Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis

5 authors, including:

Angela Leis
University Pompeu Fabra
38 PUBLICATIONS   276 CITATIONS

SEE PROFILE

Francesco Ronzano
University Pompeu Fabra
68 PUBLICATIONS   1,893 CITATIONS

SEE PROFILE

Miguel Angel Mayer
Hospital del Mar Medical Research Institute & Universitat Pompeu Fabra
178 PUBLICATIONS   1,515 CITATIONS

SEE PROFILE

Laura I Furlong
Consorci MAR Parc de Salut de Barcelona
181 PUBLICATIONS   5,712 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   eTOX-eTRANSAFE View project

Project   European Medical Information Framework View project

Original Paper

Detecting Signs of Depression in Tweets in Spanish: Behavioral and Linguistic Analysis

Angela Leis, PsyM; Francesco Ronzano, PhD; Miguel A Mayer, MD, PhD; Laura I Furlong, PhD; Ferran Sanz, Prof Dr

Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain

Corresponding Author:
Ferran Sanz, Prof Dr
Research Programme on Biomedical Informatics
Hospital del Mar Medical Research Institute
Department of Experimental and Health Sciences, Universitat Pompeu Fabra
Carrer Dr Aiguader 88
Barcelona, 08003
Spain
Phone: 34 933 160 540
Fax: 34 933 160 550
Email: ferran.sanz@upf.edu

Abstract

Background: Mental disorders have become a major concern in public health, and they are one of the main causes of the overall disease burden worldwide. Social media platforms allow us to observe the activities, thoughts, and feelings of people's daily lives, including those of patients suffering from mental disorders. There are studies that have analyzed the influence of mental disorders, including depression, in the behavior of social media users, but they have been usually focused on messages written in English.

Objective: The study aimed to identify the linguistic features of tweets in Spanish and the behavioral patterns of Twitter users who generate them, which could suggest signs of depression.

Methods: This study was developed in 2 steps. In the first step, the selection of users and the compilation of tweets were performed. A total of 3 datasets of tweets were created, a depressive users dataset (made up of the timeline of 90 users who explicitly mentioned that they suffer from depression), a depressive tweets dataset (a manual selection of tweets from the previous users, which included expressions indicative of depression), and a control dataset (made up of the timeline of 450 randomly selected users). In the second step, the comparison and analysis of the 3 datasets of tweets were carried out.

Results: In comparison with the control dataset, the depressive users are less active in posting tweets, doing it more frequently between 23:00 and 6:00 ($P<.001$). The percentage of nouns used by the control dataset almost doubles that of the depressive users ($P<.001$). By contrast, the use of verbs is more common in the depressive users dataset ($P<.001$). The first-person singular pronoun was by far the most used in the depressive users dataset (80%), and the first- and the second-person plural pronouns were the least frequent (0.4% in both cases), this distribution being different from that of the control dataset ($P<.001$). Emotions related to sadness, anger, and disgust were more common in the depressive users and depressive tweets datasets, with significant differences when comparing these datasets with the control dataset ($P<.001$). As for negation words, they were detected in 34% and 46% of tweets in among depressive users and in depressive tweets, respectively, which are significantly different from the control dataset ($P<.001$). Negative polarity was more frequent in the depressive users (54%) and depressive tweets (65%) datasets than in the control dataset (43.5%; $P<.001$).

Conclusions: Twitter users who are potentially suffering from depression modify the general characteristics of their language and the way they interact on social media. On the basis of these changes, these users can be monitored and supported, thus introducing new opportunities for studying depression and providing additional health care services to people with this disorder.

XSL•FO

RenderX

Introduction

Background

Mental health is an essential component of our health. The World Health Organization (WHO) defines mental health as a "state of well-being in which people realize their potential, cope with the normal stresses of life, work productively, and contribute to their communities" [1]. Good mental health is about being cognitive, emotionally and socially healthy and it helps to determine the way we think and feel, in relation with others and how we make choices. Several factors, such as genetic, sociocultural, economic, political and environmental aspects, shape and influence our mental health. In the last few years, mental disorders have become a major concern in public health, and they are one of the main causes of the overall disease burden worldwide. They have devastating consequences for both patients and their families [2-7]. According to the WHO, depressive disorders are the most common among the mental illnesses [8]. Such disorders conditions are characterized by sadness, loss of interest and pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, feelings of tiredness, and poor concentration [8]. In 2018, at the global level, more than 300 million people were suffering from depression, and it is the main contributor to global disability. Depression has several consequences, both personal and social costs [9,10]. In some cases, depression can lead to suicide ideation and attempts [2,11]. The prevalence of this disorder changes depending on age, but it affects the whole population, from children and adolescents to elderly people. From 2005 to 2015, the number of people with depression increased by around 18% [12]. In this context, social media platforms allow to observe the activities, thoughts, and feelings of people's daily lives and thereby investigate their emotional well-being. This domain has become a new growing area of interest in public health and health care research [13-16]. People with depression often use social media to talk about their illness and treatment, share information and experiences, seek social support and advice, reduce social isolation, and manage their mental illness [15-21]. In addition, access to mobile devices facilitates the use of social media platforms, such as Twitter and Facebook, at any time and at any place. Social media, such as Twitter, is by nature social, and we can consequently find social patterns in Twitter feeds, thereby revealing key aspects of mental and affective disorders [22]. Social media has become an important source of health-related information, which allows us to detect and predict affective disorders and which can be used as an additional tool for mental health monitoring and infoveillance [23-26]. Furthermore, the application of different methodologies based on natural language processing and machine learning technologies has proved to be effective in supporting and automating the identification of early signs of mental illness by analyzing the content shared on the Web by individuals [13-15,27]. This human interaction with social media contributes to build the so-called digital phenotype, reshaping disease expression in terms of the lived experience of individuals and detecting early manifestations of several conditions [28]. Twitter is an internet microblogging social media service that allows users to post short messages about facts, feelings and opinions,

and, as shown in previous studies, users' health conditions [15]. Twitter is one of the most important social media platforms in terms of number of users, with more than 330 million active users worldwide [29]. Since November 2017, the maximum number of characters of a tweet has been increased from 140 to 280. By analyzing huge amounts of text, researchers can link everyday language use with social behavior and personality [30,31]. Language, as a means of communication, constitutes an essential element for providing valuable insights about people's interests, feelings and concerns [32]. For this reason, the analysis of the messages posted on social media platforms may provide information about many personality traits, lifestyles, and psychological disorders [13,33,34]. The potential anonymity of social media encourages its users to be more willing to report health information, such as details of their mental disorders and treatments received. In addition, it is seen as a way to communicate and receive support from others with similar experiences, avoiding the isolation and fighting the social stigma of these conditions [12,15,17,19,32,35]. Nevertheless, users suffering from depression may also feel uncomfortable socializing and consuming information on social media platforms [36]. Several features of the messages, such as number and frequency of tweets, distribution throughout the day or during the night hours, and their seasonal character, can be used for the detection and monitoring of mental disorders, such as depression [20]. This knowledge can help health care professionals and health institutions and services in the decision-making processes to ensure better management of patients suffering from depression.

Objectives

There are many studies that have used data mining and machine learning techniques on social media platforms to automatically identify people with mental health problems, such as depression, posttraumatic stress disorder, schizophrenia, or eating disorders, usually focusing the studies on messages written in English [20,37-39]. As far as we know, on social media, there are no studies about mental disorders that analyze messages written in Spanish. Taking into account that Spanish speaking countries, such as Spain and Mexico, are among the 10 most active Twitter users in the world, with more than 6 million and 7 million users, respectively [40], we focused our research on the expression of depression in Spanish language tweets. The aim of this study was to identify the linguistic features of tweets written in Spanish and the behavioral patterns of the corresponding Twitter users that could suggest signs of depression.
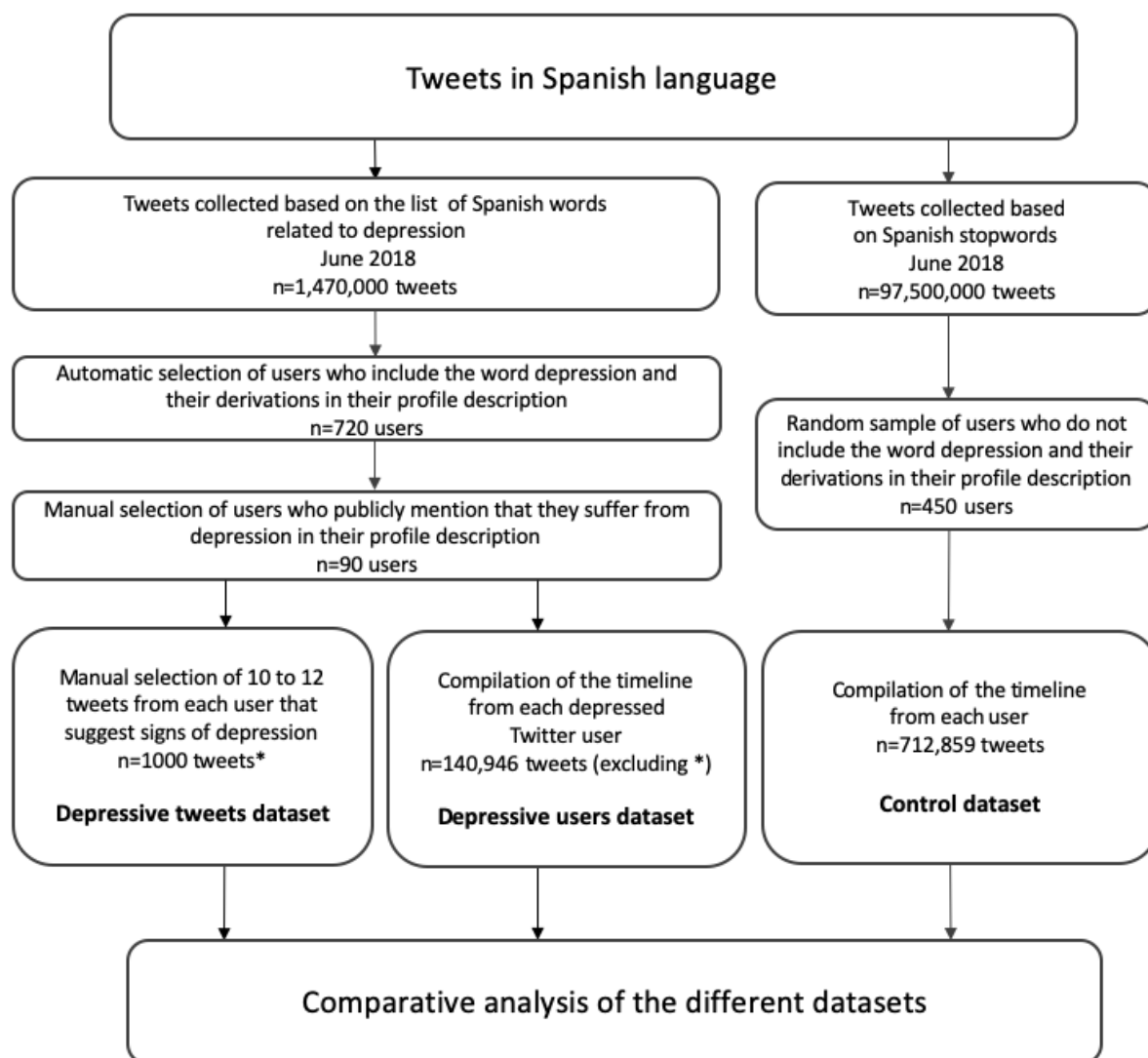
Methods

Study Steps

This study was designed and developed in 2 steps, with the aim of analyzing the linguistic patterns and behavioral features of Twitter users suffering from depression in comparison with the general population of Twitter users. The study was focused on tweets written in Spanish. In the first step, the selection of users and the compilation of tweets were performed. Given the design and purpose of the study, we decided to use the Twitter Application Programming Interface (API) [41]. Using this API, 3 datasets of tweets were created:

1. The *depressive users* dataset was made up of the timeline of 90 users who publicly mentioned on their Twitter profile that they suffer from depression.
2. The *control* dataset was made up of the timeline of 450 randomly selected Twitter users.
3. The *depressive tweets* dataset was constituted by a manual selection of tweets from the depressive users dataset, which specifically included expressions indicative of depression.

In the second step, comparison and analysis of the 3 datasets of tweets (control, depressive users, and depressive tweets datasets) were carried out to spot their distinguishing features. In the rest of this section, we will describe the methodology in detail. The flow diagram of the study is depicted in Figure 1.

Data Collection and User Selection

The selection of the tweets and their users was based on the filtered real-time streaming support provided by the Twitter API. In the first step, we selected the users who showed potential signs of depression on Twitter on the basis of the 20 most frequent words in Spanish expressed by patients suffering from depression in clinical settings. These words were jointly identified and selected by a psychologist and a family physician with clinical experience and were based on the definition and general features of depression according to the Diagnostic and Statistical Manual of Mental Disorders [42]. The list of words used and their English translations are shown in Textbox 1.

Figure 1. Flow diagram of the study process.

Textbox 1. List of Spanish words related to depression and their English translations.

- agobiado/a (overwhelmed)
- agotado/a (exhausted)
- angustiado/a (distressed)
- ansiedad (anxiety)
- ansioso/a (anxious)
- cansado/a (tired)
- decaído (low)
- depresión (depression)
- depresivo/a (depressed as a condition)
- deprimido/a (depressed as state)
- desanimado/a (discouraged)
- desesperado/a (desperate)
- desmotivado/a (demotivated)
- insomnio (insomnia)
- llorar (cry)
- nervioso (nervous)
- preocupado/a (worried)
- solo/a (lonely)
- triste (sad)
- vacío/a (empty)

During June 2018, 1,470,000 tweets, including 1 or more occurrences of the words listed in Textbox 1, were collected. From this collection of tweets and to select the users who publicly stated in the textual description associated to their profile that they suffered from depression, all the profile descriptions, including 1 or more occurrences of the word "depr" and all the possible derivations related to the word depression in Spanish, such as "depre," "depresión," "depresivo," "depresiva," "deprimido," and "deprimida," were considered. From the 720 users who included 1 or more of these words in their description profile, 90 users who stated they suffered from depression or were receiving treatment for depression were selected for the analysis. This selection was performed by a psychologist, verifying that the statements were related to real expressions of depression, excluding quotes, jokes, or fake ones. For each of these depressed Twitter users, we collected all the most recent tweets from their timeline, up to a maximum of about 3200 tweets. Thus, a total of 189,669 tweets were collected, a figure that was reduced to 140,946 after discarding the retweets. These 140,946 tweets constituted the *depressive users dataset*. Examples of sentences appearing in the user profiles that were used for selecting the depressive users are:

- "Paciente psiquiátrico con depresión crónica" (*Psychiatric patient with chronic depression;* example of a profile sentence that indicates depression).
- "Colecciono errores traducidos a tweets depresivos y a uno que otro impulso de amor" (*I gather errors translated into depressing tweets and into one or another love impulse*;

example of a profile sentence that does not indicate depression).

Once the users with profile sentences indicating depression had been retrieved, their Twitter timelines were collected. Only those users having in their timeline at least 10 tweets that suggested signs of depression were retained for further analyses. For each user, the selection of these tweets was performed by manually inspecting the tweets of the user's complete timeline in reverse temporal order, starting from the most recent one to the oldest tweet of the timeline retrieved by means of the Twitter API . Finally, a total number of 1000 tweets issued by the 90 depressive users, suggesting signs of depression, were detected and used for the analysis. This set of tweets provided us with the *depressive tweets dataset*, which was used to analyze linguistic features of tweets showing signs of depression. It has to be mentioned that these 1000 tweets were not to be included in the depressive users dataset (see Figure 1). At the same time, more than 97,500,000 tweets were also collected in June 2018: such tweets were gathered by listening to the public Twitter stream during this time span by only considering tweets with Spanish textual contents (as detected by Twitter language identification support).

Given that Twitter requires more restrictive filters than just the language of the tweets, we used a list of the most frequently used Spanish words (stopwords) to retrieve all tweets that included 1 or more of these words. The vast majority of Spanish tweets should match this criterion. A sample of 450 users who did not mention in their profile the word depression and its

derivations were selected randomly from the 97,500,000 tweets. The complete timelines of these users were compiled (1,141,021 tweets), which were reduced to 712,589 once retweets were removed. These 712,589 tweets constituted the *control dataset*. To identify the language of a tweet, we relied on the language automatically identified by Twitter for each tweet, selecting tweets in Spanish. It has to be noted that these data can contain some tweets from unidentified depressive users.

Data Analysis

A comparison of the 3 datasets was performed to determine the existence of differential linguistic and behavioral features. The different features that were analyzed are shown in Table 1.

The textual content of each tweet was analyzed by means of the following sequence of steps:

- Tokenization performed by means of a custom Twitter tokenizer included in the Natural Language Toolkit [43].
- Part-of-Speech (POS) tagging performed by means of the Freeling Natural Language Processing tool in order to analyse the usage patterns of grammatical categories (eg, adjectives, nouns, or pronouns) in the text of tweets [44].

- Identification of negations performed by relying on a custom list of Spanish negation expressions, such as *nada* (nothing), *nadie* (nobody), *no* (no), *nunca* (never), and alike.
- Identification of occurrences of positive and negative words inside the text of each tweet by means of 2 Spanish polarity lexicons: the Spanish Sentiment Lexicon and the Spanish SentiCon Lexicon [45,46]. We exploited 2 lexicons to consider and compare 2 approaches of modeling polarity in Spanish texts, thus reducing any language modeling bias that the use of a single language resource could introduce.
- Identification of words and expressions associated to the basic emotions [47] by using the Spanish Emotion Lexicon [48]. Such emotions are *alegría* (happiness), *enojo* (anger), *miedo* (fear), *repulsión* (disgust), *sorpresa* (surprise), and *tristeza* (sadness).

All the tools and aforementioned resources are publicly available. The statistical analyses were carried out with the R version 3.4.3 (R Development Core Team) and SPSS Statistics version 23.0 (IBM), applying the relevant test for each type of comparison to be carried out.

Table 1. Characteristics of the tweets analyzed.

Feature	Analyses performed
Distributions over time	• Tweets throughout the day (per hour) • Tweets throughout the week
Part-of-Speech	• Number of words by grammatical categories (part-of-speech tags) • Number of personal pronouns
Counts	• Number of characters • 200 most frequent words (word cloud) • Number of hashtags, links, mentions, and emojis
Emotion analysis	• Emotion types and frequencies
Negations	• Negation words types and frequencies
Polarity analyses	• Polarity of tweets on the basis of Spanish Sentiment Lexicon and Spanish SentiCon Polarity

Ethical Approval

The protocol used in this study was approved by the Ethics Committee of Parc Salut Mar (approval number 2017/7234/1).

Results

Distribution Over Time

Regarding the distribution of tweets over time, the number of tweets per hour and throughout the week of control and depressive users datasets were compared. The tweet hours were adjusted by the user's time zone. As shown in Figure 2, the

depressive users are less active in generating tweets than the control ones, reaching both groups the same activity level between 23:00 and 6:00. The comparison of the temporal distributions of tweets between both datasets was carried out by means of a repeated measures analysis of variance (Greenhouse-Geisser F=6.605; *P*<.001). As shown in Figure 3, the activity throughout the week of the depressive users dataset presented more regular activity than the control dataset, whose users' activity showed a sharp drop during the weekend. The differences between these datasets were statistically significant (Greenhouse-Geisser F=4.153; *P*=.008).

Figure 2. Number of tweets and retweets per hour of the control and depressive users datasets (mean±standard error of mean). SEM: standard error of mean.
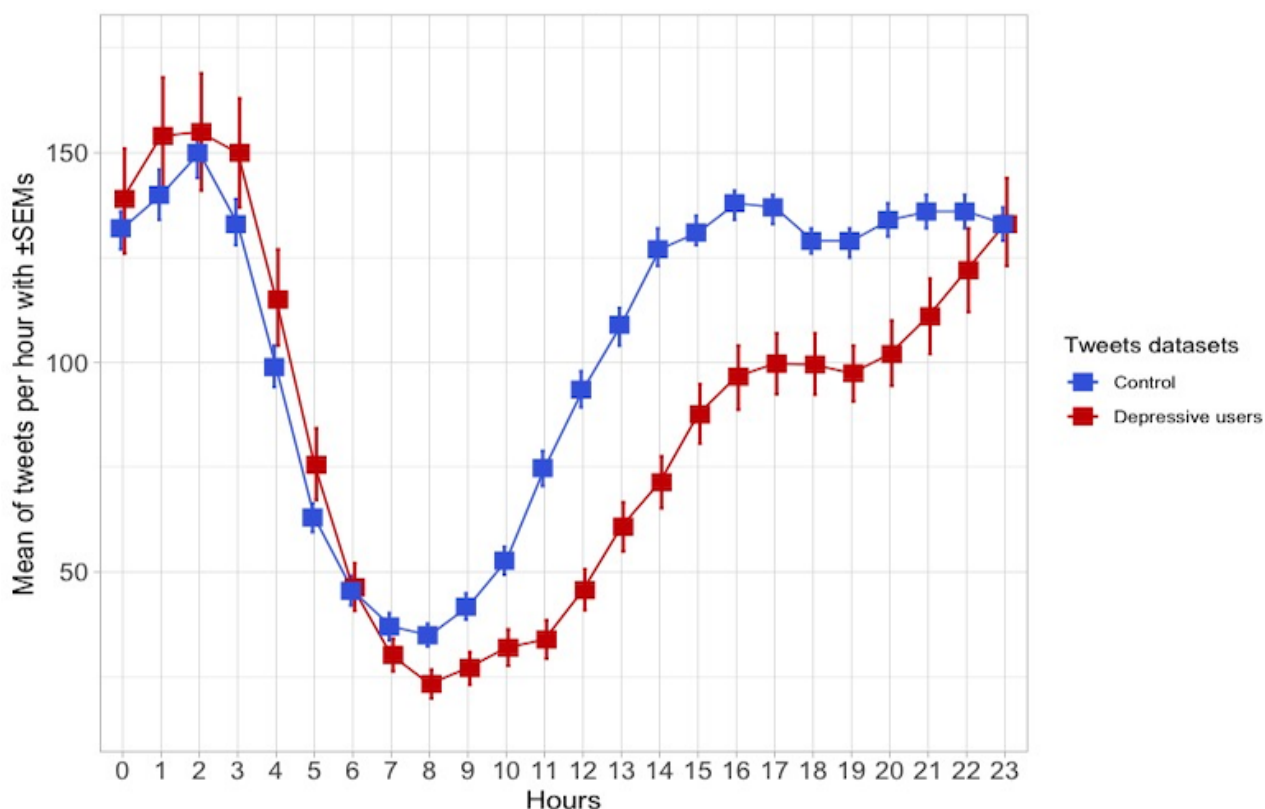
Figure 3. Number of tweets and retweets throughout the week of the control and depressive users datasets (mean±standard error of mean). SEM: standard error of mean.
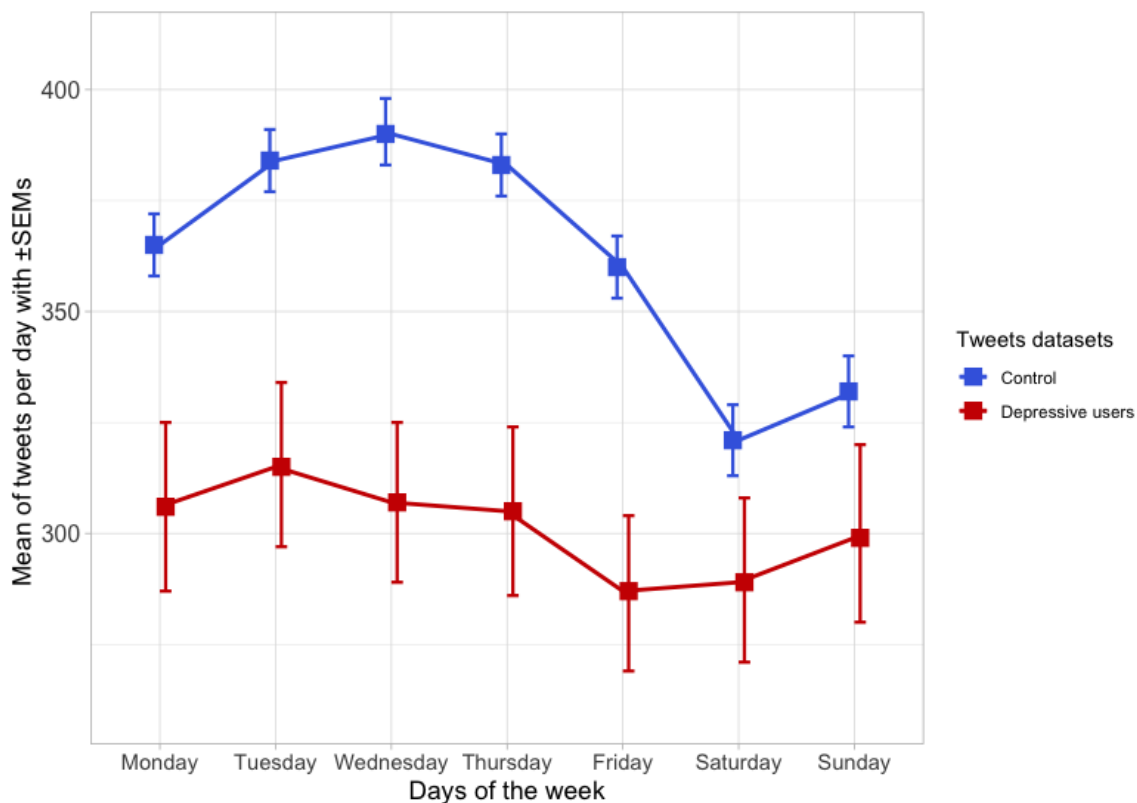
Part-of-Speech

As for the analysis of POS corresponding to the number of words by grammatical categories in each tweet, we compared the 3 datasets of tweets: the control, depressive users, and depressive tweets datasets. As previously stated, the tweets of the depressive tweets dataset were removed from the depressive users dataset. The frequencies of words in each group are shown in Table 2. The number of nouns used in the control group almost doubles that of the depressive users dataset. By contrast, verbs are more frequently used in the depressive users dataset than in the control dataset. There were statistically significant differences between the control and the depressive users datasets (χ^2_7=1,242,600; P<.001), between the control and the depressive tweets datasets (χ^2_7=2,105.7; P<.001), and between the depressive users and the depressive tweets datasets (χ^2_7=15,888; P<.001).

In relation to the different types of pronouns in the control dataset, we detected 396,181 personal pronouns (51.38%; 396,181/770,955), the first-person singular (38.37%; 152,013/396,181) being the most used. A similar profile was observed in the depressive users dataset, where 124,614 pronouns were found (55.16%; 124,614/225,913), the first-person singular remaining the most used (57.59%; 71,768/124,614). In the depressive tweets dataset, 865 personal pronouns (53.16%; 865/1,627) were identified, and the first-person singular pronoun was by far the most used (80.00%; 692/865). The frequencies of personal pronouns in the different datasets are shown in Figure 4. There were statistically significant differences between the control and the depressive users datasets (χ^2_5=15,912; P<.001), between the control and the depressive tweets datasets (χ^2_5=638.7; P<.001), and between the depressive users and the depressive tweets datasets (χ^2_5=183.9; P<.001).

In relation to the number of characters per tweet, the mean of characters per tweet in the control and depressive users datasets was 83.48 (SD 40.57) and 65.76 (SD 36.99) characters, respectively, with statistically significant differences between them (t_{213770}=161.6; P<.001). On the other hand, the mean in the depressive tweets dataset was 67.51 (SD 38.28), which was not statistically significant and different in comparison with the depressive dataset ($t_{1012.3}$=1.45; P=.15). The 200 most frequent words that appeared in the control and depressive users datasets are depicted in the 2 word clouds shown in Multimedia Appendix 1. The 10 most frequent words that appeared in the control dataset were the following: *hoy* (today), *día* (day), *ver* (to see), *quiero* (I want), *gracias* (thank you), *mejor* (better), *siempre* (always), *vida* (life), *ahora* (now), and YouTube. In the depressive users dataset, the 10 most frequent words were the following: *quiero* (I want), *vida* (life), *siempre* (always), *siento* (I feel), *nadie* (nobody), *mierda* (shit), *never* (nunca), and *día* (day). It should be noted that in the depressive tweets dataset, although there are several words in common with the depressive users dataset, we can find additional words that are not present in the other datasets, such as *vacío/a* (empty), *matar* (to kill), *desaparecer* (to disappear), *suicidar* (commit suicide), *muerta* (dead), *desastre* (disaster), *inútil* (useless,), *deprimida* (depressed as state in women), *depresiva* (depressed as a condition in women), and *insomnio* (insomnia). The word cloud of the depressive tweets dataset is shown in Multimedia Appendix 2. In relation to the use of links, hashtags, and mentions in tweets, the frequency of them in the control and depressive users datasets were 35.32% (251,728/712,584), 13.13% (93,575/712,588), and 44.00% (313,574/712,577) and 18.07% (25,475/140,946), 1.44% (2030/140,946), and 9.27% (13,060/140,942), respectively. The number of tweets, including emojis, were 13.61% (97,038/712,589) in the control dataset and 5.72% (8069/140,947) in the depressive users dataset.

Table 2. Part-of-Speech (POS) frequencies in tweets of control, depressive users, and depressive tweets datasets.

| Type of POS	POS in the control dataset, n (%)	POS in the depressive users dataset, n (%)	POS in the depressive tweets dataset, n (%)
Noun	2,298,544 (28.48)	270,104 (17.77)	1776 (15.07)
Verb	1,660,700 (20.58)	400,755 (26.36)	3391 (28.77)
Pronouns	770,955 (9.55)	225,913 (14.86)	1627 (13.80)
Adjectives	593,327 (7.35)	83,089 (5.47)	588 (4.99)
Determiner	1,068,130 (13.23)	177,795 (11.70)	1342 (11.39)
Adverbs	496,988 (6.16)	140,963 (9.27)	1351 (11.46)
Adpositions	854,573 (10.59)	123,867 (8.15)	1052 (8.93)
Conjunctions	327,852 (4.06)	97,541 (6.42)	659 (5.59)

Figure 4. Frequency of the different types of personal pronouns in the control, depressive users, and depressive tweets datasets. 1S: first-person singular; 1P: first-person plural; 2S: second-person singular; 2P: second-person plural; 3S: third-person singular; 3P: third-person plural.
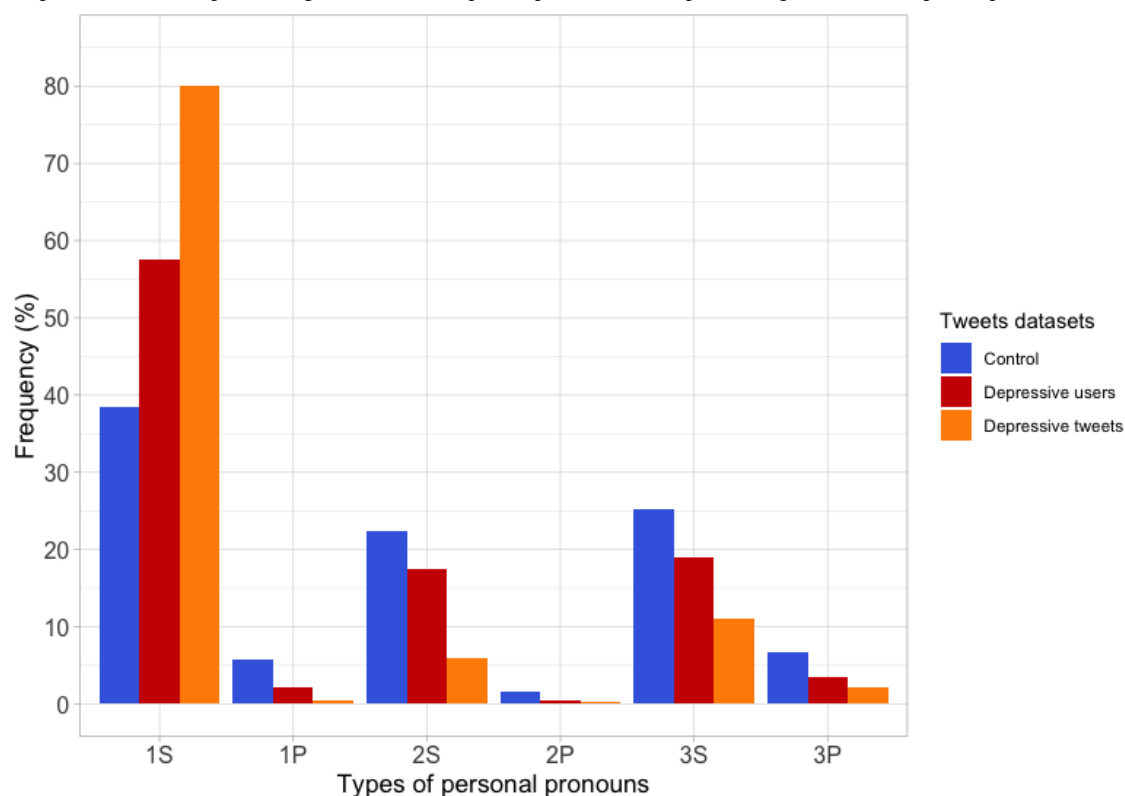
Emotion Analysis

Regarding the distribution of emotions, in the control dataset and in the depressive users dataset, the most frequent emotion was happiness (53.30%; 203,029/380,874 and 41.60%; 40,535/97,425) followed by sadness, which was more frequent in the depressive users dataset (17.59%; 67,033/380,874 and 25.49%; 24,834/97,425). In the depressive tweets dataset, the most frequent emotion was sadness (34.00%; 303/891). There were statistically significant differences between the control and the depressive users datasets (χ^2_5=6838.2; P<.001), between the control and the depressive tweets datasets (χ^2_5=296.8; P<.001), and between the depressive users and the depressive tweets datasets (χ^2_5=65.6; P<.001). The frequencies of the different emotions are shown in Figure 5.

Figure 5. Frequency distributions of emotion categories in the tweets of the 3 datasets.
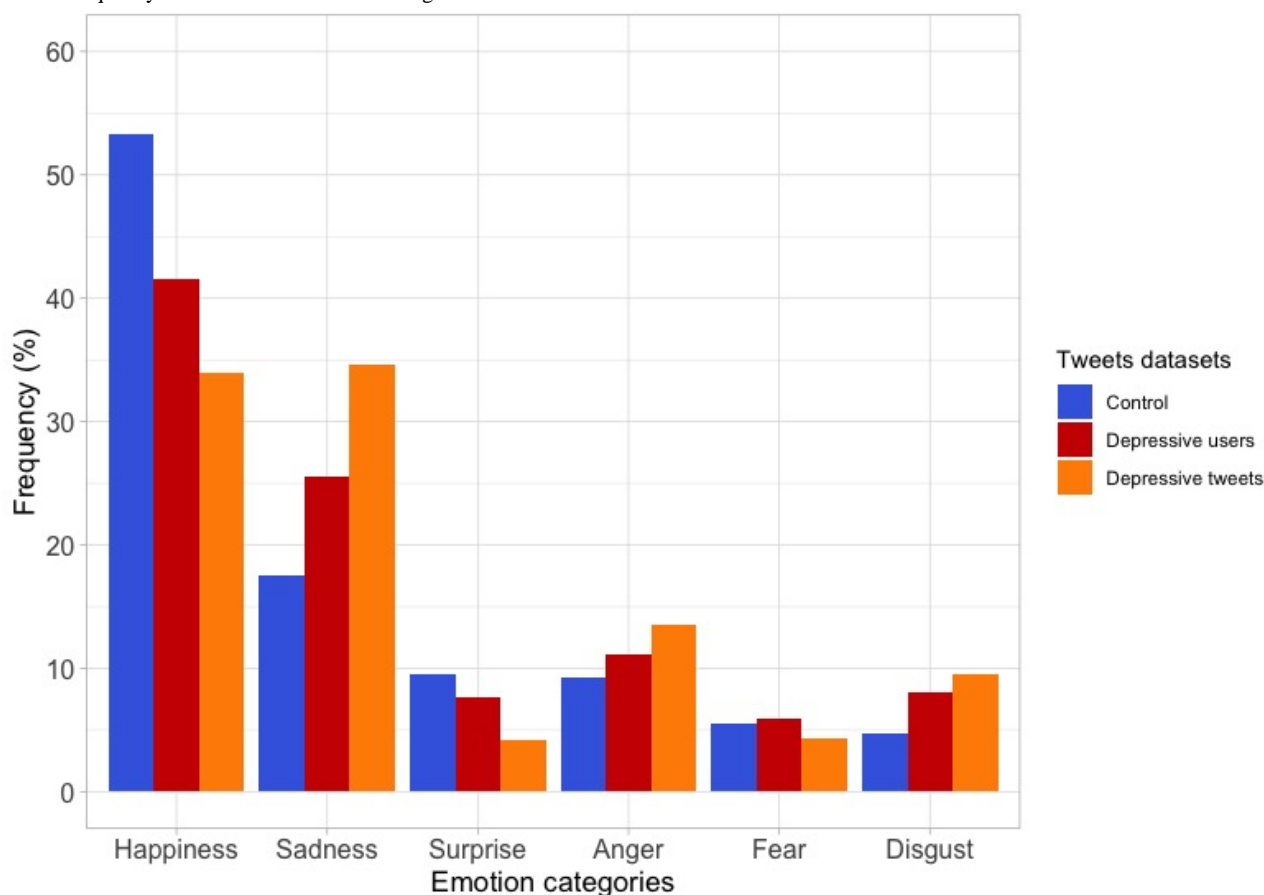
Negation Words

Regarding the use of negation words, they were detected in 21.74% (154,953/712,588) of the tweets in the control dataset, in 34.15% (48,137/140,946) of the depressive users dataset, and in 45.50% (455/1000) of the depressive tweets dataset. The mean of negation words was 0.28 (SD 0.59) in the control dataset, it was 0.49 (SD 0.82) in the depressive users dataset, and it was 0.67 (SD 0.91) in the depressive tweets dataset. There were statistically significant differences between the control and the depressive users datasets (Mann-Whitney U=4.3657e+10; P<.001), between the control and the depressive tweets datasets (Mann-Whitney U=266,990,000; P<.001), and between the depressive users and the depressive tweets datasets (Mann-Whitney U=62,002,000; P<.001).

Polarity Analysis

In relation to the polarity of tweets, 2 analyses were performed using 2 Spanish sentiment lexicons: the Senti Lexicon (including positive and negative categories) and the SentiCo Polarity (including positive, moderate positive, moderate negative, and negative categories). According to the Senti Lexicon, the analysis of tweets showed that the control dataset shows polarity in 33.47% (245,367/733,029) of the tweets, being positive in 56,54% (138,726/245,367) of them. In contrast, the depressive users dataset shows polarity in 41,31% (61,132/147,996) of the tweets, being positive in 46.14% (28,205/61,132) of them. Finally, the depressive tweets dataset shows polarity in 58.90% (589/1000) of the tweets, with positive polarity in 34.97% (206/589) of them. There were statistically significant differences between the control and the depressive users datasets (χ^2_1=2134; P<.001), between the control and the depressive tweets datasets (χ^2_1=110.3; P<.001), and between the depressive users and the depressive tweets datasets (χ^2_1=28.8; P<.001). When using the SentiCo Polarity tool, the control dataset presented 20.97% (152,228/725,717) of tweets with polarity, 29.32% (42,820/146,033) in the depressive users and 33.34% (348/1,044) in the depressive tweets dataset. The distributions of polarities are shown in Figure 6. There were statistically significant differences between the control and the depressive users datasets (χ^2_3=8820.8, P<.001), between the control and the depressive tweets datasets (χ^2_3=308.8; P<.001), and between the depressive users and the depressive tweets datasets (χ^2_3=52.4; P<.001).

Figure 6. Polarities of the tweets according to the SentiCo Polarity tool in the 3 datasets.
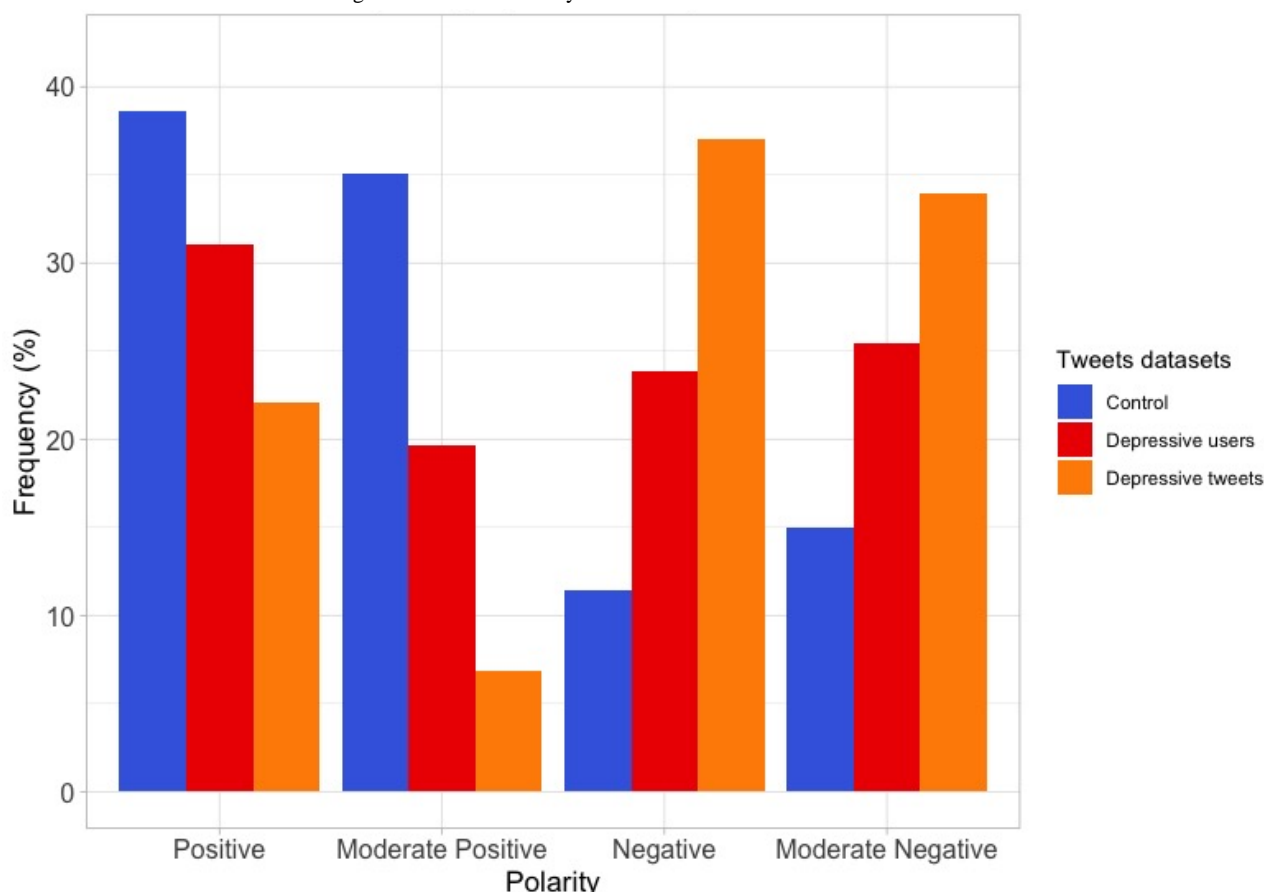
Discussion

Principal Findings

The diagnosis of depression is complex because of the heterogeneous nature of this disease and the diverse manifestation of the symptoms among individuals, which result in a great number of depressive disorder cases that are undetected and untreated, making the prevention, diagnosis, and treatment of the depressive disorders a complicated task [15,49,50]. For these reasons and taking into account that people diagnosed with depression are increasing worldwide, new strategies for detecting and monitoring this disease would be very useful. In this study, we analyzed the behavioral and linguistic patterns of tweets in Spanish that suggest signs of depression. The results contribute to the growing body of scientific literature that analyzes the messages posted on social media using languages other than English. We have introduced a new approach that comprises analyzing the timelines of self-qualified depressed users, as well as their tweets related to depression, which are manually selected. Our results show that the tweets of depressive users have different features in comparison with those of a control dataset, even when their tweets that are not related to depression are considered (depressive users dataset). In addition, the differences with the control dataset become more evident when we consider the manual selection of tweets related to depression (depressive tweets dataset).

Different Distributions of Tweets Over Time

As for the distribution of tweets over time, the users of the depressive dataset, although being less active in using Twitter, used to be more active during the night than the users of the control dataset. This can be explained as a result of insomnia, one of the most frequent symptoms of depression. This finding is consistent with previous studies carried out with English speakers, which demonstrated that individuals with depression are more active during the night [20]. Moreover, the daily mood changes, such as the morning and evening worsening that are typical in several forms of depression, could explain the lower activity of the depressive users [51]. In relation to the distribution of tweets throughout the week, the users of the depressive dataset showed a more regular activity throughout the week, tending to be more active on Saturdays, Sundays, and Mondays than those of the control dataset, whose activity showed a drop during the weekend. This trend may be related to the lowered social activity of the people suffering from depressive disorders, having a reduced participation in social leisure activities during the weekend and spending more time at home, sharing their feelings and thoughts on social media platforms [16].

Different Style of Writing

The analysis of POS and the number of words by grammatical categories show that, generally, the users of the depressive dataset used more verbs, adverbs, and pronouns but less nouns than the control dataset. The same features are also present in the depressive tweets dataset. These findings suggest that the

language of people suffering from depression is characterized by a different style of writing that some authors describe as poorly structured, indicating less interest in what surrounds them, people, objects, or things [52]. They focus on talking about actions, and this is correlated with sensitive disclosure. Consistent with many previous studies [20,30,35,53-55], the use of first-person singular pronoun is more frequent among the users of the depressive dataset, with respect to those of the control dataset, and this difference increases in the depressive tweets dataset. The increased use of this pronoun demonstrates the attention to self-focus that is associated with the negative emotional states of depression and the reduced attentional resources, highlighting the psychological distancing to connect with others [56]. This social isolation may also explain that the first- and second-person plural pronouns are the least used. Language can be used as a measure of different individual features, on the basis of the fact that people's word choice is stable over time and consistent across topics or context. For this reason, the language style appears to be a useful predictor of some mental health conditions, such as depression [20,35]. In addition, the number of characters written in the depressive users and depressive tweets datasets was smaller than the number of characters written in the control dataset, and this might be related to reduced interest and poorer language. According to the most frequent words that appeared in the depressive users and depressive tweets datasets, there are specific words that are linked to clinical symptoms and the way that depressive patients word their mood, such as words that may be related to suicide ideation. Consequently, they can be used as a signal to detect potentially depressed users on Twitter [36]. Similarly, we observed the frequent use of adjectives in feminine form in the depressive tweets dataset, which would suggest that a high proportion of the depressive users are women, a fact that is in agreement with clinical and epidemiological evidences [8,11,12,42].

Predominant Emotions

Emotions are one of the key aspects that characterize many mental health conditions and, particularly, when people are suffering from depression. An analysis of the 6 emotions that are commonly considered (happiness, sadness, surprise, anger, fear, and disgust) [47] was performed to determine the existence of differences among the datasets. Happiness is the most frequent emotion in the control and depressive users datasets, although an important reduction was observed in the depressive tweets dataset. The surprise emotion is less frequent in depressive users and, specially, in the depressed tweets datasets than the control dataset, and this fact can be related to the depressive mood, in which there is a decrease in interest in almost everything. Fear does not seem to be a differential emotion in the groups of tweets analyzed in this study.

Regarding negative emotions, we observed an increase in the frequency of words related to the sadness emotion in the depressive tweets dataset, doubling that of the control dataset. This feature had also been observed in other studies [14,35,57]. Moreover, anger is more frequent in the depressive user and depressive tweets datasets than in the control dataset. Although Twitter is used many times for expressing anger about personal or political aspects, this emotion is particularly frequent in

patients suffering from depression, who tend to feel irritable, wronged, or angry at the world [14,16,35,58]. At the same time, disgust, an emotion that is known to be associated with the depressive disorders [59], was found to be more frequent in the depressive users and depressive tweets datasets.

Negative Focused Emotion Language

In our analysis, the presence of negation words is more frequent in the depressed users (34.15%; 48,137/140,946) and depressive tweets (45.50%; 455/1000) datasets than in the control dataset (21.74%; 154,953/712,588), indicating that there is an increased use of negatively focused emotion language, which is typical in depressive patients and feelings [31,54,55,60].

Negative Polarity

The classification of tweets, on the basis of the emotional positivity or negativity of their words, is another analysis that has been carried out. In this study, we used 2 types of polarity lexicons, the Senti Lexicon (SentiLex) and the Sentico Polarity (SentiCo). In both cases, the negative polarity was higher in the depressive users and depressive tweets datasets, even tripling the negativeness of the control dataset when using the Sentico Polarity lexicon. These findings are consistent with other studies, indicating that people suffering from depression tend to focus more on negative aspects of their life [20,35], and thus their tweets contain much more negative emotional words compared with the control dataset [14]. In addition, the self-focus state that characterizes depression is associated with negative emotions [32,56,57].

Limitations and Future Directions

This study presents some limitations that have to be pointed out. On the one hand, the tweets of the depressive datasets come only from Twitter users who speak publicly about feelings and emotions that can be related with depression. This is an indirect and inaccurate way of detecting users suffering from depressive disorders. Without clinically assessing these people, there is no way to verify if the diagnosis is genuine or if they suffer from another mental disorder. On the other hand, it is possible that Twitter users self-disclose their mental health using words or expressions not included in the list of keywords used in this study for streaming tweets about depression [22,61-63]. In this respect, it is possible that a wider list could have yielded a greater coverage [21,36]. Privacy policies of social media restrict the access to users who did not grant access to their profile, and this may have generated biases in the composition of the depressive users and the depressive tweets datasets. In addition, tweets may incorporate biases because of the self-management and anonymity of the Web-based identities [61]. Moreover, Twitter users may be not be representative of the general population, and some studies have shown that they are often urban people with high levels of education [64-66]. More information about the socioeconomic and demographic details of Twitter users is needed [67]. The control dataset was a randomly selected sample of Twitter users, and it is consequently representative of the users of this social media. However, there is a possibility that users in this group may also have depression or other mental illness even though they did not mention this in their profile description. There is also the

possibility that the users included in the control group are fake accounts. Only original tweets were analyzed, and perhaps retweets, which are not included in our linguistic study, reflect users' emotions that can be related to depression status [68]. Finally, depression is a very complex mental disorder, and our study only provides a general observation of this disorder. Additional research might be carried out to examine specific depression types and determine if there are social media features that can contribute to classifying users or tweets to the different diagnosis of depression [69]. Similarly, in future works, we plan to study the linguistic features and the behavioral patterns of depression in different linguistics contexts. The possible relationship between depression and seasonality could be of interest for future studies in the context of monitoring Twitter activity [70].

Conclusions

The prevalence of common mental disorders worldwide, such as depression, requires the ability of health care systems to provide adequate diagnosis, monitoring, and treatment. The wide popularity of social media platforms introduces new opportunities for the screening of depression. The introduction of new methods of analysis for the automatic detection of signals of depression on social media platforms, such as Twitter or Facebook, has the potential of being used as a complementary tool for the assessment of these patients, assisting health care professionals in the detection and monitoring of mental health disorders. Although the analysis of tweets as a way to determine the existence of depression cannot be used as a replacement for diagnosis, it has the potential as a screening tool for depressive disorders, with a lower cost than other traditional procedures. In addition, it can be helpful to health professionals for managing and monitoring patients more efficiently. Similarly, it can be useful for particular patients, as they feel more comfortable disclosing their symptoms on Twitter than in clinical settings. In this study, we have shown that several behavioral and linguistic features of the tweets in Spanish can be used as a complementary tool to detect signals of depression of their authors, corroborating and extending the findings obtained by studies carried out on English tweets. As we described in this study, signs of depression of Twitter users are not exclusively spotted by identifying and analyzing tweets that explicitly mention expressions related to depression. Moreover, Twitter users who are potentially suffering from depression globally modify the core traits of their language, independently from the fact that the tweets are related or not related to the expression of depression. On the basis of these changes, these users can be monitored and supported. The results of this paper, jointly with other studies on the matter, support the potential of social media as an important instrument for extending and enhancing mental health services available to people with mental disorders. By means of interdisciplinary collaborations, it is possible to develop digital apps and services providing personalized alerts and psychosocial support in the mental health domain.

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Word clouds showing the 200 most frequent words in the control (left) and depressive users (right) datasets.

[PNG File, 494KB - jmir_v21i6e14199_fig5.png]

Multimedia Appendix 2

Word cloud showing the 200 most frequent words in the depressive tweets dataset.

[PNG File, 195KB - jmir_v21i6e14199_fig6.png]

References

1. World Health Organization. 2013. Mental Health Action Plan 2013-2020 URL: http://apps.who.int/iris/bitstream/10665/89966/1/9789241506021_eng.pdf?ua=1 [accessed 2018-04-25] [WebCite Cache ID 778TsWziL]
2. Marcus M, Yasamy MT, van Ommeren M, Chisholm D, Saxena S, WHO Department of Mental Health and Substance Abuse. World Health Organization. 2012. Depression: A Global Public Health Concern URL: http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf [accessed 2018-04-25] [WebCite Cache ID 778NmzRq1]

XSL•FO
RenderX

3.   Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the global burden of disease study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]

4.   Trautmann S, Rehm J, Wittchen HU. The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016;17(9):1245-1249 [FREE Full text] [doi: 10.15252/embr.201642951] [Medline: 27491723]

5.   Patel V, Chisholm D, Parikh R, Charlson FJ, Degenhardt L, Dua T, et al. Global priorities for addressing the burden of mental, neurological,substance use disorders. In: Patel V, Chisholm D, Dua T, Laxminarayan R, Medina-Mora ME, Vos T, editors. Disease Control Priorities: Mental, Neurological, and Substance Use Disorders, Third Edition (Volume 4). Washington, DC: The World Bank; 2016:1-27.

6.   Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. Lancet 2013 Nov 9;382(9904):1575-1586. [doi: 10.1016/S0140-6736(13)61611-6] [Medline: 23993280]

7.   Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Dec 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]

8.   World Health Organization. 2018. Depression: Key Facts URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2018-04-23] [WebCite Cache ID 778TC0BMP]

9.   Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006 Nov;3(11):e442 [FREE Full text] [doi: 10.1371/journal.pmed.0030442] [Medline: 17132052]

10.   Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry 2016 Feb;3(2):171-178. [doi: 10.1016/S2215-0366(15)00505-2] [Medline: 26851330]

11.   World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates URL: http://apps.who.int/iris/bitstream/10665/254610/1/WHO-MSD-MER-2017.2-eng.pdf?ua=1 [accessed 2018-04-25] [WebCite Cache ID 77CsSorTr]

12.   Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]

13.   Paul MJ, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011 Presented at: AAAI'11; July 17-21, 2011; Barcelona, Spain p. 265-272.

14.   Park M, Cha C, Cha M. Depressive Moods of Users Portrayed in Twitter. In: Proceedings of the ACM SIGKDD Workshop on Health Informatics. 2012 Presented at: HI-KDD'12; August 12-16, 2012; Beijing, China p. 1-8.

15.   Nguyen T, O'Dea B, Larsen M, Phung D, Venkatesh S, Christensen H. Using linguistic and topic analysis to classify sub-groups of online depression communities. Multimed Tools Appl 2015 Dec 21;76(8):10653-10676. [doi: 10.1007/s11042-015-3128-x]

16.   Cavazos-Rehg PA, Krauss MJ, Sowles S, Connolly S, Rosas C, Bharadwaj M, et al. A content analysis of depression-related Tweets. Comput Human Behav 2016;54:351-357 [FREE Full text] [doi: 10.1016/j.chb.2015.08.023] [Medline: 26392678]

17.   Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: Analyzing the Language of Mental Health on Twitter Through Self-Reported Diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: CLPSYCH'15; June 5, 2015; Denver, Colorado p. 1-10 URL: https://www.aclweb.org/anthology/W15-1201 [doi: 10.3115/v1/W15-1201]

18.   Naslund JA, Aschbrenner KA, McHugo GJ, Unützer J, Marsch LA, Bartels SJ. Exploring opportunities to support mental health care using social media: a survey of social media users with mental illness. Early Interv Psychiatry 2019 Jun;13(3):405-413. [doi: 10.1111/eip.12496] [Medline: 29052947]

19.   Berry N, Lobban F, Belousov M, Emsley R, Nenadic G, Bucci S. #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. J Med Internet Res 2017 Dec 5;19(4):e107 [FREE Full text] [doi: 10.2196/jmir.6173] [Medline: 28381392]

20.   De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting Depression via Social Media. In: Proceedings of the Seventh International Conference on Weblogs and Social Media. 2013 Presented at: AAAI'13; July 8-11, 2013; Cambridge, MA p. 128-138.

21.   Wilson ML, Ali S, Valstar MF. Finding Information About Mental Health in Microblogging Platforms: A Case Study of Depression. In: Proceedings of the 5th Information Interaction in Context Symposium. 2014 Presented at: IIiX'14; August 26-30, 2014; Regensburg, Germany p. 8-17. [doi: 10.1145/2637002.2637006]

22.   Coppersmith G, Dredze M, Harman C. Quantifying Mental Health Signals in Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: CLPSYCH'14; June 5, 2014; Baltimore, Maryland p. 51-60. [doi: 10.3115/v1/W14-3207]

23. Radzikowski J, Stefanidis A, Jacobsen KH, Croitoru A, Crooks A, Delamater PL. The measles vaccination narrative in Twitter: a quantitative analysis. JMIR Public Health Surveill 2016;2(1):e1 [FREE Full text] [doi: 10.2196/publichealth.5059] [Medline: 27227144]

24. Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, et al. Zika in Twitter: temporal variations of locations, actors, and concepts. JMIR Public Health Surveill 2017 Apr 20;3(2):e22 [FREE Full text] [doi: 10.2196/publichealth.6925] [Medline: 28428164]

25. Finch KC, Snook KR, Duke CH, Fu K, Tse ZT, Adhikari A, et al. Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns. Nat Hazards 2016 Apr 19;83(1):729-760. [doi: 10.1007/s11069-016-2327-8]

26. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

27. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]

28. Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. Nat Biotechnol 2015 May;33(5):462-463. [doi: 10.1038/nbt.3223] [Medline: 25965751]

29. Statista. 2018. Number of Monthly Active Twitter Users Worldwide From 1st Quarter 2010 to 2nd Quarter 2018 (in Millions) URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ [accessed 2018-10-20] [WebCite Cache ID 778OhlXHI]

30. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009 Dec 8;29(1):24-54 [FREE Full text] [doi: 10.1177/0261927X09351676]

31. Ramirez-Esparza N, Chung CK, Kacewicz E, Pennebaker JW. The Psychology of Word Use in Depression Forums in Englishin Spanish: Texting Two Text Analytic Approaches. In: Proceedings of the 2nd International Conference on Weblogs and Social Media. 2008 Presented at: AAAI'08; March 30–April 2, 2008; Seattle, Washington URL: https://www.aaai.org/Papers/ICWSM/2008/ICWSM08-020.pdf

32. Chung C, Pennebaker J. The psychological functions of function words. In: Fiedler K, editor. Social Communication: Frontiers of Social Psychology. First Edition. New York: Psychology Press; 2007:343-359.

33. Prieto VM, Matos S, Álvarez M, Cacheda F, Oliveira JL. Twitter: a good place to detect health conditions. PLoS One 2014;9(1):e86191 [FREE Full text] [doi: 10.1371/journal.pone.0086191] [Medline: 24489699]

34. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. BMC Cancer 2013 Oct 29;13:508 [FREE Full text] [doi: 10.1186/1471-2407-13-508] [Medline: 24168075]

35. Pavalanathan U, De Choudhury M. Identity Management and Mental Health Discourse in Social Media. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW'15 Companion; May 18-22, 2015; Florence, Italy p. 315-321. [doi: 10.1145/2740908.2743049]

36. De Choundhury M, Counts S, Horvitz E. Social Media as a Measurement Tool of Depression in Populations. In: Proceedings of the 5th Annual ACM Web Science Conference. 2013 Presented at: WebSci'13; May 2-4, 2013; Paris, France p. 47-56 URL: https://dl.acm.org/citation.cfm?id=2464480 [doi: 10.1145/2464464.2464480]

37. Coppersmith G, Harman C, Dredze M. Measuring Post Traumatic Stress Disorder in Twitter. In: Proceedings of the 8th International Conference on Weblogs and Social Media. 2014 Presented at: AAAI'14; June 1-4, 2014; Ann Arbor, Michigan p. 579-582 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8079/8082

38. Birnbaum ML, Ernala SK, Rizvi AF, De Choudhury M, Kane JM. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. J Med Internet Res 2017 Dec 14;19(8):e289 [FREE Full text] [doi: 10.2196/jmir.7956] [Medline: 28807891]

39. Arseniev-Koehler A, Lee H, McCormick T, Moreno MA. #Proana: pro-eating disorder socialization on Twitter. J Adolesc Health 2016;58(6):659-664. [doi: 10.1016/j.jadohealth.2016.02.012] [Medline: 27080731]

40. Statista. 2019. Leading Countries Based on Number of Twitter Users as of January 2019 (in Millions) URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [WebCite Cache ID 77Ctc6VPr]

41. Twitter Developer. URL: https://developer.twitter.com/en.html [accessed 2018-11-10] [WebCite Cache ID 778Qhq4Gu]

42. American Psychiatric Association. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Publishing; 2013.

43. Natural Language Tool Kit. URL: https://www.nltk.org/api/nltk.tokenize.html [accessed 2018-10-10] [WebCite Cache ID 778Rzo6pe]

44. Padró L, Stanilovsky E. FreeLing 3.0: Towards Wider Multilinguality. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation. 2012 Presented at: LREC'12; May 21-27, 2012; Istanbul, Turkey p. 2473-2479 URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/430_Paper.pdf

45. Perez-Rosas V, Banea C, Mihalcea R. Learning Sentiment Lexicons in Spanish. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation. 2012 Presented at: LREC'12; May 21-27, 2012; Istanbul, Turkey p. 3077-3081 URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/1081_Paper.pdf

46. Cruz FL, Troyano JA, Pontes B, Ortega FJ. Building layered, multilingual sentiment lexicons at synset and lemma levels. Expert Syst Appl 2014 Oct;41(13):5984-5994 [FREE Full text] [doi: 10.1016/j.eswa.2014.04.005]

47. Ekman P, Friesen WV, O'Sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K, et al. Universals and cultural differences in the judgments of facial expressions of emotion. J Pers Soc Psychol 1987 Oct;53(4):712-717. [doi: 10.1037/0022-3514.53.4.712] [Medline: 3681648]

48. Sidorov G, Miranda-Jiménez S, Viveros-Jiménez F, Gelbukh A, Castro-Sánchez NA, Castillo F, et al. Empirical Study of Opinion Mining in Spanish Tweets. In: Proceedings of the 11th Mexican International Conference on Artificial Intelligence. 2012 Presented at: MICAI'12; October 27-November 4, 2012; San Luis Potosí, Mexico p. 1-4.

49. Nambisan P, Luo Z, Kapoor A, Patrick TB, Cisler RA. Social Media, Big Data, and Public Health Informatics: Ruminating Behavior of Depression Revealed Through Twitter. In: Proceedings of the 2015 48th Hawaii International Conference on System Sciences. 2015 Presented at: HICSS'15; January 5-8, 2015; Kauai, Hawaii p. 2906-2913. [doi: 10.1109/HICSS.2015.351]

50. Cassano P, Fava M. Depression and public health: an overview. J Psychosom Res 2002 Oct;53(4):849-857. [Medline: 12377293]

51. Morris DW, Rush AJ, Jain S, Fava M, Wisniewski SR, Balasubramani GK, et al. Diurnal mood variation in outpatients with major depressive disorder: implications for DSM-V from an analysis of the sequenced treatment alternatives to relieve depression study data. J Clin Psychiatry 2007 Sep;68(9):1339-1347. [doi: 10.4088/JCP.v68n0903] [Medline: 17915971]

52. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16; May 7-12, 2016; San Jose, California p. 2098-2110 URL: http://europepmc.org/abstract/MED/29082385 [doi: 10.1145/2858036.2858207]

53. Bucci W, Freedman N. The language of depression. Bull Menninger Clin 1981 Jul;45(4):334-358. [Medline: 6176285]

54. Rude S, Gortner EM, Pennebaker J. Language use of depressed and depression-vulnerable college students. Cogn Emot 2004;18(8):1121-1133. [doi: 10.1080/02699930441000030]

55. Pyszczynski T, Greenberg J. Self-regulatory perseveration and the depressive self-focusing style: a self-awareness theory of reactive depression. Psychol Bull 1987 Jul;102(1):122-138. [doi: 10.1037/0033-2909.102.1.122] [Medline: 3615702]

56. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]

57. Sonnenschein AR, Hofmann SG, Ziegelmayer T, Lutz WL. Linguistic analysis of patients with mood and anxiety disorders during cognitive behavioral therapy. Cogn Behav Ther 2018;47(4):315-327. [doi: 10.1080/16506073.2017.1419505] [Medline: 29345528]

58. Painuly N, Sharan P, Mattoo SK. Relationship of anger and anger attacks with depression: a brief review. Eur Arch Psychiatry Clin Neurosci 2005 Aug;255(4):215-222. [doi: 10.1007/s00406-004-0539-5] [Medline: 16133740]

59. Neacsiu AD, Rompogren J, Eberle JW, McMahon K. Changes in problematic anger, shame, and disgust in anxious and depressed adults undergoing treatment for emotion dysregulation. Behav Ther 2018;49(3):344-359 [FREE Full text] [doi: 10.1016/j.beth.2017.10.004] [Medline: 29704965]

60. Bernard JD, Baddeley JL, Rodriguez BF, Burke PA. Depression, language, and affect: an examination of the influence of baseline depression and affect induction on language. J Lang Soc Psychol 2016;35(3):317-326. [doi: 10.1177/0261927X15589186]

61. De Choudhury M, Sharma SS, Logar T, Eekhout W, Nielsen RC. Gender and Cross-Cultural Differences in Social Media Disclosures of Mental Illness. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW'17; February 25-March 1, 2017; Portland, Oregon p. 353-369. [doi: 10.1145/2998181.2998220]

62. De Choudhury M, De S. Mental Health Discourse on Reddit: Self-Disclosure, Social Support, and Anonymity. In: Proceedings of the 8th International Conference on Weblogs and Social Media. 2014 Presented at: AAAI'14; June 1-4, 2014; Ann Arbor, Michigan p. 71-80 URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8075/8107

63. Mowery D, Smith H, Cheney T, Stoddard G, Coppersmith G, Bryan C, et al. Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. J Med Internet Res 2017 Dec 28;19(2):e48 [FREE Full text] [doi: 10.2196/jmir.6895] [Medline: 28246066]

64. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]

65. Yom-Tov E, Johansson-Cox I, Lampos V, Hayward AC. Estimating the secondary attack rate and serial interval of influenza-like illnesses using social media. Influenza Other Respir Viruses 2015 Jul;9(4):191-199 [FREE Full text] [doi: 10.1111/irv.12321] [Medline: 25962320]

66. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]

67. Chary M, Genes N, McKenzie A, Manini AF. Leveraging social networks for toxicovigilance. J Med Toxicol 2013 Jun;9(2):184-191 [FREE Full text] [doi: 10.1007/s13181-013-0299-6] [Medline: 23619711]

68.   Seabrook EM, Kern ML, Fulcher BD, Rickard NS. Predicting depression from language-based emotion dynamics: longitudinal analysis of Facebook and Twitter status updates. J Med Internet Res 2018 Dec 8;20(5):e168 [FREE Full text] [doi: 10.2196/jmir.9267] [Medline: 29739736]
69.   Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Dec 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
70.   Alvarez-Mon MA, Asunsolo Del Barco A, Lahera G, Quintero J, Ferre F, Pereira-Sanchez V, et al. Increasing interest of mass communication media and the general public in the distribution of Tweets about mental disorders: observational study. J Med Internet Res 2018 Dec 28;20(5):e205 [FREE Full text] [doi: 10.2196/jmir.9582] [Medline: 29807880]

Abbreviations

API: Application Programming Interface
POS: Part-of-Speech
WHO: World Health Organization