Text-Based Detection and Understanding of Changes in Mental Health

Yaoyiran Li, Rada Mihalcea and Steven R. Wilson

University of Michigan, Ann Arbor, MI, USA
{yaoyiran,mihalcea,steverw}@umich.edu

Abstract. Previous work has investigated the identification of mental health issues in social media users, yet the way that users' mental states and related behavior change over time remains relatively understudied. This paper focuses on online mental health communities and studies how users' contributions to these communities change over one year. We define a metric called the Mental Health Contribution Index (MHCI), which we use to measure the degree to which users' contributions to mental health topics change over a one-year period. In this work, we study the relationship between MHCI scores and the online expression of mental health symptoms by extracting relevant linguistic features from user-generated content and conducting statistical analyses. Additionally, we build a classifier to predict whether or not a user's contributions to mental health subreddits will increase or decrease. Finally, we employ propensity score matching to identify factors that correlate with an increase or a decrease in mental health forum contributions. Our work provides some of the first insights into detecting and understanding social media users' changes in mental health states over time.

Keywords: Natural Language Processing · Mental Health · Social Media

1 Introduction and Related Work

Mental health issues pose a major challenge for modern society [6, 23, 5] and early detection and intervention are fundamental to preventing the progression of mental illnesses [32, 38, 55, 27, 26, 1, 49]. To aid the work already being done by medical practitioners, social media data are increasingly being used to analyze and predict mental illnesses [31, 42, 22, 14, 39, 43, 3, 15, 28]. As much of the data shared by users on social media is unstructured text, language attributes have proven to be effective features for the analysis and prediction of mental health states of social media users [37, 13, 2, 35, 11, 40, 51]. In this paper, we focus on the discussion website, Reddit[1], where user-generated content is organized into topical, user-created boards called "subreddits". Previous work has studied mental health-related (MH) subreddits in terms of self-disclosure, social support and anonymity [22, 43]. Prior work has also focused on both the detection and prediction of mental health problems using social media data [51, 17, 19, 40]. Our work, however, focuses on the changes of individuals' mental states with the aim to address the following two tasks: (1) to build a classification system that is

[1] http://www.reddit.com

able to predict changes in the level of user interaction with online MH communities, and (2) to discover the underlying factors that correlate with those changes. While we would ideally like to track individuals' true mental states over time, this information is difficult to obtain directly. Therefore, in this work, we focus directly on the change in users' contributions to online MH communities as measured by the number of posts that these users contribute to MH subreddits. We show that this measure is, in fact, significantly related to different linguistic expressions of MH symptoms, and should serve as a beneficial index to consider in the study of MH discussions in online communities.

We collect users from MH subreddits and define a Mental Health Contribution Index (MHCI) to separate users into three categories: those who make more, less, or about the same number of contributions to MH subreddits in the second half of a given year (between 2017-03-01 and 2017-09-01) compared with those in the first half of that year (between 2016-09-01 and 2017-03-01). Here the term *contribution* means the number of posts contributed to target subreddits. When deciding the length of the time period, we follow the approach taken in previous work [21, 20, 18, 19, 35]. For the users in our study, we seek to address three research questions:

- **RQ1**. How do users, grouped by their MHCI scores, express different symptoms of MH problems throughout the year in general?
- **RQ2**. Can we build a classifier to predict if a user's contributions to MH subreddits will increase or decrease during the second half of the year?
- **RQ3**. What factors from the first six months correlate with either an increase or a decrease in MH contributions in the second half of the year?

2 Data Collection and Annotation

We crawl our data through the Python Reddit API PRAW[2] and annotate the data produced by a set of redditors (users of the Reddit site). The process consists of the following steps: (1) Identify MH subreddits and redditors; (2) Define a Mental Health Contribution Index (MHCI) based on posting and commenting actions; and (3) Filter target redditors according to their MHCI scores.

We start with fifteen well-studied MH subreddits reported in [21], from which we crawl posts and information about the users who wrote them. The subreddits used in our study are: r/depression, r/SuicideWatch, r/socialanxiety, r/mentalhealth, r/BPD, r/ptsd, r/bipolarreddit, r/rapecounseling, r/StopSelfHarm, r/survivorofabuse, r/EatingDisorders, r/hardshipmates, r/panicparty, r/psychoticreddit and r/traumatoolbox. As done in previous work [22, 43, 21], we do *not* assume that every user found on these subreddits is an active mental health patient. We conducted the crawl on 2017-09-01, collecting data from the past year, so that time span of our data is from 2016-09-01 to 2017-09-01. We crawl all the authors of posts found on these subreddits in the given time span. After removing moderators and suspended users, we are left with a final set of 53,416 unique users.

Next, we define the Mental Health Contribution Index (MHCI) which measures the change in contributions to MH subreddits against Non-MH subreddits for a given redditor between two 6-month time periods. The *contribution* to MH subreddits is defined

[2] https://praw.readthedocs.io/en/latest/

as the total number of posts contributed in these subreddits, which is regarded as a reflection of the attention paid to MH topics, or the extent to which the user may resort to help and exchange ideas with others. First, we define t_1 as the time period between 2016-09-01 and 2017-03-01; t_2 as the time period between 2017-03-01 and 2017-09-01; m_i^r as the number of posts contributed to MH subreddits in $t_i, i = 1, 2$ for a redditor r; and n_i^r as the number of posts contributed to NonMH subreddits in $t_i, i = 1, 2$ for a redditor r. We then define the MHCI score for a redditor r as follows:

$$MHCI(r) = \alpha \frac{m_2^r + 1}{m_1^r + 1} + \beta \frac{(m_2^r + 1)(n_1^r + 1)}{(m_1^r + 1)(n_2^r + 1)} \tag{1}$$

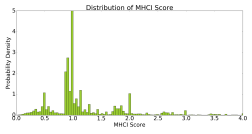

Fig. 1: Distribution of MHCI score in 53,416 redditors.

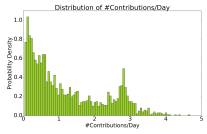

Fig. 2: Distribution of number of contributions per day over 1,767 redditors.

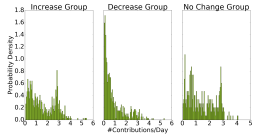

Fig. 3: Distribution of number of contributions per day over three types of users respectively.

The MHCI score is a weighted sum of the ratio between absolute MH contributions and the ratio between relative MH contributions in t_1 and t_2. We also introduce a score, $MHCI'(r)$, that is the same as $MHCI(r)$, but it includes the number of total posts *plus comments* (rather than only posts). The distribution of the MHCI score for all 53,416 redditors is shown in Figure 1. The criteria we adopt to filter users with increased, decreased and no change in contributions to MH subreddits are as follows:

$$increase\ group: MHCI(r) > 2.5, MHCI^{'}(r) > 5$$

$$decrease\ group: MHCI(r) < 0.4, MHCI^{'}(r) < 0.2$$

$$no\ change\ group: 0.9 < MHCI(r) < 1.1, 0.75 < MHCI^{'}(r) < 1.25$$

We exclude all samples in "gray zones" (e.g. $0.4 < MHCI(r) < 0.9$) to reduce the risk of misclassification [16, 52, 19, 47, 33]. For filtered users, annotators are asked to manually look into all their posts and comments. Only users with self-reported diagnoses of MH problems are kept [13, 12]. Other users who provide help, seek support for helping other people instead of themselves, distribute questionnaires for study and do not reveal their own MH states are ruled out. Based on the criteria above, we derive 641 redditors in the *increase* group, 758 redditors in the *decrease* group, and 368 redditors in the *no change* group; these 1,767 redditors form our final dataset which contains 113,630 posts and 692,203 comments where 13,162 posts and 66,961 comments are from MH subreddits and the rest are from Non-MH subreddits. Figure 2 and 3 show the activity of our redittors. We identify 703 subreddits that were contributed to by at least 12 of our 1,767 redditors in the first half-year, which we will use to answer RQ3. We anonymized our data and all individuals who participated in our project, including annotators, strictly follow the ethical guidelines as suggested in previous work [4, 30].

3 RQ1: Changes of MH Symptoms

3.1 Method

In this section, we measure the differences in language use between the *increase* group and the *decrease* group over the time span of one year. We calculate the frequencies of using words belonging to different semantic categories in the Linguistic Inquiry and Word Count (LIWC) [45] lexicon during the year and conduct Welch's t-test to measure if there is a significant change in the usage of words from these categories. In our calculation, a positive t-stat represents the increased use of a semantic category, and a negative t-stat represents decreased use. We point out known MH symptoms as linguistic features and analyze the way that those symptoms change over the year.

Table 1: Welch's t-test results between contents of two six-month periods for *increase* group users (with high MHCI scores). The significance levels are $\alpha = 0.05/M$(*), $0.01/M$(**), $0.001/M$(***). $M = 60$ is adopted for Bonferroni correction.

Category	Time Period 1	Time Period 2	t-stat	p
negate	0.0223	0.0242	21.569	***
death	0.0022	0.0024	6.420	***
health	0.0079	0.0100	39.353	***
affect	0.0633	0.0662	19.994	***
leisure	0.0128	0.0116	-18.577	***
interrogative	0.0174	0.0176	3.001	-
adverb	0.0618	0.0640	20.505	***
conjunction	0.0703	0.0721	11.791	***
pronoun	0.1685	0.1779	42.618	***
verb	0.1827	0.1901	32.261	***
1st person singular	0.0595	0.0651	39.597	***
1st person plural	0.0051	0.0047	-10.564	***
2nd person	0.0202	0.0217	17.077	***
3rd person singular	0.0131	0.0126	-7.080	***
positive emotion	0.0368	0.0365	-2.795	-
negative emotion	0.0258	0.0287	30.963	***
sad	0.0046	0.0058	28.839	***
anxiety	0.0037	0.0046	23.795	***

Table 2: Welch's t-test results between contents of two six-month periods for *decrease* group users (with low MHCI scores). The significance levels are $\alpha = 0.05/M$(*), $0.01/M$(**), $0.001/M$(***). $M = 60$ is adopted for Bonferroni correction.

Category	Time Period 1	Time Period 2	t-stat	p
negate	0.0248	0.0231	-16.581	***
death	0.0027	0.0023	-9.489	***
health	0.0110	0.0081	-44.124	***
affect	0.0691	0.0623	-28.521	***
leisure	0.0103	0.0120	22.981	***
interrogative	0.0179	0.0177	-1.748	-
adverb	0.0658	0.0619	-23.448	***
conjunction	0.0735	0.0711	-13.696	***
pronoun	0.1873	0.1687	-71.382	***
verb	0.1944	0.1833	-41.384	***
1st person singular	0.0739	0.0573	-99.840	***
1st person plural	0.0047	0.0055	20.309	***
2nd person	0.0217	0.0197	-20.652	***
3rd person singular	0.0120	0.0123	3.634	**
positive emotion	0.0364	0.0366	1.747	-
negative emotion	0.0317	0.0269	-41.996	***
sad	0.0077	0.0049	-54.379	***
anxiety	0.0051	0.0039	-28.349	***

3.2 Results and Analysis

Previous work suggests that some certain linguistic features extracted from social media users can be viewed as symptoms of mental disorders [19, 18, 24] and this fact is also supported by research in psychiatry [26, 1, 49]. Compared with control group, individuals with mental issues show more frequent use of categories such as 'negative emotion', '1st person singular pronouns', 'negate' and less frequent use of categories such as 'positive emotion', '3rd person singular pronouns' and '1st person plural pronouns'. Our results in Table 1 and 2 show that, in general, high MHCI users express increased MH symptoms and low MHCI users express decreased MH symptoms in second half of the year:

Emotional Symptoms. High MHCI users show increased use of 'negative emotion', 'sad' and 'anxiety' categories and decreased use of 'leisure' and 'positive emotion' categories. However, low MHCI users exhibit the opposite trend.

Linguistic Symptoms. High MHCI users show an increased use of verbs, which is considered to positively correlate with sensitivity [29]. Increased use of '1st person singular' and decreased use of '1st person plural' and '3rd person singular pronouns' indicates that high MHCI users become more socially isolated and self-attentional in the second half of the year [8]. The opposite trend is also true for the low MHCI users.

Subjectivity Symptoms. From the 'negate' category, we observe that high MHCI users tend to express more negative opinions in the second half of the year, while low MHCI users express fewer negative opinions during the same time period.

4 RQ2: A Classification Task

4.1 Method

In this section, we propose an initial framework for a classification task between high and low MHCI users based on only the texts that these users have written. The training and test data consists of all posts and comments contributed by the 641 users from the *increase* group and 758 users from the *decrease* group during t_1 and t_2. For each user, we extract features v_1 and v_2 from raw texts contributed during t_1 and t_2 respectively. Then, $v_2 - v_1$, which is expected to capture information on changes of language style for a user between the two halves of the year, serves as an input to a classifier. The output is binary variable representing whether or not a user's MH contributions increase or decrease in t_2 compared with t_1.

There are four types of features adopted in our work: Average Word2Vec, Average GloVe, Doc2Vec and LIWC. Although Word2Vec and GloVe provide pre-trained word embeddings, we train Word2Vec, GloVe and Doc2Vec embeddings from scratch on our data because of the uniqueness of our corpus. Here we introduce our features briefly:

Average Word2Vec. We use the Continuous Bag-of-Words Model (CBOW) to train 300-dimensional word embeddings which uses a neural network [36] with negative sampling [34]. We set the window size to 5 and the initial learning rate to 0.025. Then, for each user, we average together all word embeddings to produce a single feature vector for each of t_1 and t_2 to produce v_1 and v_2.

Average GloVe. We also train 300-dimensional GloVe word embeddings which are based on global word-word co-occurrence statistics [46]. The window size is set as 15 and the initial learning rate is set as 0.05. For each user, v_1 and v_2 are derived by averaging together all word embeddings in t_1 and t_2 respectively.

Doc2Vec. We implement Distributed Memory Model with Paragraph Vectors (PV-DM) [34] for training document embeddings with a window size of 10 and an initial learning rate of 0.025. For each user, the model will directly output 300-dimensional feature vectors, v_1 and v_2, that summarize all the user's posts and comments in t_1 and t_2, respectively.

LIWC. We use all semantic categories in the LIWC lexicon [45]. For each user, we count frequencies of use of those categories in t_1 and t_2 and derive v_1 and v_2.

We adopt Logistic Regression (LR), Support Vector Machine (SVM) and a custom Neural Network (NN) with two hidden layers of size 100 and 20 as our classifiers, and they are implemented through SciKit-learn package [44].

4.2 Results and Analysis

We evaluate our models with 10-fold cross validation, and report the average accuracy, precision, recall, f1-score and their 95% confidence interval of the score estimate (i.e. 2 times standard deviation) in Table 3. When using a single-source feature, we find that averaged word embeddings can capture the changes of mental states to some extent, but they are inferior to using LIWC categories. Doc2Vec, however, beats LIWC in accuracy, recall and f1-score. When we combine Doc2Vec with LIWC, however, the performance increases even further for all metrics when using the neural network classifier.

Feature and Classifier	Accuracy	Precision	Recall	F1-Score
Word2Vec+LR	0.7713 (+/- 0.0662)	0.7542 (+/- 0.0758)	0.7442 (+/- 0.0986)	0.7485 (+/- 0.0760)
Word2Vec+SVM	0.7820 (+/- 0.0688)	0.7521 (+/- 0.0760)	0.7831 (+/- 0.0911)	0.7668 (+/- 0.0747)
Word2Vec+NN	0.7649 (+/- 0.0842)	0.7454 (+/- 0.0946)	0.7457 (+/- 0.1500)	0.7395 (+/- 0.1095)
GloVe+LR	0.7756 (+/- 0.0672)	0.7638 (+/- 0.1046)	0.7442 (+/- 0.0584)	0.7530 (+/- 0.0655)
GloVe+SVM	0.7692 (+/- 0.0586)	0.7485 (+/- 0.0822)	0.7505 (+/- 0.0658)	0.7489 (+/- 0.0608)
GloVe+NN	0.7527 (+/- 0.0747)	0.7436 (+/- 0.0907)	0.7209 (+/- 0.1167)	0.7269 (+/- 0.0877)
LIWC+LR	0.8142 (+/- 0.0412)	0.7941 (+/- 0.0605)	0.8066 (+/- 0.1327)	0.7980 (+/- 0.0574)
LIWC+SVM	0.8185 (+/- 0.0491)	0.8052 (+/- 0.0584)	0.7989 (+/- 0.1219)	0.8004 (+/- 0.0631)
LIWC+NN	0.8235 (+/- 0.0419)	**0.8170 (+/- 0.0639)**	0.8238 (+/- 0.1066)	0.8143 (+/- 0.0774)
Doc2Vec+LR	0.8234 (+/- 0.0668)	0.8107 (+/- 0.0610)	0.8019 (+/- 0.1187)	0.8054 (+/- 0.0803)
Doc2Vec+SVM	0.8113 (+/- 0.0350)	0.7955 (+/- 0.0399)	0.7925 (+/- 0.0769)	0.7934 (+/- 0.0446)
Doc2Vec+NN	**0.8241 (+/- 0.0377)**	0.8088 (+/- 0.0563)	**0.8268 (+/- 0.0691)**	**0.8181 (+/- 0.0342)**
Doc2Vec+LIWC+LR	0.8392 (+/- 0.0610)	0.8284 (+/- 0.0733)	0.8207 (+/- 0.1025)	0.8235 (+/- 0.0699)
Doc2Vec+LIWC+SVM	0.8306 (+/- 0.0666)	0.8104 (+/- 0.0689)	0.8238 (+/- 0.1060)	0.8163 (+/- 0.0758)
Doc2Vec+LIWC+NN	**0.8614 (+/- 0.0535)**	**0.8587 (+/- 0.0544)**	**0.8519 (+/- 0.0763)**	**0.8558 (+/- 0.0333)**

Table 3: Classification results with 10-fold cross-validation. We report here the average accuracy, precision, recall, f1-score and their 95% confidence interval of the score estimate (i.e. 2 times standard deviation).

5 RQ3: Factors Correlate with Changes in MH Contributions

5.1 Method

In this section, we borrow the propensity score matching (PSM) method from the study of causal inference to find if contributions to certain subreddits in t_1 correlate with increased (high MHCI) or decreased (low MHCI) contributions to MH subreddits in t_2. PSM aims to isolate the effect of other observable covariates known as confounding variables [9, 7, 10] and thus PSM may be preferable to traditional statistical approaches like regression models which can give rise to confounding effects [21, 20]. In our case, for each subreddit, we form pairs of users having similar *probabilities of contributing* to the subreddit, but *only one user in each pair actually contributed*. Then we conduct Welch's t-test to measure if the subreddit may have an effect on those users. Our dataset includes $n = 1,767$ users and $m = 703$ treatments (subreddits). Confounding variables are unigram features for every user, X_i, $i = 1$ to n. $T_{j,i}$, $j = 1$ to m (binary), represents if user i received treatment j. A Propensity Score is defined as the probability of receiving treatment j given a confounding variable, $P_j(X_i) = Prob(T_{j,i} = 1|X_i), \forall i$,

which is fitted with logistic regression model [44]. User labels are denoted as Y_i, $i = 1$ to n. Then our algorithm is summarized in Algorithm 1:

Algorithm 1 Measuring the effects of treatments on MHCI

Require: $X_i, Y_i, T_{j,i}, i = 1$ to n, $j = 1$ to m
  for $j = 1$ **to** m **do**
    step 1: Split data into a training set and a test set.
    step 2: Fit $model_j$ to the training data.
    step 3: Form treatment group and control group in test set based on Propensity Score Matching.
    step 4: Conduct Welch's t-test on treatment and control groups.
  end for
  return t-stats for all m treatments

5.2 Results and Analysis

Table 4: Top 15 treatments that correlate with an increase in MH contributions.

Treatment	t-stat
r/WikiLeaks	3.464
r/vancouver	3.464
r/trypophobia	2.752
r/Marijuana	2.738
r/Ask_Politics	2.449
r/cordcutters	2.449
r/piercing	2.291
r/cars	2.254
r/announcements	2.190
r/MeanJokes	2.038
r/AskUK	2.070
r/Bandnames	2.000
r/solotravel	2.000
r/whatisthisthing	1.981
r/Bitcoin	1.951

Table 5: Top 15 treatments that correlate with a decrease in MH contributions.

Treatment	t-stat
r/depression	14.191
r/BipolarReddit	5.740
r/SuicideWatch	5.554
r/StopGaming	4.472
r/bipolar	4.354
r/pics	4.157
r/mentalhealth	4.057
r/pornfree	3.464
r/rapecounseling	3.314
r/baseball	3.162
r/socialanxiety	3.004
r/comics	2.758
r/LongDistance	2.738
r/Rateme	2.662
r/BPD	2.660

Table 4 and Table 5 show the top 15 treatments that correlate with an increase in MH contributions and ones that correlate with a decrease in MH contributions, respectively. Our analysis will pay more attention to treatments that correlate with decreased MH contributions since their t-stats are relatively higher. We then discuss the results in the following aspects:

Support Communities. Support communities are shown to correlate with decreased MH contributions in t_2 which are shown to be correlated with reduced MH symptoms in RQ1. MH support subreddits include 'r/depression', 'r/BipolarReddit', 'r/SuicideWatch', 'r/bipolar', 'r/mentalhealth', 'r/socialanxiety' and 'r/BPD' (Borderline Personality Disorder). Recent work shows the reciprocity between social media users who disclose their own MH issues and their online audience, which is consistent with our findings

[25, 53]. Other support communities include 'r/rapecounseling' (help with sexualized violence), 'r/StopGaming' (help with video game addiction) and 'r/pornfree' (help with addiction to porn).

Interesting Pictures, Comics and Memes. Some subreddits focus on sharing images, captioned photos etc. that are intended to be funny. This category includes 'r/pics' and 'r/comics' and both correlate with decreased MH contributions in t_2.

Story Sharing and Friend Making. These subreddits correlate with decreased MH contributions. 'r/LongDistance' is a subreddit to share stories about long-distance relationships and 'r/Rateme' for users to rate everyone else. As a support, psychological research shows that social interaction is helpful to reduce stress [41].

Politics. There are two subreddits related to politics in Table 4 and 5. They are 'r/WikiLeaks' and 'r/Ask_Politics', and both correlate with increased MH contributions in t_2.

Other Subreddits. 'r/baseball' correlates with reduced MH contributions in t_2. 'r/Marijuana', 'r/trypophobia' (a community for those with a common fear of irregular clusters of holes or bumps found in the world) and 'r/piercing' (for discussion of various body piercings and jewelry) correlate with increased MH contributions. Some support for these findings can be found in psychological research on sports [48], drug use [50] and body piercing [54].

Our findings may provide some insights on what factors may influence individual's mental health. Those factors mainly include discussions in certain subreddits like 'r/Rateme'. Currently, we cannot make any claim on if engagement in a certain subeddit has a causal relationship with mental health, but we still deem it valuable for mental health researchers to investigate promising correlations such as those that we have discovered in this work.

6 Conclusion

This paper provides some insights into detecting and understanding changes in contributions to online mental health communities over time. We propose the MHCI index and filter users whose contributions to MH subreddits increase, decrease and stay about the same over two consecutive six-month periods. Our findings show that increased MH contributions correlate with increased MH linguistic symptoms while decreased MH contributions generally show the opposite trend. Further, we propose a framework for building classifiers to distinguish between high and low MHCI redditors and demonstrate the effectiveness of word embeddings and document embeddings in this task. Our work also reveals the underlying correlation between users' engagement in discussions in different subreddits and changes in those users' MH contributions over time.

7 Acknowledgement

We thank all anonymous reviewers for their constructive suggestions on our work. We also thank Dr. Márcio Duarte Albasini Mourão for helpful discussions with us on RQ1. This work is partly supported by the Michigan Institute for Data Science, by the National Science Foundation under grant #1344257 and by the John Templeton Foundation under grant #48503.

References

1. Abdel-Khalek, A.M.: Can somatic symptoms predict depression? Social Behavior and Personality: an international journal **32(7): 657-666** (2004)
2. Amir, S., Coppersmith, G., Carvalho, P., et al.: Quantifying mental health from social media with neural user embeddings. Proceedings of Machine Learning for Healthcare 2017 (2017)
3. Benton, A., Mitchell, M., Harman, C.: Multitask learning for mental health conditions with limited social media data. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (2017)
4. Benton, A., Coppersmith, G., Dredze, M.: Ethical research protocols for social media health research. Proceedings of the First Workshop on Ethics in Natural Language Processing (2017)
5. Bijl, R., De Graaf, R., E., H.: The prevalence of treated and untreated mental disorders in five countries. Health Aff (Millwood) **22**, 122–133 (2003)
6. Bloom, D., Cafiero, E., Jan-Llopis, E., Abrahams-Gessel, S., Bloom, L., Fathima, S., Feigl, A., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A., Weinstein, C.: The global economic burden of non-communicable diseases. In: Geneva: World Economic Forum. Geneva, 2011 (2011)
7. Blundell, R., et al.: Alternative approaches to evaluation in empirical microeconomics. Portuguese Economic Journal **1**, 91–115 (2002)
8. Boals, A., Klein, K.: Word use in emotional narratives about failed romantic relationships and subsequent mental health. Journal of Language and Social Psychology **24**, 252–268 (2005)
9. Caliendo, M., Kopeinig, S.: Some practical guidance for the implementation of propensity score matching. Journal of Economic Surveys **22(1)**, 31–72 (2008)
10. Caliendo, M., et al.: The microeconometric estimation of treatment effects-an overview. Working Paper, J.W.Goethe University of Frankfurt (2005)
11. Chancellor, S., Lin, Z., Goodman, E.L., Zerwas, S., De Choudhury, M.: Quantifying and predicting mental illness severity in online pro-eating disorder communities. Proceedings of The 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing (2016)
12. Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K.: From adhd to sad: Analyzing the language of mental health on twitter through self-reported diagnoses. Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (2015)
13. Coppersmith, G., Harman, C., Dredze, M.: Measuring post traumatic stress disorder in twitter. Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. (ICWSM 2014) (2014)
14. Coppersmith, G., Dredze, M., Harman, C.: Quantifying mental health signals in twitter. ACL Workshop on Computational Linguistics and Clinical Psychology (2014)
15. Corrigan, P.: How stigma interferes with mental health care. American Psychologist **59(7)**, 614–625 (2004)
16. Coste, J., Pouchot, J.: A grey zone for quantitative diagnostic and screening tests. Int J Epidemiol **32(2): 304-13** (2003)
17. De Choudhury, M., Counts, S., Horvitz, E.: Predicting postpartum changes in emotion and behavior via social media. Proceedings of the 2013 ACM annual conference on Human factors in computing systems (2013)
18. De Choudhury, M., Counts, S., Horvitz, E., Hoff, A.: Characterizing and predicting postpartum depression from facebook data. Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing (2014)

19. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. AAAI Conference on Weblogs and Social Media (2013)
20. De Choudhury, M., Kiciman, E.: The language of social support in social media and its effect on suicidal ideation risk. Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM 2017) (2017)
21. 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. San Jose, California, USA May 07 - 12, 2016 (2016)
22. De Choudury, M., De, S.: Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. (ICWSM 2014). Ann Arbor, MI, Jun 2-Jun 4, 2014 (2014)
23. Demyttenaere, K., Bruffaerts, R., Posada-Villa, J., et al.: Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. The Journal of theAmerican Medical Association,JAMA **291(21)**, 2581–2590 (2004)
24. Ernala, S. K.and Rizvi, A.F., et al.: Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. Proceedings of the ACM Human-Computer Interaction.CSCW Online First. (2018)
25. Ernala, S.K., Birnbaum, M., R., A., Kane, J.D.C.M.: Characterizing audience engagement and assessing its impact on social media disclosures of mental illnesses. Proceedings of the 12th International AAAI Conference on Web and Social Media (2018)
26. Etkin, A., Wager, T.D.: Functional neuroimaging of anxiety: a meta-analysis of emotional processing in ptsd, social anxiety disorder, and specific phobia. American Journal of Psychiatry **164(10): 14761488** (2007)
27. Field, T.A., B.E., Jones, L.: The new abcs: A practitioner's guide to neuroscience-informed cognitive-behavior therapy. Journal of Mental Health Counseling **37(3)**, 206220 (2015)
28. Goffman, E.: Stigma: Notes on the management of spoiled identity. Prentice-Hall (1963)
29. Houghton, D., Joinson, A.: Linguistic markers of secrets and sensitive self-disclosure in twitter. pp. 3480–3489. System Science (HICSS), 2012 45th Hawaii International Conference on System Sciences (2012)
30. Hovy, D., Spruit, S.L.: The social impact of natural language processing. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (2016)
31. Johnson, G., Ambrose, P.: Neo-tribes: the power and potential of online communities in health care. Communications of the ACM **49(1)**, 107–113 (2006)
32. Kessler, R., Price, R.: Primary prevention of secondary disorders: A proposal and agenda. American Journal of Community Psychology **21(5)**, 607–633 (1993)
33. Kroenke, K., Spitzer, R., Williams, J.: The phq-9: validity of a brief depression severity measure. J Gen Intern Med **16(9):606.** (2001)
34. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. Proceedings of the 31st International Conference on Machine Learning (2014)
35. Loveys, K., Crutchley, P., Wyatt, E., Coppersmith, G.: Small but mighty: Affective micropatterns for quantifying mental health from social media language. Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology (2017)
36. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. ICLR Workshop Papers (2013)
37. Mitchell, M., Hollingshead, K., Coppersmith, G.: Quantifying the language of schizophrenia in social media. Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (2015)
38. Mrazek, P., Haggerty, R.: Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research. National Academies Press (1994)

39. Nelson, B., McGorry, P.D., Wichers, M., Wigman, J.T.W., Hartmann, J.A.: Moving from static to dynamic models of the onset of mental disorder: A review. JAMA Psychiatry **74(5)**, 528534 (2017)
40. Nguyen, T., Phung, D., Dao, B., Venkatesh, S., Berk, M.: Affective and content analysis of online depression communities. IEEE Transactions on Affective Computing **5(3): 217226** (2014)
41. Ono, E., Nozawa, T., Ogata, T., Motohashi, M., Higo, N., Kobayashi, T., et al.: Relationship between social interaction and mental health. IEEE/SICE International Symposium on System Integration (SII) (2011)
42. Park, M., McDonald, D.W., Cha, M.: Perception differences between the depressed and non-depressed users in twitter. Proceedings of the 7th International AAAI Conference on Weblogs and Social Media. (ICWSM 2013) (2013)
43. Pavalanathan, U., De Choudhury, M.: Identity management and mental health discourse on social media. In: Proceedings of WWW'15 Companion: 24th International World Wide Web Conference, Web Science Track. Florence, Italy, May 18-22, 2015 (2015)
44. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., et al.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12:2825-2830** (2011)
45. Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K.: The development and psychometric properties of liwc2015 (2015)
46. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2014)
47. Radloff, L.: The ces-d scale: A self-report depression scale for research in the general population. Applied Psychological Measurement **1:385-401** (1977)
48. Raglin, J.: Exercise and mental health.beneficial and detrimental effects. Sports Medicine **9**, 323329 (1990)
49. Robinson, M.S., Alloy, L.B.: Negative cognitive styles and stress-reactive rumination interact to predict depression: A prospective study. Cognitive Therapy and Research **27(3): 275-291** (2003)
50. Shedler, J., Block, J.: Adolescent drug use and psychological health: A longitudinal inquiry. American Psychologist **45(5)**, 612–630 (1990)
51. Shen, J., Rudzicz, F.: Detecting anxiety on reddit. Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology (2017)
52. Sox, H., Blatt, M., Hinggins, M., Marton, K.: Medical Decision Making. Boston: Butterworth-Heinemann (1987)
53. Sprecher, S., Treger, S., Wondra, J.D., Hilaire, N., Wallpe, K.: Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. Cognitive Therapy and Research **49(5): 860-866** (2003)
54. Stirn, A.: Body piercing: medical consequences and psychological motivations. The Lancet **361:12051215** (2003)
55. Taylor, E.: Assessing, Diagnosing, and Treating Serious Mental Disorders: A Bioecological Approach for Social Workers. Oxford University Press (2014)