Social Media, Big Data and Public Health Informatics: Ruminating behavior of depression revealed through Twitter

Priya Nambisan, PhD
University of Wisconsin - Milwaukee
nambisap@uwm.edu

Zhihui Luo, PhD
University of Wisconsin - Milwaukee
jakeluo@uwm.edu

Akshat Kapoor, M.S.
University of Wisconsin - Milwaukee
akapoor@uwm.edu

Timothy B Patrick, PhD
University of Wisconsin - Milwaukee
tp5@uwm.edu

Ron A Cisler, PhD
University of Wisconsin - Milwaukee
rac@uwm.edu

Abstract

Undiagnosed and untreated depressive disorders have become a serious public health issue and it is prevalent among people of all ages, gender and race. Social media sites, such as Twitter, have become a major venue for people to express/disclose their thoughts and feelings. The tweets from these micro-blogging sites could be used to screen for and potentially detect depression. To date, studies in this area have focused on developing and validating the terms and vocabulary used by users with depression, or evaluating tweets related to depression by using terms that are synonymous with depression. This approach has not produced reliable findings. In this study, we depart from this approach and instead, base our analysis on research on depressive disorders, which indicates the critical significance of repetitive thoughts and ruminating behavior of people with depression. The current study and findings hold important implications for research on depression, social media, and public health informatics.

1. Introduction

Undiagnosed and untreated depressive disorders have become a serious public health issue and more and more people are being diagnosed with depression each year [1]. Studies are also finding that depression, if left untreated, has serious consequences, such as suicide or violence [1, 2]. Many people suffering from depression experience loss of appetite, binge eating, rumination, lack of sleep or too much sleep, total loss of physical abilities, withdrawal and seclusion and other physiological and psychological symptoms [1]. Depressive disorders can also be accompanied by multiple medical problems, unexplained physical symptoms, chronic pain, all of which could lead to higher and more frequent use of medical services [1].

Depression is highly prevalent among people of all ages, gender and race and further, increasingly there is evidence of a rise in depression among younger population and senior citizens. Over 20 million people in the United States suffer from some form of mood disorder, however, only 50% is being diagnosed and treated properly [3]. Approximately 14.8 million adults (>18yrs) or 6.7% of the US population are being diagnosed with Major depressive disorders (MDD) every year [3]. While the rate of diagnosis is increasing, depression diagnosis is still very difficult as clinical depression can manifest in different ways and can vary from one individual to another [4]. Sometimes it can mimic other diseases and quite often, co-existing conditions can confound an accurate diagnosis of clinical depression [4]. While active research is being undertaken to develop a more accurate test to clearly diagnose depressive disorders, most physicians rely on interview based methods and ruling out other diseases that could cause the same kind of symptoms.

Depression among younger population, which is reaching the rates of adult depression [5,6], is far more troubling as they are typically ill-equipped to detect it and can result in poor self-esteem and blaming of oneself, rather than seeking help [7]. However, this population is far more inclined to express themselves freely and disclose their thoughts and feelings spontaneously on social media micro-blogging sites such as Twitter.

Launched in 2006, Twitter has become an increasingly popular social media platform that allows users to read and publicly express their current thoughts or opinions on any subject in a message of 140 characters or less,
called a tweet. Based on the company’s estimate there are over 271 million monthly active users and about 500 million tweets sent per day. This produces extensive data and a potential to study them for various health related aspects, such as detecting moods from tweets or understanding how people use such mediums for health enhancement and support.

Twitter use has been studied in several health related contexts, especially for public health campaigns [8] and public health surveillance [9]. Velardi et al [9] mined tweets and used a symptom-driven approach for early detection and analysis of epidemics. Studies have also shown that Twitter can be used as an effective source of social support, when dealing with health related issues, such as weight loss: in a study involving active social media users, it was found that people who tweeted about their weight loss attempts on Twitter received more social support than on Facebook, or even in real life [10]. Recent events have also spurred an increasing interest in social media activity of those who engage in mass shootings and other crimes. This has increased the interest in understanding whether these posts/tweets could be used to diagnose or detect mental illness and other depression-related behaviors.

Research on depression-related tweets has focused on investigating whether tweets can be used to diagnose depression. Most of these studies have adopted a simple ‘vocabulary-based approach’ -- specifically, developing and validating the terms and vocabulary used by Twitter users with depression, which include several symptoms of depression as well as terms that are synonymous with depression [e.g. 11,12]. These studies certainly indicate that tweets can be used as a tool to detect depression, however, the use of vocabulary and terms are not always reliable and are quite tedious. As such, the findings from these studies have been rather ad-hoc in nature and not very useful for formulating public health policies. Hence, it is imperative that we draw from the study of depressive disorders to analyze tweets made by Twitter users with depression.

Studies in the area of depression clearly indicate that one of the main symptomatic characteristics of depression is repetitive thoughts and ruminating behavior [13]. Repetitive thoughts are usually about the symptoms or concomitant characteristics of depression, such as those related to sleep (either too much sleep or lack of sleep), pain, suicidal thoughts, violence etc. These types of repetitive thoughts and ruminating behavior can manifest in a person’s tweets and can serve as a better indicator of the presence of depression than the vocabulary used. Hence, in this study we specifically look to validate whether people with depression show such behavior through their tweets. specifically, we investigate whether

1. Twitter users with depression tweet significantly greater number of times regarding symptoms or concomitant characteristics of depression;
2. Twitter users with depression tweet significantly more about symptoms or concomitant characteristics of depression than Twitter users without depression (control group);
3. some of these symptomatic characteristics are more associated with the tweets of people with depression than others;

The study findings will enable analysis of tweets to contribute to diagnosis of depression in a more systematic and reliable way.

2. Research Background

2.1. Repetitive thoughts in Depression

Repetitive thoughts about issues and problems in one’s life and about one’s self is a normal process engaged in by people from time to time [14]. This can be a consequence of specific traumatic events or of one’s experiences more generally. Prolonged recurrent thoughts sometimes can become a constant mental process. These repetitive thoughts can be constructive, in which the individual tries to make sense of these experiences. These repetitive thoughts can also be destructive and lead to anxiety and depression. It is not clear as to whether these repetitive thoughts lead to depression or whether depression causes the individual to have these repetitive thoughts. However, studies have shown a clear positive relationship between repetitive thoughts and depression [15,16,17,18,19,20,21,22]. Earlier studies had concluded that repetitive thoughts, especially, rumination and worry, negatively impacts affective states [23,24] and this has led to several remedies that will stop repetitive thoughts [25]. However, recent studies suggests that repetitive thoughts and rumination are part of depression and instead of trying to forcefully avoid it, it might be better to use it in such a way to reach a positive affective state [26]. Talking to someone or writing it in a journal are some of the ways to turn repetitive rumination to positive mood state (e.g. 27, 16]. Hence, research in this area has concluded that repetitive thoughts can have both constructive and unconstructive outcomes [28]. Repetitive thoughts, usually about symptoms and upsetting events can be a necessary solution for
coming to terms with those symptoms and events [15,16,17,18] and also help in distancing oneself from the event by taking it to an abstract level. If individuals have repetitive thoughts and are not able to reach a level of understanding of the situation (or if they cannot distance themselves from it), then they fall into a spiral and this will lead to a more depressive state [19,20,21,22].

2.2. Worry and Rumination

Worry and rumination are two forms of repetitive thought that often co-exist with depression. Worry is usually about future events and rumination is about past events. They both may have similar outcomes, but rumination seems to be higher among individuals with depression, whereas worrying, to some level, is common in the general population. Rumination is one of the main activities that individuals with depressive disorders engage in. Rumination involves asking questions such as “why am I such a loser”, “why are bad things always happening to me”, etc. Some studies show that people who ruminate have better treatment outcomes than people who have thoughts about dangerous activities [29]. Positive outcomes of rumination are especially high when patients undertake it believing that trying to make sense of bad experiences can get them to recover from depression. However, rumination could also lead to negative thoughts and may exacerbate depression as well as indicate the onset of depression [29].

2.3. Repetitive thoughts about Sleep & Pain

Repetitive thoughts about symptoms of depression are quite common in depression. Sleep disorders seem to dominate the list of symptoms. Similar to many other symptoms in depression, it is not clear whether sleep disorders cause depression or whether depression causes sleep issues. People with depression complain about either lack of sleep or too much sleep [30]. Most of these studies were cross-sectional studies that showed a positive association between sleep and depression. One longitudinal study showed that sleep issues did predict future depression [31], however, other issues such as feelings of worthlessness, mood disturbance, thoughts of death etc. were stronger predictors of future depression in another study [32]. It is, however, a concomitant characteristic of depression.

Chronic pain is also another concomitant symptom of depression [33,34]. Chronic pain may not be associated with any physical injury or may not have a physiological cause. Again, as with other symptoms the causal relationship between pain and depression still remains elusive [35]. Evidence, so far, shows clearly that pain is a co-morbidity of depression [33,34]. While people with depression may not always have pain as a co-morbidity, people with chronic pain almost always have depression as a co-morbidity [34].

2.4. Suicidal thoughts in Depression

Another concomitant characteristic of depression is suicidal thoughts. Suicidal thoughts are dangerous and it can have serious consequences. Depression is one of the major causes of suicide and people with depression often have recurrent suicidal thoughts. However, all people with depression may not have suicidal thoughts and these thoughts can also be part of other mental diseases such as schizophrenia and bipolar disorders. Currently, we have statistics on the number of suicides and the number of attempted suicides (especially, if the suicide attempt resulted in seeking medical help). However, we do not have much data on how many people have suicidal thoughts and what percentage of this result in actual suicide.

2.5. Tweets of people with Depression

Analyzing tweets of a person who has been diagnosed with depression can provide a window to the kind of repetitive thoughts that people with depression usually have. Many of the ruminations of the depressed individual can be captured from their tweets; however, we do not know whether the depressed users tweet about depression-related matters and whether such tweets could be analyzed to understand the tweet-patterns of the typical depressed individual.

Researchers from Brigham Young University [36] found that tweets about suicide mirrored with the actual suicide rates in that region, implying that these tweets should be taken seriously and some type of interventions should be developed in parallel with the suicide hotline, specifically targeting social media users.

Social media sites such as Twitter are primarily used by the younger population, who are also sometimes reluctant or incapable of articulating what they feel with regards to depression and suicidal thoughts in a face-to-face setting. There haven’t been any studies that have looked into Twitter data to understand the kinds of repetitive thoughts depressed individuals have regarding their symptoms (or about suicide).

From the above literature, we can conclude that people with depression will tweet about symptomatic
characteristics such as sleep, pain and suicide, much more than people without depression.

Hence, we hypothesize that:
1) Twitter users with depression will exhibit an overall higher level of ruminating behavior compared to that by Twitter users without depression.
2) Tweeting about symptomatic characteristics of depression (such as sleep and pain) will be positively associated with having been diagnosed with depression.
3) Tweeting suicidal thoughts will be positively associated with having been diagnosed with depression.
4) Twitter users with depression are more likely to tweet repeatedly about symptomatic characteristics such as sleep, pain and suicidal thoughts compared to that by Twitter users without depression.

3. Methods

We collected data directly from Twitter. Our first step was to create a dataset of Twitter users who have at some point or the other tweeted that they have been diagnosed with depression or are taking anti-depressant medications. Drawing on prior research on depression, we developed a set of phrases (Table 1) that would enable us to get the tweets with those phrases or words in it. The depression/study group was created by searching for Twitter users who had tweeted that they have been diagnosed with depression, or suffering from depression or implied as such (e.g. tweeted about anti-depressant drug use) during the week of 04/28/2014 through 05/04/2014. Processing 2 and twitter4j 4.0 were used to aggregate individual tweets via the Twitter API.

The search terms yielded 713 tweets. These tweets were then manually validated by reading each tweet and eliminating any non-relevant tweets as well as duplicate tweets made by the same user. Tweets made by non-individuals such as help groups, or other commercial, non-individual entities, etc. were not included. Re-tweets were also not included in the analysis (re-tweets are tweets originating from somebody else and not from the people with depression). The 713 tweets were made by 334 Twitter users and we manually evaluated their tweets to confirm that they are actually tweeting about being diagnosed with depression rather than casually mentioning about being depressed that day, or talking about someone else’s depression. After validation, we were left with 316 unique Twitter users who had tweeted about being diagnosed with depression during that week.

Table 1: Terms & phrases used for searching for depression diagnosis tweets

<table>
<thead>
<tr>
<th>Diagnosis related terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>I AND diagnosed AND depression</td>
</tr>
<tr>
<td>I am suffering AND depression</td>
</tr>
<tr>
<td>I AND depression meds</td>
</tr>
<tr>
<td>I AND medication AND depression</td>
</tr>
<tr>
<td>I AND have AND depression AND meds</td>
</tr>
<tr>
<td>I AND take AND depression meds</td>
</tr>
<tr>
<td>I AND Prozac AND depressed</td>
</tr>
<tr>
<td>I AND Prozac AND depression</td>
</tr>
<tr>
<td>I AND taking depression meds</td>
</tr>
<tr>
<td>I AND therapy AND depression</td>
</tr>
<tr>
<td>I battling AND depression</td>
</tr>
<tr>
<td>I AND clinical AND depression</td>
</tr>
<tr>
<td>I AND living with depression</td>
</tr>
<tr>
<td>How to tell AND depression</td>
</tr>
</tbody>
</table>

For these 316 Twitter users with depression (study group), we downloaded their entire Twitter timeline with all their tweets (all tweets ever made by these 316 Twitter users) for further analysis. The depression/study group made a total of 575255 tweets. NVivo 10 and NCapture were utilized to extract user timelines via Twitter API. MS Access 2013 was used to store aggregated tweets and Hyperion Interactive Reporting Studio 11 was used to search and analyze the tweets for tweet contents. The tweet contents that we were specifically focused on in this study were tweets about symptomatic characteristics and dangerous thoughts, which are considered to be the two categories that people with depression ruminate about the most [6]. For symptomatic characteristics, we chose two topics – sleep and pain. For dangerous thoughts, we picked suicidal thoughts.

We developed a control group of 316 Twitter users by randomly searching for Twitter users who had tweeted at least once during the same week as the depression/study group. We got a total of 479,840 tweets. Commercial and other non-individual users were not included. Twitter timelines for each user was downloaded and validated by verifying that no user in this group had ever tweeted about being diagnosed with depression. Manual validation was performed to ensure that the same search terms used for the depression group were not present.
The timeline for all users was analyzed by searching for the two categories of rumination using a set of pre-compiled search terms (based on prior research on depression). The search results were further manually validated by reading the tweets and eliminating any false positives or non-relevant tweets in the search results.

4. Results

We used Fisher’s exact test (two-tailed) to compare the categorical variables within each group. The first finding was that Hypothesis H1 was supported - the depression group had much higher number of overall tweets than control group and the difference was statistically significant (p<.01). Hypothesis H2 was also supported. There was a significant relationship between having tweeted about suffering from depression and tweeting about sleep related thoughts (p = 0.01). Twitter users diagnosed with depression had significantly more pain-related tweets (p<0.0001). These results are also consistent with the general population, where chronic pain and depression have been shown to coexist [34].

Hypothesis H3 was also supported. Tweets expressing suicidal thoughts were significantly associated with the depression group. Compared to control group, more people in the depression group tweeted about suicidal thoughts (p<0.0001).

Hypotheses 4 was also supported. We found that there were higher number of tweets from the depression group regarding the symptomatic characteristics of depression such as sleep, pain and suicidal thoughts (see table 3).

Among users tweeting about any of these activities (suicide/sleep/pain), no significant association was found between gender and having tweeted about depression or not.

Table 3 - Fisher’s exact test comparing the number of tweets related to pain, sleep & suicide between groups (Hypothesis H1, H2, H3)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Depression Group</th>
<th>Control Group</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Pain related Tweets</td>
<td>1657</td>
<td>746</td>
<td>0</td>
</tr>
<tr>
<td>No. of sleep related tweets</td>
<td>2423</td>
<td>1443</td>
<td>0</td>
</tr>
<tr>
<td>No. of Suicidal tweets</td>
<td>447</td>
<td>105</td>
<td>0</td>
</tr>
</tbody>
</table>

An independent t-test was performed to compare the average tweets (continuous variable) per Twitter user among the two groups, for each of the three symptoms. There was a significant difference in the average tweets per Twitter user between the depression group and the control group, (Table 4) for each of the ruminating thoughts: suicidal (p=0.001), sleep (p=0.0001) and pain (p=0.0001).

Table 4 - Fisher’s exact test comparing average no. of tweets per user in the depression vs. control group (Hypothesis H4)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Depression Group</th>
<th>Control Group</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicidal Tweets</td>
<td></td>
<td></td>
<td>.001</td>
</tr>
<tr>
<td>Average no. of tweets per user</td>
<td>3.04</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>Std. dev.</td>
<td>3.87</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>147</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Sleep Related Tweets</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average no. of tweets per user</td>
<td>9.14</td>
<td>6.01</td>
<td></td>
</tr>
<tr>
<td>Std. dev.</td>
<td>9.94</td>
<td>5.7</td>
<td></td>
</tr>
</tbody>
</table>
Pain Related Tweets

Average no. of tweets per user 6.63 3.66 .0001
Std. dev. 6.76 3.4
n 250 204

Tweets frequency was further analyzed by dividing the Twitter users based on the number of times they tweeted about each of the symptom: 1 tweet, 2 tweets, 3 tweets, 4 tweets and 5 or more tweets. Chi-square test was performed to compare the categorical data (counts of users in each category) among the two groups (Table 5). There was a significant difference among the two groups on the basis of how many times the users in each group for each of the three symptoms: suicidal tweets (p<0), sleep (p=0.001641) and pain (p=0.00001).

Table 5: Chi square test to compare tweet frequency between the groups (Hypothesis H1, H2, H3)

<table>
<thead>
<tr>
<th></th>
<th>Depression Group</th>
<th>Control Group</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suicidal Tweets</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 tweet</td>
<td>64 [3.16]</td>
<td>55 [6.44]</td>
<td>&lt;0</td>
</tr>
<tr>
<td>2 tweets</td>
<td>35 [0.38]</td>
<td>12 [0.77]</td>
<td></td>
</tr>
<tr>
<td>3 tweets</td>
<td>17 [0.60]</td>
<td>4 [1.22]</td>
<td></td>
</tr>
<tr>
<td>4 tweets</td>
<td>11 [1.77]</td>
<td>0 [3.62]</td>
<td></td>
</tr>
<tr>
<td>5 tweets or more</td>
<td>20 [2.47]</td>
<td>1 [5.05]</td>
<td></td>
</tr>
<tr>
<td><strong>Sleep Related Tweets</strong></td>
<td></td>
<td></td>
<td>0.0016</td>
</tr>
<tr>
<td>1 tweet</td>
<td>31 [0.49]</td>
<td>36 [0.54]</td>
<td></td>
</tr>
<tr>
<td>2 tweets</td>
<td>29 [2.26]</td>
<td>44 [2.50]</td>
<td></td>
</tr>
<tr>
<td>3 tweets</td>
<td>23 [0.19]</td>
<td>25 [0.21]</td>
<td></td>
</tr>
<tr>
<td>4 tweets</td>
<td>19 [1.75]</td>
<td>30 [1.94]</td>
<td></td>
</tr>
<tr>
<td>5 tweets or more</td>
<td>163 [3.56]</td>
<td>105 [3.93]</td>
<td></td>
</tr>
<tr>
<td><strong>Pain Related Tweets</strong></td>
<td></td>
<td></td>
<td>0.00001</td>
</tr>
<tr>
<td>1 tweet</td>
<td>41 [4.36]</td>
<td>62 [5.34]</td>
<td></td>
</tr>
<tr>
<td>2 tweets</td>
<td>27 [2.65]</td>
<td>40 [3.25]</td>
<td></td>
</tr>
<tr>
<td>3 tweets</td>
<td>27 [0.00]</td>
<td>22 [0.00]</td>
<td></td>
</tr>
<tr>
<td>4 tweets</td>
<td>24 [0.00]</td>
<td>20 [0.00]</td>
<td></td>
</tr>
<tr>
<td>5 tweets or more</td>
<td>131 [6.34]</td>
<td>60 [7.77]</td>
<td></td>
</tr>
</tbody>
</table>

5. Limitations

While we ruled out depression among the control group, this does not mean that they do not have depression. It is highly unlikely that they have depression, since they haven’t tweeted about it. However, more rigorous measures should be taken to rule out depression in the control group. In addition, existence of other mental illnesses also needs to be ruled out in the control group, which we did not do in this study, since existence of other mental illnesses were not ruled out in the depression group as well. Future studies need to address other mental illnesses that may co-exist with depression that can also impact symptomatic tweets. However, depression might exist concomitantly with several chronic diseases as well, so including all of them might make it too complicated.

6. Conclusions and Implications

Unlike previous studies we used diagnosis-related tweets to create our sample study group. We also compared the tweets with a control group to further validate that the tweets from depressed users will be significantly associated with tweets regarding symptoms and characteristics of depression.

The main contribution of this study is showing that depressed Twitter users exhibit the same ruminating behavior especially on topics such as sleep, pain and suicidal thoughts, as they do offline. This characteristic can be used to identify and diagnose depression using the tweets in a much more effective and efficient way.

Future studies, from this data, could delve deeper into the conversations of the depressed users to understand more about the emotions and sentiments of these users while they are tweeting about suicidal thoughts or their symptoms. While tweets look like tiny bits of disconnected feeds, it has been widely used for conversations [37]. Several studies report on the patterns of conversation and pattern of emotions in these conversations [38,39,40].

The findings from this study indicate that microblogging sites such as Twitter is used by people with depression to express their thoughts and for ruminating about the symptoms and issues that bother them. Hence, future studies on depression could use microblogging as a way to elicit ruminating thoughts from individuals, esp. younger population.

Future studies on depression using social media could also include other social media sites such as Facebook and Instagram and triangulate the data around an
individual user to understand patterns of ruminating behavior during depression. Such data could provide insights into the activities of people with depression and can also be used in the treatment of depression.

Finally, the findings here have important implications for public health informatics. So far, social media has been used for disseminating public health information, but not much for gathering public health information. The study findings here as well as other studies [11] indicate that such micro-blogging sites may be a good source for gathering data on mental well-being of people. More research is required in this area to understand how interventions can be developed to help people who tweet repeatedly about depression symptoms. For example, something like a hotline number when there are tweets regarding suicidal thoughts could be useful. Google, for example, provides such an intervention if the user searches ‘suicidal thoughts’; such an intervention could be integrated into social media sites. The analysis of tweets cannot be a replacement for diagnosis at a physician’s office, but it can be used as a potential screening tool for depression, or it can be used for specific population groups who are more comfortable disclosing their symptoms in Twitter than in an office setting.

7. References


