Towards Assessing Changes in Degree of Depression through Facebook

H. Andrew Schwartz[†] **Johannes Eichstaedt**[†] **Margaret L. Kern**[†] **Gregory Park**[†]
Maarten Sap[†] **David Stillwell**[‡] **Michal Kosinski**[‡] and **Lyle Ungar**[†]
[†]Psychology and Computer & Information Science, University of Pennsylvania
[‡]Psychometrics Centre, University of Cambridge
`hansens@seas.upenn.edu`

Abstract

Depression is typically diagnosed as being present or absent. However, depression severity is believed to be continuously distributed rather than dichotomous. Severity may vary for a given patient daily and seasonally as a function of many variables ranging from life events to environmental factors. Repeated population-scale assessment of depression through questionnaires is expensive. In this paper we use survey responses and status updates from 28,749 Facebook users to develop a regression model that predicts users' degree of depression based on their Facebook status updates. Our user-level predictive accuracy is modest, significantly outperforming a baseline of average user sentiment. We use our model to estimate user changes in depression across seasons, and find, consistent with literature, users' degree of depression most often increases from summer to winter. We then show the potential to study factors driving individuals' level of depression by looking at its most highly correlated language features.

1 Introduction

Depression, a common mental disorder, greatly contributes to the economic, social, and physical burden of people worldwide. Along with other mental disorders it has been related to early termination of education, unstable marriages, teenage pregnancy, financial problems, role impairment, heart disease, and other negative outcomes (Kessler and Bromet, 2013; Lichtman et al., 2014)

Currently, depression is primarily assessed through surveys. Diagnoses require a medical or psychological evaluation, and are typically classified into discrete categories (absent, mild, moderate, severe). Clinicians rely on retrospective reports by patients to monitor symptoms and treatment. Unobtrusive assessments based on language use in Facebook and social media usage could amend both the self-help resources available to patients as well as repertoire of clinicians with richer information. Such a tool could allow for more frequent and fine grained (i.e., continuously scored) assessment and could provide contextualized information (e.g. specific words and online activities that are contributing to the user's depression score).

Here, we predict and characterize one's *degree of depression* (*DDep*) based on their language use in Facebook. Datasets connecting surveyed depression with language in Facebook are rare at best. To operationalize *DDep*, we use the depression facet scores of the "Big 5" item pool (Goldberg, 1999) from the MyPersonality dataset. This provides a continuous value outcome, for which we fit a regression model based on ngrams, LDA topics, and lexica usage. By predicting continuous values, rather than classes, one can track changes in *DDep* of varying size across time; we find significantly more users' *DDep* increases from summer to winter than vice-versa.

Our primary contribution is the exploration of predicting continuous-valued depression scores from individuals' social media messages. To the best of our knowledge this has not previously been studied, with other social media and depression work focused on discrete classes: present or absent. We compare our predictive model of *DDep* to one derived from a state-of-the-art sentiment lexicon and look at changes across seasons. Finally, we characterize *DDep* by looking at its top ngram and topic correlates.

2 Background

2.1 Depression

Depression is generally characterized by persistent low mood, poor concentration, fatigue, and little interest in normally enjoyable activities. Depression can range from mild to severe, and can occur as an acute episode (major depressive episode), extend chronically over time (major depressive disorder, persistent depressive disorder), reoccur after a period of remission (recurrent depression), or occur at specific periods (seasonal affective disorder, postpartum depression, premenstrual dysphoric disorder). Prevalence rates vary; the World Health Organization estimates that over 350 million people worldwide have a depressive disorder, with many more reporting at least some symptoms (Organization, 2012). In the U.S., in the World Health Mental Survey, over half of the respondents (62%) endorsed at least one diagnostic stem questions for depression, with 19.2% meeting criteria for at least one major depressive episode (Kessler et al., 2010).

Although depression has long been defined as a single disease with a set of diagnostic criteria, it often occurs comorbidly with other psychological and physical disorders. Anxiety, anger, and other psychological disorders often co-occur with depression, and some have suggested that anxiety and depression are different manifestations of the same underlying pathology (Mineka et al., 1998). An expert panel convened by the American Heart Association recently recommended that depression be considered a formal risk factor for heart disease (Lichtman et al., 2014). Depression has been related to a range of physical conditions, including asthma, cancer, cardiovascular disease, diabetes, and chronic pain (Kessler and Bromet, 2013), although the causal direction is confounded; it may be that other factors cause both depression and physical illness (Friedman and Kern, 2014).

As noted previously, assessing degree of depression as a continuous value allows us to look at changes in depression across time. There has been longstanding interest and discussion of seasonal patterns of depression, with observations of seasonal depressive patterns apparent in ancient times, and the first systematic description occurring in 1984 (Westrin and Lam, 2007). Commonly called Seasonal Affective Disorder (SAD), the DSM-V now refers to this pattern as recurrent major depressive disorder with a seasonal pattern. A clinical diagnosis of seasonal depression requires that two major depressive episodes have occurred in the past two years, with the onset and remission showing a regular temporal pattern (predominantly with onset occurring in the fall/winter and full remission in spring/summer).

Patients with depression often have common symptoms of low energy, reduced or intensified psychomotor movements, low concentration, indecisiveness, and thoughts of death, as well as related symptoms such as fatigue, insomnia, and weight gain. A challenge in diagnosis is that it relies on a patient's historical report, and other possible causes such as physical illness must be ruled out. Further, with stigmas against mental illness and feats about seeking treatment, many cases go unrecognized, causing considerable burden on the individual and society as a whole. Prevalence rates vary, but rigorous reviews suggest a prevalence of .4% in the U.S., although estimates have been reported as high as 10% (Blazer et al., 1998; Magnusson and Partonen, 2005).

There are a number of different hypotheses about the pathophysiology of S A D, including circadian, neurotransmitter, and genetic causes (Lam and Levitan, 2000). Reviews suggest that light therapy is an effective and well-tolerated treatment, with effects equal to or larger than antidepressants (Golden et al., 2005; Lam and Levitan, 2000; Thompson, 2001; Westrin and Lam, 2007). Attempts to explain why light therapy is so effective have included shifting photoperiods (light-dark cycles, with less light in the winter), changes in melotonin secretion, and circadian phase shifts (Lam and Levitan, 2000).

One related explanation for the photoperiod effect is latitude, with the prevalence of seasonal depression increasing with growing distance from the equator. Although there has been some support for this hypothesis in the U.S. (Rosen et al., 1990), findings in other countries have been mixed (Mersch et al., 1999). Although latitude may play some role, other factors such as climate, genetic vulnerability, and the sociocultural context may have a stronger impact.

Altogether, inconsistent results suggest that there is considerable variation in the magnitude, causes and manifestations of seasonal depression, much of which is not fully understood, in part due to diagnostic issues (Lam and Levitan, 2000). A

Dislike myself.
Am often down in the dumps.
Have frequent mood swings.
Feel desperate.
Feel comfortable with myself. (-)
Seldom feel blue. (-)
Am very pleased with myself. (-)

Table 1: The seven items of the depression facet from the 100-item International Personality Item Pool (IPIP) proxy to the NEO-PI-R (Goldberg, 1999). (-) indicates a reverse coded item.

weekly or even daily depression assessment tool would allow us to more fully understand the seasonal and other temporal changes in depression.

We use the "depression facet" scores derived from a subset of the "big-5" personality items. Specifically, depression is one of several facets (e.g. anger, depression, anxiety, self-consciousness, impulsiveness, vulnerability) of the neuroticism personality factor. Neuroticism refers to individual differences in the tendency to experience negative, distressing emotions, and behavioral and cognitive styles that result from this (McCrae and John, 1992). It includes traits such as tension, depression, frustration, guilt, and self-consciousness, and is associated with low self-esteem, irrational thoughts and behaviors, ineffective coping styles, and somatic complaints.

Various scales have been developed to measure neuroticism, such as the Eysenck Personality Questionnaire (Eysenck and Eysenck, 1975) and the NEO-PI-R (Costa and McCrae, 1992). Some items on these scales overlap with self-reported items that screen for depression (e.g., personality item: "I am often down in the dumps"; depression screening item: "how often have you been feeling down, depressed, or hopeless?"; see Table 1.), such that the personality items effectively provide a proxy measure of depressive tendencies.

2.2 Related Work

Depression has been linked with many online behaviors. In fact, even Internet usage itself seems to vary as a function of being depressed(Katikalapudi et al., 2012). Other behaviors include social networking (Moreno et al., 2011) and differences in location sharing on Facebook (Park et al., 2013).

Most related to our work, are those using linguistic features to assess various measures of de-

pression. For example, De Choudhury et al. (2013) used online posting behavior, network characteristics, and linguistic features when trying to predict depression rather than find its correlates. They used crowdsourcing to screen Twitter users with the CES-D test (Beekman et al., 1997), while others analyzed one year of Facebook status updates for DSM diagnostic critera of a Major Depressive Episode (Moreno et al., 2011). In addition, Park et al. (2013) predicted results of the Beck Depression Inventory (Beck et al., 1961).

While previous works have made major headway toward automatic depression assessment tools from social media, to the best of our knowledge, none have tried to predict depression as a continuum rather than a discrete, present or absent, attribute. For instance, Neuman et al. (2012) classified blog posts based on whether they contained signs of depression, and De Choudhury et al. (2013) classified which newfound mothers would suffer from postpartum depression.

3 Predicting Degree of Depression

3.1 Method

Dataset. We used a dataset of 28,749 nonclinical users who opted into a Facebook application ("MyPersonality"; Kosinski and Stillwell, 2012) between June 2009 and March 2011, completed a 100-item personality questionnaire (an International Personality Item Pool (IPIP) proxy to the NEO-PI-R (Goldberg, 1999), and shared access to their status updates containing at least 500 words. Users wrote on average of 4,236 words (69,917,624 total word instances), and a subset of 16,507 users provided gender and age, in which 57.0% were female and the mean age was 24.8.

The dataset was divided into *training* and *testing* samples. In particular, the testing sample consisted of a random set of 1000 users who wrote at least 1000 words and completed the personality measure, while the training set contained the 27,749 remaining users.

Degree of depression. We estimated user-level degree of depression (*DDep*) as the average response to seven *depression* facet items, which are nested within the larger Neuroticism item pool. For each item, users indicated how accurately short phrases described themselves (e.g., "often feel blue", "dislike myself"; responses ranged from 1 = very inaccurate to 5 = very accu-
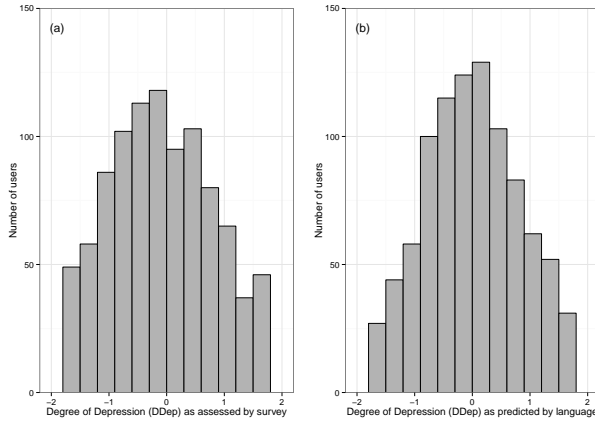
Figure 1: Histograms of (a) survey-assessed and (b) predicted user-level degree of depression $DDep$.
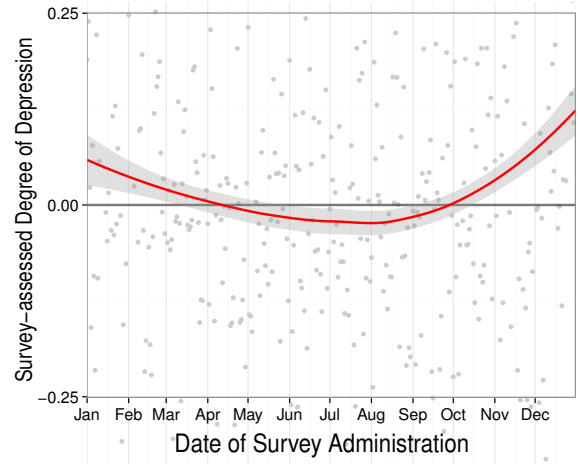

Figure 2: Seasonal trends in degree of depression as assessed by surveys. Red line is a LOESS smoothed trend (+/- 1 SE) over the average of scores from users who completed the survey on that day.

rate). Figure 1a shows the distribution of survey-assessed DDep (standardized). The items can be seen in Table 1.

Figure 2 shows the daily averages of survey-assessed $DDep$, collapsed across years. A LOESS smoother over the daily averages illustrates a seasonal trend, with depression rising over the winter months and dropping during the summer.

Regression modeling. In order to get a continuous value output from our model, we explored regression techniques over our training data. Since this first work exploring regression was concerned primarily with language content, our features for predicting depression were based entirely on language use (other social media activity and friend networks may be considered in future work). These features can be broken into four categories:

ngrams: Ngrams of order to 1 to 3, found via HappierFunTokenizer, and restricted to those used by at least 5% of users (resulting in 10,450 ngrams). The features were encoded as relative frequency of mentioning each ngram (ng):

$$rel\_freq(user, ng) = \frac{freq(user, ng)}{\sum\limits_{ng' \in ngs} freq(user, ng')}$$

topics: 2000 LDA derived Facebook topics.[1] Usage was calculated as the probability of the topic given the user:

$$usage(top|user) = \sum\limits_{ng \in topic} p(top|ng) * rel\_freq(user, ng)$$

lexica: 64 LIWC categories (Pennebaker et al., 2007) as well as the sentiment lexicon from NRC Canada (Mohammad et al., 2013).[2] Usage of a lexicon (lex) was calculated similar to the LDA topics, where w is the weight of the word in the lexicon in the case of sentiment and always 1 in the case of LIWC which has no weights:

$$usage(lex, user) = \sum\limits_{ng \in lex} w(ng, lex) * rel\_freq(user, ng)$$

number of words: Encoded simply as the integer value for that user.

We used penalized linear regression to fit our features to *DDep*. We experimented with a few penalization types over the training set and settled on $L2$ ("ridge regression"), using Principal Components Analysis to first reduce the ngram and topic features to 10 % of their original size. In order to ensure users tested provided an adequate amount of features, we only tested over those with at least 1,000 words. However, we found that including more users in our training set at the expense of words per user increased model accuracy. Thus, we only required our training data users to mention 500 words, essentially allowing more noise in order to increase the number of training examples.

We also experimented with training models on two sets of messages: *all messages* and the subset of messages written in the same three-month season as the survey administration (*season only*

---

[1]downloaded from wwbp.org/data.html

[2]downloaded from www.saifmohammad.com

121

Model	Season test (r)	All test (r)
$Baseline_{sentiment}$	.124	.149
$Season$	.321	.340
All	.351	**.386**

Table 2: Accuracy of various models against test sets containing only messages from the *season* and year in which the user took the survey as well as a test using all of user's messages. Models: $Baseline_{sentiment}$ a model based on a state-of-the-art sentiment lexicon (Mohammad et al., 2013); $Season$: model trained on messages sent only during the same season and year in which each user took the survey; All model trained on all messages of each user.

messages). Because the degree of depression may vary over time, we reasoned that messages written closer to survey administration might better reflect the degree of depression assessed by the survey. When generating predictions on users in the test set, we applied both the all messages model and the season only messages model to features from all messages and then to just the features from the same season as the survey administration.

3.2 Evaluation and Results

We evaluated accuracy using the Pearson correlation coefficient r between our predictions and survey-assessed $DDep$. As a baseline, we built a regression model simply using the NRC sentiment (Mohammad et al., 2013) feature.

Accuracies are shown in Table 2. Accuracy was highest ($r = .386$) when we trained a model over all messages from users in the training set and then applied this model to all messages by users in the test set. Though our model allows for seasonal change in depression, we suspect the test across all messages was more accurate than that of only using the season in which the users depression was assessed due to the larger amount messages and language features provided to the model.

Both models (season-only messages, and all messages) gave significant ($p < 0.05$) improvement over the baseline ($r = .149$) and though these accuracies may look small, it's worth noting that a correlation above $r = 0.3$ is often regarded as a strong link between a behavior and a psychological outcome (Meyer et al., 2001). Still, we fit many behavior variables (i.e., language use features) to an outcome and so we might hope
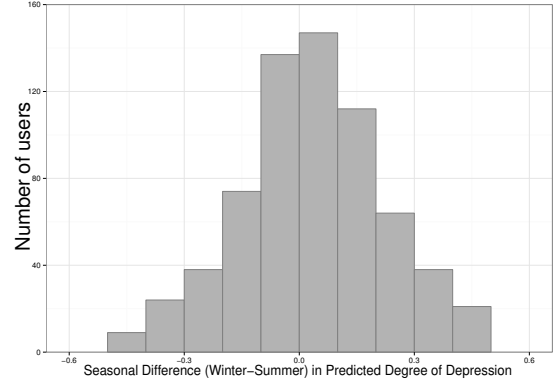

Figure 3: Histogram of differences between winter and summer predictions of user-level $DDep$. Average user-level predicted $DDep$ values were significantly higher in the winter months ($t = 4.63$, $p < .001$).

for higher variance explained. We suspect having more users to train on and taking more features into account could improve results. For example, people who nearly stopped writing for a season would be thrown out of our analyses since it is completely based on language content, even though they are more likely to be depressed (social isolation is a common symptom in depression). Similarly, we do not use demographics in our models, even though women are more likely to become depressed than men.

To assess individual seasonal changes in degree of depression, we predicted summer and winter $DDep$ values for each user with at least 1000 words across both summer-only and winter-only messages, respectively. We then compared the differences across the seasonal predictions; Figure 3 shows the distribution of user-level seasonal differences across 676 users with sufficient language for both seasonal predictions. In line with the trends seen in survey data, average user-level $DDep$ values, as predicted by language, were significantly higher in the winter months ($t = 4.63$, $p < .001$).

4 Differential Language Analysis

Figure 4 shows the 100 ngrams most highly correlated with depression score across the 21,913 Facebook users in our dataset writing at least 1,000 words. Unlike typical word clouds, the clouds represent language that differentiates users scoring high on depression. The size of a word represents its correlation with depression (larger

= stronger), the color its relative frequency (grey = rarely used, blue = moderately used, red = frequently used).

The f-word emerges as both the most correlated feature (as indicated by the size of the word) and is highly frequent (indicated by the red color). Together with words such as 'pissed' and 'bloody', these curse words suggest hostility or aggression. Similarly, words such has 'hate' and 'lonely' suggest negative social relationships.

Perhaps surprisingly, the words 'depression' and 'depressed' emerge as highly correlated features. These face valid features occur infrequently (as indicated by their grey color), yet are strongly associated with depressive tendencies, demonstrating the high statistical power of our approach applied to this large dataset in identifying significant but rarely used language features. The both frequent and highly correlated word 'why' hints at signs of hopelessness and meaninglessness, a core feature of depressive disorders.

As illustrated in Figure 5, extending the words and phrase results, automatically derived topics demonstrate substantial overlap with the major clinical symptoms of major depressive disorder (American Psychiatric Association et al., 2013). Hopelessness and meaninglessness are seemingly expressed by 'hopeless' and 'helpless'. Perhaps the most noticable symptom of depression, depressed mood, is expressed in topics mentioning 'feel', 'crap', 'sad', and 'miserable'.

Depression often affects psychomotor function, either in terms of fatigue and low energy or inversely as insomnia and hyperactivity. Such symptoms are reflected in words such as 'tired', and 'sleep'. Depression is often expressed somatically through bodily symptoms, captured through 'hurt', 'my head' and 'pain'.

One of the most predictive questions on depressive screening questionnaires asks about suicidal thought, which appears with topics related to thoughts of death, with words such as 'kill', 'die', and 'dying'.

Topics also reflected hostility, aggression, and negative relationships with other people. Loneliness has emerged as one of the strongest predictors of physical morbidity and mortality (Hawkley and Cacioppo, 2010), and both 'lonely' and 'alone' appear as some of the most correlated single words. Given such striking descriptive results, future work might try to detect depression associ-

Figure 5: Top ten topics most positively correlated with depression (from $r = .14$ at top to $r = .11$ at bottom). All are significant at a Bonferroni-corrected threshold of $p < 0.001$. Word size corresponds to prevalence within the topics.

ated conditions as well such as insomnia, loneliness, and aggression.

5 Conclusion

Depression can be viewed as a continuous construct that changes over time, rather than simply as being a disease that one has or does not have. We showed that regression models based on Facebook language can be used to predict an individual's degree of depression, as measured by a depression facet survey. In line with survey seasonal trends and the broader literature, we found that language-based predictions of depression were higher in the winter than the summer, suggesting that our
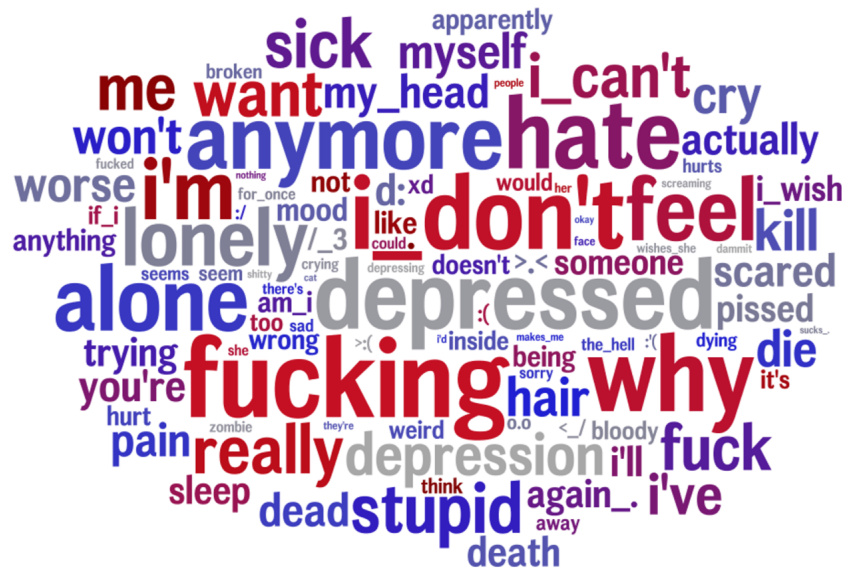
Figure 4: The 100 ngrams most correlated with *DDep* (ranging from $r = .05$ to $r = .10$). All are significant at a Bonferroni-corrected threshold of $p < 0.001$. Ngram size corresponds to correlation strength (larger words are more distinguishing). Color corresponds to relative frequency (red if frequent, blue moderate, grey infrequent).

continuous predictions are capturing small, yet meaningful within-person changes. With further development of regression models, many users write enough on Facebook that we could estimate changes in their level of depression on a monthly or even weekly basis. Such estimates, correlated with word use over time offers potential both for research at the group-level ("What are the social and environmental determinants of depression?", "How well are talk or medication-based interventions working?") as well as, eventually, for medical and therapeutic application at the individual level ("How well am I doing and what depression-relevant thoughts or behaviors have I disclosed in the past week?").

References

APA American Psychiatric Association, American Psychiatric Association, et al. 2013. Diagnostic and statistical manual of mental disorders.

Aaron T Beck, Calvin H Ward, Mock Mendelson, Jeremiah Mock, and JK Erbaugh. 1961. An inventory for measuring depression. *Archives of general psychiatry*, 4(6):561.

Aartjan TF Beekman, DJH Deeg, J Van Limbeek, AW Braam, MZ De Vries, W Van Tilburg, et al. 1997. Criterion validity of the Center for Epidemiologic Studies Depression scale (CES-D): results from a community-based sample of older subjects in The Netherlands. *Psychological medicine*, 27(1):231–236.

Dan G Blazer, Ronald C Kessler, and Marvin S Swartz. 1998. Epidemiology of recurrent major and minor depression with a seasonal pattern. The National Comorbidity Survey. *The British Journal of Psychiatry*, 172(2):164–167.

Paul T Costa and Robert R McCrae. 1992. Professional manual: revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI). *Odessa, FL: Psychological Assessment Resources*.

Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013a. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the 2013 ACM annual conference on Human factors in computing systems*, pages 3267–3276. ACM.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013b. Predicting depression via social media. In *AAAI Conference on Weblogs and Social Media*.

Hans Jurgen Eysenck and Sybil Bianca Giuletta Eysenck. 1975. *Manual of the Eysenck Personality Questionnaire (junior and adult)*. Hodder and Stoughton.

Howard S Friedman and Margaret L Kern. 2014. Personality, Well-Being, and Health*. *Psychology*, 65(1):719.

Lewis R Goldberg. 1999. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, 7:7–28.

Robert N Golden, Bradley N Gaynes, R David Ekstrom, Robert M Hamer, Frederick M Jacobsen, Trisha Suppes, Katherine L Wisner, and Charles B Nemeroff. 2005. The efficacy of light therapy in the treatment of mood disorders: a review and meta-analysis of the evidence. *American Journal of Psychiatry*, 162(4):656–662.

Louise C Hawkley and John T Cacioppo. 2010. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine*, 40(2):218–227.

R Katikalapudi, Sriram Chellappan, Frances Montgomery, Donald Wunsch, and Karl Lutzen. 2012. Associating Internet usage with depressive behavior among college students. *Technology and Society Magazine, IEEE*, 31(4):73–80.

Ronald C. Kessler and Evelyn J. Bromet. 2013. The Epidemiology of Depression Across Cultures. *Annual Review of Public Health*, 34(1):119–138, Mar.

Ronald C Kessler, Howard G Birnbaum, Victoria Shahly, Evelyn Bromet, Irving Hwang, Katie A McLaughlin, Nancy Sampson, Laura Helena Andrade, Giovanni de Girolamo, Koen Demyttenaere, et al. 2010. Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: results from the WHO World Mental Health Survey Initiative. *Depression and anxiety*, 27(4):351–364.

M. Kosinski and D.J. Stillwell. 2012. myPersonality Project. *http://www.mypersonality.org/wiki/*.

Raymond W Lam and Robert D Levitan. 2000. Pathophysiology of seasonal affective disorder: a review. *Journal of Psychiatry and Neuroscience*, 25(5):469.

Judith H Lichtman, Erika S Froelicher, James A Blumenthal, Robert M Carney, Lynn V Doering, Nancy Frasure-Smith, Kenneth E Freedland, Allan S Jaffe, Erica C Leifheit-Limson, David S Sheps, et al. 2014. Depression as a Risk Factor for Poor Prognosis Among Patients With Acute Coronary Syndrome: Systematic Review and Recommendations A Scientific Statement From the American Heart Association. *Circulation*.

Andres Magnusson and Timo Partonen. 2005. Focus Points. *CNS Spectr*, 10(8):625–634.

Robert R McCrae and Oliver P John. 1992. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.

Peter Paul A Mersch, Hermine M Middendorp, Antoinette L Bouhuys, Domien GM Beersma, and Rutger H van den Hoofdakker. 1999. Seasonal affective disorder and latitude: a review of the literature. *Journal of affective disorders*, 53(1):35–48.

Gregory J Meyer, Stephen E Finn, Lorraine D Eyde, Gary G Kay, Kevin L Moreland, Robert R Dies, Elena J Eisman, Tom W Kubiszyn, and Geoffrey M Reed. 2001. Psychological testing and psychological assessment: A review of evidence and issues. *American psychologist*, 56(2):128–165.

S Mineka, D Watson, and LA Clark. 1998. Comorbidity of anxiety and unipolar mood disorders. *Annual review of psychology*, 49:377.

Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu. 2013. NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, Georgia, USA, June.

Megan A Moreno, Lauren A Jelenchick, Katie G Egan, Elizabeth Cox, Henry Young, Kerry E Gannon, and Tara Becker. 2011. Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and anxiety*, 28(6):447–455.

Yair Neuman, Yohai Cohen, Dan Assaf, and Gabbi Kedma. 2012. Proactive screening for depression through metaphorical and automatic text analysis. *Artificial intelligence in medicine*, 56(1):19–25.

World Health Organization. 2012. *Depression fact sheet.* http://www.who.int/mediacentre/factsheets/fs369/en/.

Sungkyu Park, Sang Won Lee, Jinah Kwak, Meeyoung Cha, and Bumseok Jeong. 2013. Activities on Facebook Reveal the Depressive State of Users. *Journal of medical Internet research*, 15(10).

James W. Pennebaker, C.K. Chung, M. Ireland, A. Gonzales, and R.J. Booth. 2007. The development and psychometric properties of LIWC2007. *Austin, TX, LIWC. Net*.

Leora N Rosen, Steven D Targum, Michael Terman, Michael J Bryant, Howard Hoffman, Siegfried F Kasper, Joelle R Hamovit, John P Docherty, Betty Welch, and Norman E Rosenthal. 1990. Prevalence of seasonal affective disorder at four latitudes. *Psychiatry research*, 31(2):131–144.

C Thompson. 2001. Evidence-based treatment. *Seasonal affective disorder: practice and research*, pages 151–158.

Asa Westrin and Raymond W Lam. 2007. Seasonal affective disorder: a clinical update. *Annals of Clinical Psychiatry*, 19(4):239–246.