of digital detection methods for ASD.

Acknowledgments

This study was funded by the Robert Wood Johnson Foundation Grant 73495 (to YH and JBH), National Institutes of Health/National Human Genome Research Institute Grant 5U54HG007963-04 (to JSB and JBH), the Weatherhead Center for International Affairs (to YH).

Conflicts of Interest

None declared.

References

1. Buescher AVS, Cidav Z, Knapp M, Mandell DS. Costs of autism spectrum disorders in the United Kingdom and the United States. *JAMA Pediatr* 2014 Aug;168(8):721-728. [doi: [10.1001/jamapediatrics.2014.210](https://doi.org/10.1001/jamapediatrics.2014.210)] [Medline: [24911948](https://pubmed.ncbi.nlm.nih.gov/24911948/)]
2. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators, Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *MMWR Surveill Summ* 2012 Mar 30;61(3):1-19 [[FREE Full text](#)] [Medline: [22456193](https://pubmed.ncbi.nlm.nih.gov/22456193/)]
3. Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, et al. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *Lancet* 2006 Jul 15;368(9531):210-215. [doi: [10.1016/S0140-6736\(06\)69041-7](https://doi.org/10.1016/S0140-6736(06)69041-7)] [Medline: [16844490](https://pubmed.ncbi.nlm.nih.gov/16844490/)]
4. Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, Matthews FE, et al. Prevalence of autism-spectrum conditions: UK school-based population study. *Br J Psychiatry* 2009 Jun;194(6):500-509 [[FREE Full text](#)] [doi: [10.1192/bjp.bp.108.059345](https://doi.org/10.1192/bjp.bp.108.059345)] [Medline: [19478287](https://pubmed.ncbi.nlm.nih.gov/19478287/)]
5. Brugha TS, McManus S, Bankart J, Scott F, Purdon S, Smith J, et al. Epidemiology of autism spectrum disorders in adults in the community in England. *Arch Gen Psychiatry* 2011 May;68(5):459-465. [doi: [10.1001/archgenpsychiatry.2011.38](https://doi.org/10.1001/archgenpsychiatry.2011.38)] [Medline: [21536975](https://pubmed.ncbi.nlm.nih.gov/21536975/)]
6. Centers for Disease Control and Prevention. 2012. Screening and diagnosis for healthcare providers URL: <https://www.cdc.gov/ncbddd/autism/hcp-screening.html> [WebCite Cache ID 72Evv3U4G]
7. Mandell DS, Ittenbach RF, Levy SE, Pinto-Martin JA. Disparities in diagnoses received prior to a diagnosis of autism spectrum disorder. *J Autism Dev Disord* 2007 Oct;37(9):1795-1802 [[FREE Full text](#)] [doi: [10.1007/s10803-006-0314-8](https://doi.org/10.1007/s10803-006-0314-8)] [Medline: [17160456](https://pubmed.ncbi.nlm.nih.gov/17160456/)]
8. Volkmar FR, Lord C, Bailey A, Schultz RT, Klin A. Autism and pervasive developmental disorders. *J Child Psychol Psychiatry* 2004 Jan;45(1):135-170. [Medline: [14959806](https://pubmed.ncbi.nlm.nih.gov/14959806/)]
9. Cath DC, Ran N, Smit JH, van Balkom AJLM, Comijs HC. Symptom overlap between autism spectrum disorder, generalized social anxiety disorder and obsessive-compulsive disorder in adults: a preliminary case-controlled study. *Psychopathology* 2008;41(2):101-110. [doi: [10.1159/000111555](https://doi.org/10.1159/000111555)] [Medline: [18033980](https://pubmed.ncbi.nlm.nih.gov/18033980/)]
10. Diagnostic and statistical manual of mental disorders. 5th edition. Arlington, VA: American Psychiatric Association; 2013.
11. Bryson SE, Rogers SJ, Fombonne E. Autism spectrum disorders: early detection, intervention, education, and psychopharmacological management. *Can J Psychiatry* 2003 Sep;48(8):506-516. [doi: [10.1177/070674370304800802](https://doi.org/10.1177/070674370304800802)] [Medline: [14574826](https://pubmed.ncbi.nlm.nih.gov/14574826/)]
12. Evans DW, Canavera K, Kleinpeter FL, Maccubbin E, Taga K. The fears, phobias and anxieties of children with autism spectrum disorders and Down syndrome: comparisons with developmentally and chronologically age matched children. *Child Psychiatry Hum Dev* 2005;36(1):3-26. [doi: [10.1007/s10578-004-3619-x](https://doi.org/10.1007/s10578-004-3619-x)] [Medline: [16049642](https://pubmed.ncbi.nlm.nih.gov/16049642/)]
13. Grant BF, Hasin DS, Blanco C, Stinson FS, Chou SP, Goldstein RB, et al. The epidemiology of social anxiety disorder in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *J Clin Psychiatry* 2005 Nov;66(11):1351-1361. [Medline: [16420070](https://pubmed.ncbi.nlm.nih.gov/16420070/)]
14. Rumsey JM, Rapoport JL, Sceery WR. Autistic children as adults: psychiatric, social, and behavioral outcomes. *J Am Acad Child Psychiatry* 1985 Jul;24(4):465-473. [Medline: [4019976](https://pubmed.ncbi.nlm.nih.gov/4019976/)]
15. Zandt F, Prior M, Kyrios M. Repetitive behaviour in children with high functioning autism and obsessive compulsive disorder. *J Autism Dev Disord* 2007 Feb;37(2):251-259. [doi: [10.1007/s10803-006-0158-2](https://doi.org/10.1007/s10803-006-0158-2)] [Medline: [16865546](https://pubmed.ncbi.nlm.nih.gov/16865546/)]
16. Colasanto M. A Comparative Study of Repetitive Behaviours in Pediatric Obsessive-compulsive Disorder and Autism Spectrum Disorder. University of Toronto 2015 Nov [[FREE Full text](#)]
17. Bishop DVM, Norbury CF. Exploring the borderlands of autistic disorder and specific language impairment: a study using standardised diagnostic instruments. *J Child Psychol Psychiatry* 2002 Oct;43(7):917-929. [Medline: [12405479](https://pubmed.ncbi.nlm.nih.gov/12405479/)]

18. Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry* 2006 Sep;45(9):1094-1103. [doi: [10.1097/01.chi.0000227880.42780.0e](https://doi.org/10.1097/01.chi.0000227880.42780.0e)] [Medline: [16926617](https://pubmed.ncbi.nlm.nih.gov/16926617/)]
19. Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 2000 Jun;30(3):205-223. [Medline: [11055457](https://pubmed.ncbi.nlm.nih.gov/11055457/)]
20. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994 Oct;24(5):659-685. [Medline: [7814313](https://pubmed.ncbi.nlm.nih.gov/7814313/)]
21. Filipek PA, Accardo PJ, Ashwal S, Baranek GT, Cook EH, Dawson G, et al. Practice parameter: screening and diagnosis of autism: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology* 2000 Aug 22;55(4):468-479. [Medline: [10953176](https://pubmed.ncbi.nlm.nih.gov/10953176/)]
22. Ozonoff S, Heung K, Byrd R, Hansen R, Hertz-Picciotto I. The onset of autism: patterns of symptom emergence in the first years of life. *Autism Res* 2008 Dec;1(6):320-328 [FREE Full text] [doi: [10.1002/aur.53](https://doi.org/10.1002/aur.53)] [Medline: [19360687](https://pubmed.ncbi.nlm.nih.gov/19360687/)]
23. Westman AG, Miniscalco C, Gillberg C. Autism in preschoolers: does individual clinician's first visit diagnosis agree with final comprehensive diagnosis? *Scientific World Journal* 2013;2013:716267 [FREE Full text] [doi: [10.1155/2013/716267](https://doi.org/10.1155/2013/716267)] [Medline: [24082856](https://pubmed.ncbi.nlm.nih.gov/24082856/)]
24. Grzadzinski R, Huerta M, Lord C. DSM-5 and autism spectrum disorders (ASDs): an opportunity for identifying ASD subtypes. *Mol Autism* 2013 May 15;4(1):12 [FREE Full text] [doi: [10.1186/2040-2392-4-12](https://doi.org/10.1186/2040-2392-4-12)] [Medline: [23675638](https://pubmed.ncbi.nlm.nih.gov/23675638/)]
25. Hswen Y, Naslund JA, Brownstein JS, Hawkins JB. Online Communication about Depression and Anxiety among Twitter Users with Schizophrenia: Preliminary Findings to Inform a Digital Phenotype Using Social Media. *Psychiatr Q* 2018 Dec;89(3):569-580. [doi: [10.1007/s11126-017-9559-y](https://doi.org/10.1007/s11126-017-9559-y)] [Medline: [29327218](https://pubmed.ncbi.nlm.nih.gov/29327218/)]
26. Hswen Y, Naslund JA, Chandrashekar P, Siegel R, Brownstein JS, Hawkins JB. Exploring online communication about cigarette smoking among Twitter users who self-identify as having schizophrenia. *Psychiatry Res* 2017 Dec;257:479-484 [FREE Full text] [doi: [10.1016/j.psychres.2017.08.002](https://doi.org/10.1016/j.psychres.2017.08.002)] [Medline: [28841509](https://pubmed.ncbi.nlm.nih.gov/28841509/)]
27. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol* 2015 May;33(5):462-463. [doi: [10.1038/nbt.3223](https://doi.org/10.1038/nbt.3223)] [Medline: [25965751](https://pubmed.ncbi.nlm.nih.gov/25965751/)]
28. Beykikhoshk A, Arandjelović O, Phung D, Venkatesh S, Caelli T. Using Twitter to learn about the autism community. *Soc Netw Anal Min* 2015 Jun 2;5(1). [doi: [10.1007/s13278-015-0261-5](https://doi.org/10.1007/s13278-015-0261-5)]
29. Chen X, Sykora M, Jackson T, Elayan S. What about Mood Swings? Identifying Depression on Twitter with Temporal Measures of Emotions. 2018 Presented at: Proceedings of The Web Conference 2018; April 2018; Lyon, France p. 23-27. [doi: [10.1145/3184558.3191624](https://doi.org/10.1145/3184558.3191624)]
30. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in twitter. : Association for Computational Linguistics; 2014 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June 2014; Baltimore, Maryland p. 51-60 URL: <http://aclweb.org/anthology/W14-3207> [doi: [10.3115/v1/W14-3207](https://doi.org/10.3115/v1/W14-3207)]
31. McIver DJ, Hawkins JB, Chunara R, Chatterjee AK, Bhandari A, Fitzgerald TP, et al. Characterizing Sleep Issues Using Twitter. *J Med Internet Res* 2015 Jun 08;17(6):e140 [FREE Full text] [doi: [10.2196/jmir.4476](https://doi.org/10.2196/jmir.4476)] [Medline: [26054530](https://pubmed.ncbi.nlm.nih.gov/26054530/)]
32. De Choudhury M, Gamon M, Counts S. Happy, nervous or surprised? Classification of human affective states in social media. 2012 Presented at: Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media; June 2012; Dublin, Ireland p. 4-7.
33. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Presented at: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media; 2013; Cambridge, Massachusetts p. 8-11.
34. Mazurek MO. Social media use among adults with autism spectrum disorders. *Computers in Human Behavior* 2013 Jul;29(4):1709-1714. [doi: [10.1016/j.chb.2013.02.004](https://doi.org/10.1016/j.chb.2013.02.004)]
35. Insel TR. Digital Phenotyping: Technology for a New Science of Behavior. *JAMA* 2017 Oct 03;318(13):1215-1216. [doi: [10.1001/jama.2017.11295](https://doi.org/10.1001/jama.2017.11295)] [Medline: [28973224](https://pubmed.ncbi.nlm.nih.gov/28973224/)]
36. Onnela J, Rauch SL. Harnessing Smartphone-Based Digital Phenotyping to Enhance Behavioral and Mental Health. *Neuropsychopharmacology* 2016 Jun;41(7):1691-1696. [doi: [10.1038/npp.2016.7](https://doi.org/10.1038/npp.2016.7)] [Medline: [26818126](https://pubmed.ncbi.nlm.nih.gov/26818126/)]
37. Statista. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 3rd quarter 2018 (in millions) URL: <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/> [accessed 2018-09-09] [WebCite Cache ID 72IBsyybY]
38. Goodman WK, Price LH, Rasmussen SA, Mazurek C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry* 1989 Nov;46(11):1006-1011. [Medline: [2684084](https://pubmed.ncbi.nlm.nih.gov/2684084/)]
39. McDougle CJ, Kresch LE, Goodman WK, Naylor ST, Volkmar FR, Cohen DJ, et al. A case-controlled study of repetitive thoughts and behavior in adults with autistic disorder and obsessive-compulsive disorder. *Am J Psychiatry* 1995 May;152(5):772-777. [doi: [10.1176/ajp.152.5.772](https://doi.org/10.1176/ajp.152.5.772)] [Medline: [7726318](https://pubmed.ncbi.nlm.nih.gov/7726318/)]
40. Russell AJ, Mataix-Cols D, Anson M, Murphy DGM. Obsessions and compulsions in Asperger syndrome and high-functioning autism. *Br J Psychiatry* 2005 Jun;186:525-528. [doi: [10.1192/bjp.186.6.525](https://doi.org/10.1192/bjp.186.6.525)] [Medline: [15928364](https://pubmed.ncbi.nlm.nih.gov/15928364/)]

41. American Psychiatric Association. In: Diagnostic and statistical manual of mental disorders, Fifth Edition. Arlington, VA: American Psychiatric Publishing; 2013.
42. Pew Research Center. 2018 Feb 5. Social Media Fact Sheet URL: <http://www.pewinternet.org/fact-sheet/social-media/> [accessed 2018-09-10] [WebCite Cache ID 72Jv3sDCB]
43. Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, et al. Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics* 2010 Jan;125(1):e17-e23 [FREE Full text] [doi: [10.1542/peds.2009-0958](https://doi.org/10.1542/peds.2009-0958)] [Medline: [19948568](https://pubmed.ncbi.nlm.nih.gov/19948568/)]
44. Kasari C, Gulsrud AC, Wong C, Kwon S, Locke J. Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *J Autism Dev Disord* 2010 Sep;40(9):1045-1056 [FREE Full text] [doi: [10.1007/s10803-010-0955-5](https://doi.org/10.1007/s10803-010-0955-5)] [Medline: [20145986](https://pubmed.ncbi.nlm.nih.gov/20145986/)]
45. Rogers SJ, Vismara L, Wagner AL, McCormick C, Young G, Ozonoff S. Autism treatment in the first year of life: a pilot study of infant start, a parent-implemented intervention for symptomatic infants. *J Autism Dev Disord* 2014 Dec;44(12):2981-2995 [FREE Full text] [doi: [10.1007/s10803-014-2202-y](https://doi.org/10.1007/s10803-014-2202-y)] [Medline: [25212413](https://pubmed.ncbi.nlm.nih.gov/25212413/)]

Abbreviations

ASD: autism spectrum disorder

OCD: obsessive-compulsive disorder

UTC: universal time code

Y-BOCS: Yale-Brown Obsessive-Compulsive Scale

Edited by C Dias; submitted 24.09.18; peer-reviewed by X Chen, S Henderson, L Geneviève, D Mciver; comments to author 02.11.18; revised version received 16.11.18; accepted 18.11.18; published 08.02.19

Please cite as:

Hswen Y, Gopaluni A, Brownstein JS, Hawkins JB

Using Twitter to Detect Psychological Characteristics of Self-Identified Persons With Autism Spectrum Disorder: A Feasibility Study

JMIR Mhealth Uhealth 2019;7(2):e12264

URL: <http://mhealth.jmir.org/2019/2/e12264/>

doi: [10.2196/12264](https://doi.org/10.2196/12264)

PMID:

©Yulin Hswen, Anuraag Gopaluni, John S Brownstein, Jared B Hawkins. Originally published in JMIR Mhealth and Uhealth (<http://mhealth.jmir.org>), 08.02.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mhealth and uhealth, is properly cited. The complete bibliographic information, a link to the original publication on <http://mhealth.jmir.org/>, as well as this copyright and license information must be included.