Facebook language predicts depression in medical records

Johannes C. Eichstaedt[a,1,2], Robert J. Smith[b,1], Raina M. Merchant[b,c], Lyle H. Ungar[a,b], Patrick Crutchley[a,b], Daniel Preoţiuc-Pietro[a], David A. Asch[b,d], and H. Andrew Schwartz[e]

[a]Positive Psychology Center, University of Pennsylvania, Philadelphia, PA 19104; [b]Penn Medicine Center for Digital Health, University of Pennsylvania, Philadelphia, PA 19104; [c]Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, PA 19104; [d]The Center for Health Equity Research and Promotion, Philadelphia Veterans Affairs Medical Center, Philadelphia, PA 19104; and [e]Computer Science Department, Stony Brook University, Stony Brook, NY 11794

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved September 11, 2018 (received for review February 26, 2018)

Depression, the most prevalent mental illness, is underdiagnosed and undertreated, highlighting the need to extend the scope of current screening methods. Here, we use language from Facebook posts of consenting individuals to predict depression recorded in electronic medical records. We accessed the history of Facebook statuses posted by 683 patients visiting a large urban academic emergency department, 114 of whom had a diagnosis of depression in their medical records. Using only the language preceding their first documentation of a diagnosis of depression, we could identify depressed patients with fair accuracy [area under the curve (AUC) = 0.69], approximately matching the accuracy of screening surveys benchmarked against medical records. Restricting Facebook data to only the 6 months immediately preceding the first documented diagnosis of depression yielded a higher prediction accuracy (AUC = 0.72) for those users who had sufficient Facebook data. Significant prediction of future depression status was possible as far as 3 months before its first documentation. We found that language predictors of depression include emotional (sadness), interpersonal (loneliness, hostility), and cognitive (preoccupation with the self, rumination) processes. Unobtrusive depression assessment through social media of consenting individuals may become feasible as a scalable complement to existing screening and monitoring procedures.

big data | depression | social media | Facebook | screening

Each year, 7–26% of the US population experiences depression (1, 2), of whom only 13–49% receive minimally adequate treatment (3). By 2030, unipolar depressive disorders are predicted to be the leading cause of disability in high-income countries (4). The US Preventive Services Task Force recommends screening adults for depression in circumstances in which accurate diagnosis, treatment, and follow-up can be offered (5). These high rates of underdiagnosis and undertreatment suggest that existing procedures for screening and identifying depressed patients are inadequate. Novel methods are needed to identify and treat patients with depression.

By using Facebook language data from a sample of consenting patients who presented to a single emergency department, we built a method to predict the first documentation of a diagnosis of depression in the electronic medical record (EMR). Previous research has demonstrated the feasibility of using Twitter (6, 7) and Facebook language and activity data to predict depression (8), postpartum depression (9), suicidality (10), and posttraumatic stress disorder (11), relying on self-report of diagnoses on Twitter (12, 13) or the participants' responses to screening surveys (6, 7, 9) to establish participants' mental health status. In contrast to this prior work relying on self-report, we established a depression diagnosis by using medical codes from an EMR.

As described by Padrez et al. (14), patients in a single urban academic emergency department (ED) were asked to share access to their medical records and the statuses from their Facebook timelines. We used depression-related International Classification of Diseases (ICD) codes in patients' medical records as a proxy for the diagnosis of depression, which prior research has shown is feasible with moderate accuracy (15). Of the patients enrolled in the study, 114 had a diagnosis of depression in their medical records. For these patients, we determined the date at which the first documentation of a diagnosis of depression was recorded in the EMR of the hospital system. We analyzed the Facebook data generated by each user before this date. We sought to simulate a realistic screening scenario, and so, for each of these 114 patients, we identified 5 random control patients without a diagnosis of depression in the EMR, examining only the Facebook data they created before the corresponding depressed patient's first date of a recorded diagnosis of depression. This allowed us to compare depressed and control patients' data across the same time span and to model the prevalence of depression in the larger population (~16.7%).

Results

Prediction of Depression. To predict the future diagnosis of depression in the medical record, we built a prediction model by using the textual content of the Facebook posts, post length, frequency of posting, temporal posting patterns, and demographics (*Materials and Methods*). We then evaluated the performance of this model by comparing the probability of depression estimated by our algorithm against the actual presence or absence of depression for each patient in the medical record (using 10-fold cross-validation to avoid overfitting). Varying the threshold of this probability for diagnosis

Significance

Depression is disabling and treatable, but underdiagnosed. In this study, we show that the content shared by consenting users on Facebook can predict a future occurrence of depression in their medical records. Language predictive of depression includes references to typical symptoms, including sadness, loneliness, hostility, rumination, and increased self-reference. This study suggests that an analysis of social media data could be used to screen consenting individuals for depression. Further, social media content may point clinicians to specific symptoms of depression.

Author contributions: J.C.E., R.M.M., L.H.U., and H.A.S. designed research; J.C.E., P.C., D.P.-P., and H.A.S. performed research; J.C.E. and H.A.S. contributed new reagents/analytic tools; J.C.E., P.C., D.P.-P., and H.A.S. analyzed data; and J.C.E., R.J.S., R.M.M., L.H.U., D.A.A., and H.A.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the Open Science Framework, https://osf.io/zeuyc.

[1]J.C.E.and R.J.S. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: johannes.penn@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1802331115/-/DCSupplemental.

Published online October 15, 2018.

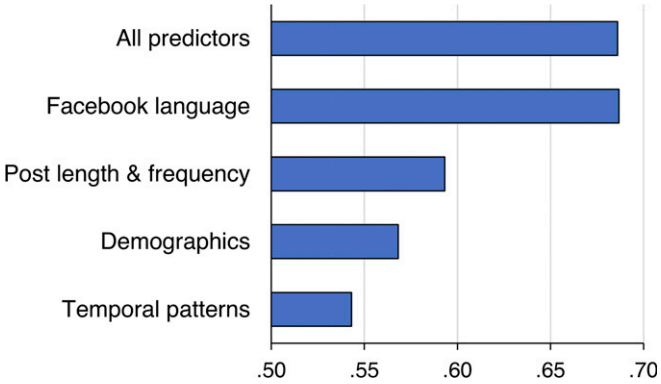

Fig. 1. Prediction performances of future diagnosis of depression in the EMR based on demographics and Facebook posting activity, reported as cross-validated out-of-sample AUCs.

uniquely determines a combination of true and false positive rates that form the points of a receiver operating characteristic (ROC) curve; overall prediction performance can be summarized as the area under the ROC curve. To yield interpretable and fine-grained language variables, we extracted 200 topics by using latent dirichlet allocation (LDA; ref. 16), a method akin to factor analysis but appropriate for word frequencies. We trained a predictive model based on the relative frequencies with which patients expressed these topics, as well as one-word and two-word phrases, obtaining an area under the curve (AUC) of 0.69, which falls just short of the customary threshold for good discrimination (0.70). As shown in Fig. 1, language features outperform other posting features and demographic characteristics, which do not improve predictive accuracy when added to the language-based model.

How do these prediction performances compare against other methods of screening for depression? Noyes et al. (17) assessed the concordance of screening surveys with diagnoses of depression recorded in EMRs as in this study*; the results are shown in Fig. 2 together with our Facebook model. The results suggest that the Facebook prediction model yields prediction accuracies comparable to validated self-report depression scales.

Previous work observed that depressed users are more likely to tweet during night hours (6). However, patients with and without a diagnosis of depression in our study differed only modestly in their temporal posting patterns (diurnally and across days of the week; AUC = 0.54). Post length and posting frequency (metafeatures) were approximately as predictive of depression in the medical record as demographic characteristics (AUCs of 0.59 and 0.57, respectively), with the median annual word count across posts being 1,424 words higher for users who ultimately had a diagnosis of depression (Wilcoxon $W = 26{,}594$, $P = 0.002$). Adding temporal pattern features and metafeatures to the language-based prediction model did not substantially increase prediction performance, suggesting that the language content captures the depression-related variance in the other feature groups.

Comparison with Previous Prediction Studies. In our sample, patients with and without a diagnosis of depression in the medical record were balanced at a 1:5 ratio to simulate true depression prevalence. In previous work, this balance has been closer to 1:1 (e.g., 0.94:1 in ref. 7, 1.78:1 in ref. 6). When limiting our sample to balanced classes (1:1), we obtain an AUC of 0.68 and F_1 score (the harmonic mean of precision and recall) of 0.66, which is comparable to the F_1 scores of 0.65 reported in ref. 7 and 0.68 reported in ref. 6 based on Twitter data and survey-reported depression. The fact that language content captures the depression-related variance in the other feature groups is consistent with what has been seen in previous work (6, 7, 18). However, this work shows that social media can predict diagnoses in medical records, rather than self-report surveys.

Predicting Depression in Advance of the Medical Record. We sought to investigate how far in advance Facebook may be able to yield a prediction of future depression. To that end, we considered language data for depressed patients from seven 6-mo windows preceding the first documentation of depression (or its matched time for controls) for the subset of 307 users who had at least 20 words in all seven windows. The results, shown in Fig. 3, suggest that the closer in time the Facebook data are to the documentation of depression, the better their predictive power, with data from the 6 mo immediately preceding the documentation of depression yielding an accuracy (i.e., AUC) of 0.72, surpassing the customary threshold of good discrimination (0.70). These results lend plausibility to the estimates of predictive power because one would expect just such a temporal trend. A minimal prediction of future depression (AUC = 0.62) above chance ($P = 0.002$) can be obtained approximately 3 mo in advance (3–9-mo window). Although this prediction accuracy is relatively modest, it suggests that, perhaps in conjunction with other forms of unobtrusive digital screening, the potential exists to develop burdenless indicators of mental illness that precede the medical documentation of depression (which may often be delayed) and which, as a result, could reduce the total extent of functional impairment experienced during the depressive episode.

Language Markers of Depression. To better understand what specific language may serve as markers of future depression and underlay the prediction performances of the aforementioned machine learning models, we determined how users with and

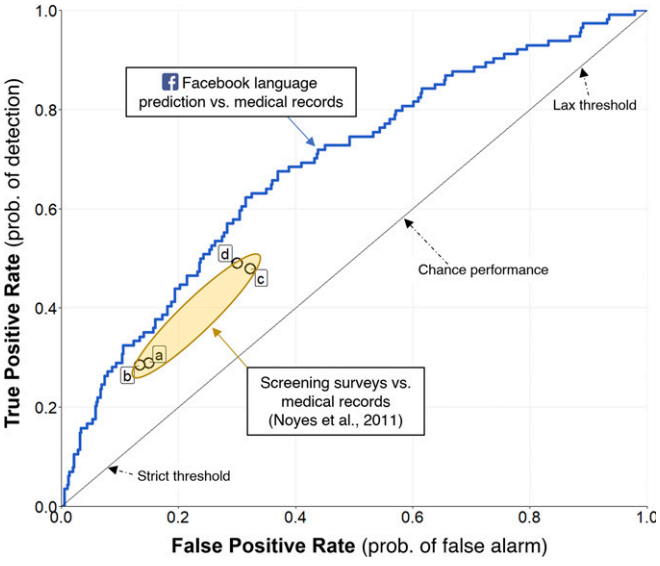

Fig. 2. ROC curve for a Facebook activity-based prediction model (all predictors combined; blue), and points as combinations of true and false positive rates reported by Noyes et al. (17) for different combinations of depression surveys (*a* and *b*, 9-item Mini-International Neuropsychiatric Interview–Major Depressive Episode Module; *c* and *d*, 15-item Geriatric Depression Scale with a cutoff >6) and time windows in Medicare claims data (*a* and *c*, within 6 mo before and after survey; *b* and *d*, within 12 mo).

*Noyes et al. (17) sought to benchmark claims data against self-report depression scales as the criterion variable in a sample of 1,551 elderly adults; we have derived the points given in Fig. 2 from the confusion matrices they published. They included the ICD-9 codes used in this study (296.2 and 311) among their "extended set" of codes.

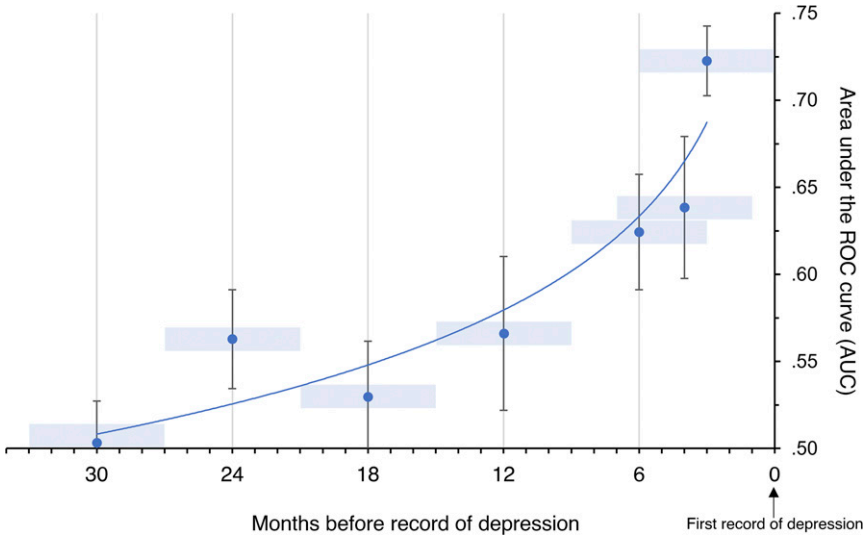

Fig. 3. AUC prediction accuracies of future depression status as a function of time before the documentation of depression in the medical record. Shown in blue are the 6-mo time windows of Facebook data used for the predictions; the blue dots indicate the AUCs obtained for these windows. Error bars indicate SEs (based on the 10 cross-validation folds). Logarithmic trendline is shown to guide the eye.

without a diagnosis of depression differed in the expression of the 200 data-driven LDA topics derived from their text.[†] In Fig. 4, we show the 10 topics most strongly associated with future depression status when controlling for age, gender, and race: 7 (of 200) topics were individually significant at $P < 0.05$ with Benjamini–Hochberg correction for multiple comparisons.

To complement this data-driven approach, we also examined the use of 73 prespecified dictionaries (lists of words) from the Linguistic Inquiry and Word Count (LIWC) software (2015; ref. 19) that is widely used in psychological research. Nine LIWC dictionaries predicted future depression status at Benjamini–Hochberg-corrected significance levels controlling for demographics (Table 1).

We observed emotional language markers of depressed mood (topic: *tears*, *cry*, *pain*; standardized regression coefficient $\beta = 0.15$; $P < 0.001$), loneliness (topic: *miss*, *much*, *baby*; $\beta = 0.14$; $P = 0.001$) and hostility (topic: *hate*, *ugh*, *fuckin*; $\beta = 0.12$; $P = 0.012$). The LIWC negative emotion ($\beta = 0.14$; $P = 0.002$; most frequent words: *smh*, *fuck*, *hate*) and sadness dictionaries ($\beta = 0.17$; $P < 0.001$; *miss*, *lost*, *alone*) captured similar information.

We observed that users who ultimately had a diagnosis of depression used more first-person singular pronouns (LIWC dictionary: $\beta = 0.19$; $P < 0.001$; *I*, *my*, *me*), suggesting a preoccupation with the self. First-person singular pronouns were found by a recent meta-analysis (20) to be one of the most robust language markers of cross-sectional depression (meta-analytic $r = 0.13$) and by a preliminary longitudinal study a marker of future depression, as observed in this study (21). Although there is substantial evidence that the use of first-person singular pronouns is associated with depression in private writings (22), this study extends the evidence for this association into the semipublic context of social media.

Cognitively, depression is thought to be associated with perseveration and rumination, specifically on self-relevant information (23), which manifests as worry and anxiety when directed toward the future (21). In line with these conceptualizations, we observed language markers suggestive of increased rumination (topic: *mind*, *alot*, *lot*; $\beta = 0.11$; $P = 0.009$) and anxiety (LIWC dictionary: $\beta = 0.08$; $P = 0.043$; *scared*, *upset*, *worry*), albeit not meeting Benjamini–Hochberg-corrected significance thresholds.

Depression often presents itself with somatic complaints in primary care settings (24, 25). In our sample, we observed that the text of users who ultimately received a diagnosis of depression contained markers of somatic complaints (topic: *hurt*, *head*, *bad*; $\beta = 0.15$; $P < 0.001$; LIWC dictionary, health: $\beta = 0.11$; $P = 0.004$; *life*, *tired*, *sick*). We also observed increased medical references (topic: *hospital*, *pain*, *surgery*; $\beta = 0.20$; $P < 0.001$), which is consistent with the finding that individuals with depression are known to visit the ED more frequently than individuals without depression (26).[‡]

Discussion

Our results show that Facebook language-based prediction models perform similarly to screening surveys in identifying patients with depression when using diagnostic codes in the EMR to identify diagnoses of depression. The profile of depression-associated language markers is nuanced, covering emotional (sadness, depressed mood), interpersonal (hostility, loneliness), and cognitive (self-focus, rumination) processes, which previous research has established as congruent with the determinants and consequences of depression. The growth of social media and continuous improvement of machine-learning algorithms suggest that social media-based screening methods for depression may become increasingly feasible and more accurate.

We chose to examine depression because it is prevalent, disabling, underdiagnosed, and treatable. As a major driver of medical morbidity and mortality, it is important to more thoroughly diagnose and treat depression across the population. Patients with depression exhibit poorer medical outcomes after acute inpatient care, increased utilization of emergency care resources, and increased all-cause mortality (25–28). Identifying patients at an earlier stage in their mental

[†]A language prediction model using only the 200 LDA topics (and not the relative frequencies of words and phrases) reaches an accuracy of AUC of 0.65, so the topics capture most of the language variance.

[‡]No topic or dictionary is negatively associated with future depression status (controlling for demographic characteristics) at significance levels corrected for multiple comparisons. The 10 LDA topics most negatively associated with depression status are shown in *SI Appendix*, Fig. S1. They cover language suggestive of gratitude, faith, school and work, and fitness and music consumption (*SI Appendix*, Table S1 includes an extended set of LIWC associations).

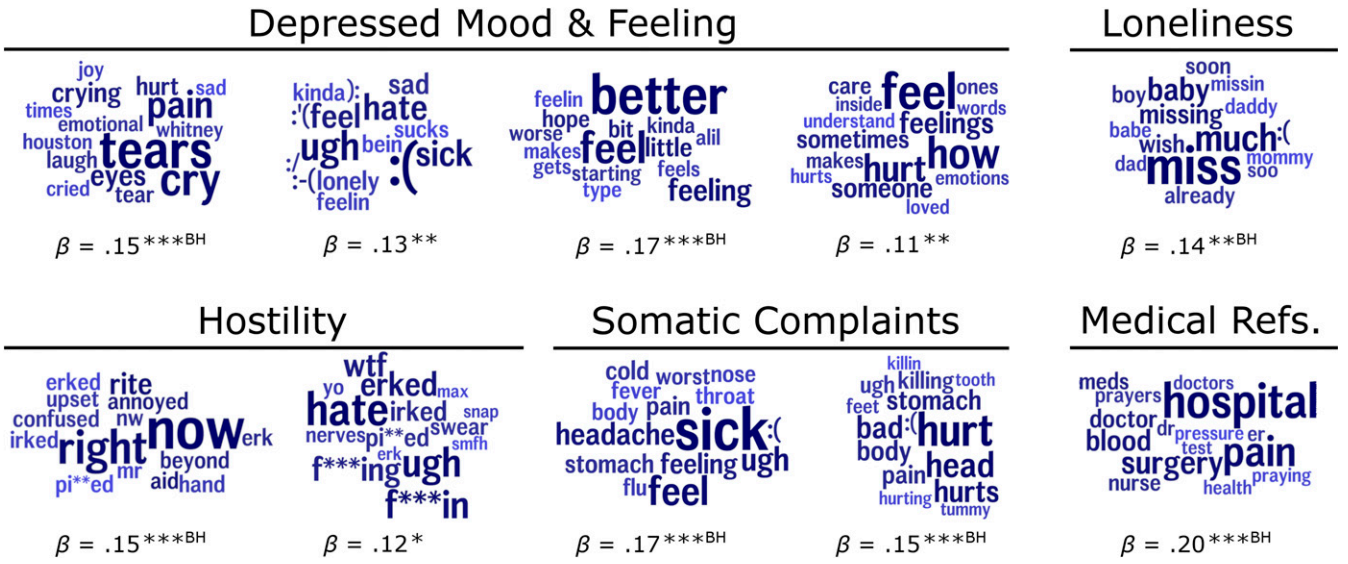

Fig. 4. Ten language topics most positively associated with a future depression diagnosis controlling for demographics (*$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$; $^{BH}P < 0.05$ after Benjamini–Hochberg correction for multiple comparisons). Font size reflects relative prevalence of words within topics. Color shading is to aid readability and carries no meaning.

illness through novel means of detection creates opportunities for patients to be connected more readily with appropriate care resources. The present analysis suggests that social media-based prediction of future depression status may be possible as early as 3 mo before the first documentation of depression in the medical record.

In the primary care setting, a diagnosis of depression is often missed (29). The reason for such underdetection is multifactorial: depression has a broad array of possible presenting symptoms, and its severity changes across time. Primary care providers are also tasked with addressing many facets of health within a clinical visit that may be as brief as 15 min. Previous research has recommended improving detection of depression through the routine use of multistep assessment processes (30). Initial identification of individuals who may be developing depression via analysis of social media may serve as the first step in such a process (using a detection threshold favoring high true positive rates). With the increasing integration of social media platforms, smartphones, and other technologies into the lives of patients, novel avenues are becoming available to detect depression unobtrusively. These methods include the algorithmic analysis of phone sensor, usage, and GPS position data on smartphones (31), and of facial expressions in images and videos, such as those shared on social media platforms (32, 33). In principle, these different screening modalities could be combined in a way that improves overall screening to identify individuals to complete self-report inventories (34) or be assessed by a clinician.

In the present study, patients permitted researchers to collect several years of retroactive social media data. These longitudinal data may allow clinicians to capture the evolution of depression severity over time with a richness unavailable to traditional clinical surveys delivered at discrete time points. The language exhibited by patients who ultimately developed depression was nuanced and varied, covering a wide array of depression-related symptoms. Changes in language patterns about specific symptoms could alert clinicians to specific depression symptoms among their consenting patients.

This study illustrates how social media-based detection technologies may optimize diagnosis within one facet of health. These technologies raise important question related to patient privacy, informed consent, data protection, and data ownership. Clear guidelines are needed about access to these data, reflecting the sensitivity of content, the people accessing it, and their purpose (35). Developers and policymakers need to address the challenge that the application of an algorithm may change social media posts into protected health information, with the corresponding expectation of privacy and the right of patients to remain autonomous in their health care decisions. Similarly, those who interpret the data need to recognize that people may change what they write based on their perceptions of how that information might be observed and used.

The key contribution of this study is that it links mental health diagnoses with social media content, and that it used this linkage to reveal associations between the content and symptoms of a prevalent, underdiagnosed, and treatable condition. This suggests that, one day, the analysis of social media language could serve as a scalable front-line tool for the identification of depressed individuals. Together with the growing

Table 1. LIWC Dictionaries Associated with Depression

LIWC dictionary	β	*P* value
Pronouns		
First pers singular (*I*, *me*)	0.19	***
Emotions		
Feel (perceptual process)	0.15	***
Negative emotions	0.14	**
Sadness	0.17	***
Cognitive processes		
Discrepancy	0.12	**
Other		
Health	0.11	**

Shown are all pronoun and psychological process LIWC 2015 dictionaries significantly associated with future depression status controlling for demographics, with strengths of associations given as standardized regression coefficients. All coefficients meet the $P < 0.05$ significance threshold when corrected for multiple comparisons by Benjamini–Hochberg method. Significantly correlated superordinate personal pronoun and pronoun dictionaries are not shown, which include the first-person singular pronoun dictionary shown here.

**$P < 0.01$, ***$P < 0.001$.

sophistication, scalability, and efficacy of technology-supported treatments for depression (36, 37), detection and treatment of mental illness may soon meet individuals in the digital spaces they already inhabit.

Materials and Methods

Participant Recruitment and Data Collection. This study was approved by the institutional review board at the University of Pennsylvania. The flow of the data collection is described in ref. 14. In total, 11,224 patients were approached in the ED over a 26-mo period. Patients were excluded if they were under 18 y old, suffered from severe trauma, were incoherent, or exhibited evidence of severe illness. Of these, 2,903 patients consented to share their social media data and their EMRs, which resulted in 2,679 (92%) unique EMRs. These EMRs were not specific to the ED but covered all patient encounters across the entire health care system. A total of 1,175 patients (44%) were able to log in to their Facebook accounts, and our Facebook app was able to retrieve any Facebook information and posts for as much as 6 y earlier, ranging from July 2008 through September 2015. These users shared a total of 949,530 Facebook statuses, which we used to model the 200 LDA topics.

From the health system's EMRs, we retrieved demographic data (age, sex, and race) and prior diagnoses (by ICD-9 codes). We considered patients as having a diagnosis of depression if their EMRs included documentation of ICD codes 296.2 (Major Depression) or 311 (Depressive Disorder, not elsewhere classified), resulting in 176 patients with any Facebook language (base rate 176/1,175 = 15.0%, or 1:5.68). Of the 176 depressed patients, 114 (63%) had at least 500 words in status updates preceding their first documentation of a diagnosis of depression. A total of 49 patients had no language preceding their first documentation, suggesting that, for 28% of the sample, their first documentation of depression preceded joining or the posting on Facebook. Notably, a depression-related ICD code could reflect self-report by the patient of a history of depression and did not necessarily imply clinical assessment or current depressive symptoms, treatment, or management [Trinh et al. (15) suggest that using ICD codes as a proxy for a diagnosis of depression is feasible with moderate accuracy].

To model the application in a medical setting and control for annual patterns in depression, for each patient with depression, we randomly selected another five patients without a history of depression who had at least 500 words in status updates preceding the same day as the first recorded diagnosis of depression. This yielded a sample of $114 + 5 \times 114 = 684$ patients who shared a total of 524,292 Facebook statuses in the included temporal window.[§] We excluded one patient from the sample for having less than 500 words after excluding unicode tokens (such as emojis), for a final sample of 683 patients.

Sample Description. Sample characteristics are shown in Table 2. Among all 683 patients, the mean age was 29.9 y (SD = 8.57); most were female (76.7%) and black (70.1%). Depressed patients were more likely to have posted more words on Facebook (difference between medians = 3,794 words; Wilcoxon $W = 27{,}712$; $P = 0.014$) and be female [χ^2 (1, $n = 583$) = 7.18; $P = 0.007$], matching national trends in presentations to urban academic EDs (26, 38, 39).

Word and Phrase Extraction. We determined the relative frequency with which users used words (unigrams) and two-word phrases (bigrams) by using our open-source Python-based language analysis infrastructure (dlatk.wwbp.org). We retained as variables the 5,381 words and phrases that were used by at least 20% of the sample across their 524,292 Facebook statuses.

Topic Modeling. As the coherence of topics increases when modeled over a larger number of statuses, we modeled 200 topics from the 949,530 Facebook statuses of all patients who agreed to share their Facebook statuses by using an implementation of LDA provided by the MALLET package (40). Akin to factor analysis, LDA produces clusters of words that occur in the same context across Facebook posts, yielding semantically coherent topics. It is appropriate for the highly nonnormal frequency distributions observed in language use. After modeling, we derived the use of 200 topics (200 values per user) for every user in the sample, which summarize their language use.

Table 2. Sample Descriptives

Sample descriptive	Depressed	Nondepressed	*P* value
No. of subjects	114	569	
Mean age (SD)	30.9 (8.1)	29.7 (8.65)	
Female, %	86.8	74.7	**
Black, %	75.4	69.1	
Mean word count (SD)	19,784 (27,736)	14,802 (21,789)	*
Median word count	10,655	6,861	*

Differences in age and mean word count were tested for significance by using *t* tests, percent female and black by using χ^2 tests with continuity correction, and median word counts by using Wilcoxon rank-sum test with continuity correction.
*$P < 0.05$, **$P < 0.01$.

Temporal Feature Extraction. We split the time of the day into six bins of 4 h in length, and, for every user, calculated the fraction of statuses posted in each of these bins. Similarly, we determined the fraction of posts made on each day of the week.

Metafeature Extraction. For every user, we determined how many unigrams were posted per year, the average length of the posts (in unigrams), and the average length of unigrams.

Dictionary Extraction. LIWC 2015 (41) provides dictionaries (lists of words) widely used in psychological research. We matched the extracted word frequencies against these dictionaries to determine the users' relative frequency of use of the 73 LIWC dictionaries.

Prediction Models. We used machine learning to train predictive models using the unigrams, bigrams, and 200 topics, using 10-fold cross-validation to avoid overfitting (similar to ref. 42). In this cross-validation procedure, the data are randomly partitioned into 10 stratified folds, keeping depressed users and their five "control users" within the same fold. Logistic regression models with a ridge penalty and their hyperparameters were fit within 9 folds and evaluated across the remaining held-out fold. The procedure was repeated 10 times to estimate an out-of-sample probability of depression for every patient. Varying the threshold of this probability for depression classification uniquely determines a combination of true and false positive rates that form the points of a ROC curve. We summarize overall prediction performance as the area under this ROC curve (i.e., AUC), which is suitable for describing classification accuracies over unbalanced classes.

Prediction in Advance of Documentation. We carried out the prediction as outlined earlier but truncated the available language data to time windows ranging from 0–6 mo before diagnosis (excluding the 24 h immediately before diagnosis) to 1–7, 3–9, 9–15, 15–21, 21–27, and 27–33 mo before the first documentation of depression in the medical records. We truncated the data analogously for control users. For this analysis, we limited the sample to those with data in each of the seven time windows, specifically thresholding at a total of 20 words total in each window. Because this lower threshold results in less stable measures of language use, we employed outlier removal, replacing feature observations that were more than 2 standard deviations from the mean with the feature's mean. This resulted in 307 patients (56 depressed) with the same users represented in each of the time windows (average word counts for depressed and nondepressed users in these windows are shown in *SI Appendix*, Fig. S2). AUCs were tested for significance against the null distribution through permutation tests with 100,000 permutations.

Language Associations. To determine if a language feature (topic or LIWC category) was associated with (future) depression status, we individually tested it as a predictor in an in-sample linear regression model controlling for demographic characteristics (binary variables for age quartile, ethnicity, and gender), and report its standardized regression coefficient (β) with the associated significance. We explored language correlations separately by gender but found that we had insufficient power to find language correlations among male users in the sample.

[§]We excluded 40 users with any Facebook language from the set of possible controls if they did not have the aforementioned ICD codes but only depression-like diagnoses that were not temporally limited, i.e., recurrent Depression (296.3) or Dysthymic Disorders (300.4), Bipolar Disorders (296.4–296.8), or Adjustment Disorders or Posttraumatic Stress Disorder (309). We additionally excluded 36 patients from the possible control group if they had been prescribed any antidepressant medications (i.e., selective serotonin reuptake inhibitors) without having been given an included depression ICD code.

Controlling for Multiple Comparisons. In addition to the customary significance thresholds, we also report whether a given language feature meets a $P < 0.05$ significance threshold corrected with the Benjamini–Hochberg procedure (43) for multiple comparisons.

Data Sharing. Medical record outcomes and the linked social media data are considered Protected Health Information and cannot be shared. However, for the main language features (200 DLA topics and 73 LIWC dictionaries), we are able to share mean levels and SDs for depressed and nondepressed users (deposited in Open Science Framework, https://osf.io/zeuyc/).

ACKNOWLEDGMENTS. We thank anonymous Reviewer 1 for her or his insightful suggestions. Support for this research was provided by a Robert Wood Johnson Foundation Pioneer Award; Templeton Religion Trust Grant TRT0048.

1. Demyttenaere K, et al.; WHO World Mental Health Survey Consortium (2004) Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. *JAMA* 291:2581–2590.
2. Kessler RC, et al.; National Comorbidity Survey Replication (2003) The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (NCS-R). *JAMA* 289:3095–3105.
3. Wang PS, et al. (2005) Twelve-month use of mental health services in the United States: Results from the national comorbidity survey replication. *Arch Gen Psychiatry* 62:629–640.
4. Mathers CD, Loncar D (2006) Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 3:e442.
5. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN (2009) Screening for depression in adult patients in primary care settings: A systematic evidence review. *Ann Intern Med* 151: 793–803.
6. De Choudhury M, Gamon M, Counts S, Horvitz E (2013) Predicting depression via social media. *ICWSM* 13:1–10.
7. Reece AG, et al. (2016) Forecasting the onset and course of mental illness with Twitter data. *Sci Rep* 7:13006.
8. Schwartz HA, et al. (2014) Towards assessing changes in degree of depression through Facebook. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (Association for Computational Linguistics, Stroudsburg, PA), pp 118–125.
9. De Choudhury M, Counts S, Horvitz EJ, Hoff A (2014) Characterizing and predicting postpartum depression from shared Facebook data. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Association for Computational Linguistics, Stroudsburg, PA), pp 626–638.
10. Homan CM, et al. (2014) *Toward Macro-insights for Suicide Prevention: Analyzing Fine-grained Distress at Scale* (Association for Computational Linguistics, Stroudsburg, PA).
11. Coppersmith G, Dredze M, Harman C (2014) Quantifying mental health signals in twitter. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (Association for Computational Linguistics, Stroudsburg, PA), pp 51–60.
12. Coppersmith G, Dredze M, Harman C, Hollingshead K, Mitchell M (2015) CLPsych 2015 shared task: Depression and PTSD on Twitter. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (Association for Computational Linguistics, Stroudsburg, PA), pp 31–39.
13. Pedersen T (2015) Screening Twitter users for depression and PTSD with lexical decision lists. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (Association for Computational Linguistics, Stroudsburg, PA), pp 46–53.
14. Padrez KA, et al. (2015) Linking social media and medical record data: A study of adults presenting to an academic, urban emergency department. *BMJ Qual Saf* 25: 414–423.
15. Trinh NHT, et al. (2011) Using electronic medical records to determine the diagnosis of clinical depression. *Int J Med Inform* 80:533–540.
16. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *J Mach Learn Res* 3: 993–1022.
17. Noyes K, Liu H, Lyness JM, Friedman B (2011) Medicare beneficiaries with depression: Comparing diagnoses in claims data with the results of screening. *Psychiatr Serv* 62: 1159–1166.
18. Preotiuc-Pietro D, et al. (2015) The role of personality, age and gender in tweeting about mental illnesses. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistics Signal to Clinical Reality* (Association for Computational Linguistics, Stroudsburg, PA), pp 21–30.
19. Pennebaker JW, Boyd RL, Jordan K, Blackburn K (2015) The development and psychometric properties of LIWC2015 (University of Texas, Austin).
20. Edwards T, Holtzman NS (2017) Ameta-analysis of correlations between depression and first person singular pronoun use. *J Res Pers* 68:63–68.
21. Zimmermann J, Brockmeyer T, Hunn M, Schauenburg H, Wolf M (2017) First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. *Clin Psychol Psychother* 24:384–391.
22. Tackman AM, et al. (2018) Depression, negative emotionality, and self-referential language: A multi-lab, multi-measure, and multi-language-task research synthesis. *J Pers Soc Psychol*, 10.1037/pspp0000187.
23. Sorg S, Vögele C, Furka N, Meyer AH (2012) Perseverative thinking in depression and anxiety. *Front Psychol* 3:20.
24. Rush AJ; Agency for Health Care Policy and Research (1993) Depression in primary care: Detection, diagnosis and treatment. *Am Fam Physician* 47:1776–1788.
25. Simon GE, VonKorff M, Piccinelli M, Fullerton C, Ormel J (1999) An international study of the relation between somatic symptoms and depression. *N Engl J Med* 341: 1329–1335.
26. Boudreaux ED, Cagande C, Kilgannon H, Kumar A, Camargo CA (2006) A prospective study of depression among adult patients in an urban emergency department. *Prim Care Companion J Clin Psychiatry* 8:66–70.
27. Fan H, et al. (2014) Depression after heart failure and risk of cardiovascular and all-cause mortality: A meta-analysis. *Prev Med* 63:36–42.
28. Zheng D, et al. (1997) Major depression and all-cause mortality among white adults in the United States. *Ann Epidemiol* 7:213–218.
29. Perruche F, et al. (2010) Anxiety and depression are unrecognized in emergency patients admitted to the observation care unit. *Emerg Med J* 28:662–665.
30. Mitchell AJ, Vaze A, Rao S (2009) Clinical diagnosis of depression in primary care: A meta-analysis. *Lancet* 374:609–619.
31. Saeb S, et al. (2015) Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study. *J Med Internet Res* 17:e175.
32. Levi G, Hassner T (2015) Emotion recognition in the wild via convolutional neural networks and mapped binary patterns. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (Association for Computational Linguistics, Stroudsburg, PA), pp 503–510.
33. Ringeval F, et al. (2017) AVEC 2017: Real-life depression, and affect recognition workshop and challenge. *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge* (Association for Computing Machinery, New York), pp 3–9.
34. Gilbody S, Sheldon T, House A (2008) Screening and case-finding instruments for depression: A meta-analysis. *CMAJ* 178:997–1003.
35. Grande D, Mitra N, Shah A, Wan F, Asch DA (2013) Public preferences about secondary uses of electronic health information. *JAMA Intern Med* 173:1798–1806.
36. Foroushani PS, Schneider J, Assareh N (2011) Meta-review of the effectiveness of computerised CBT in treating depression. *BMC Psychiatry* 11:131.
37. Newman MG, Szkodny LE, Llera SJ, Przeworski A (2011) A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: Is human contact necessary for therapeutic efficacy? *Clin Psychol Rev* 31:89–103.
38. Rhodes KV, et al. (2001) Better health while you wait: A controlled trial of a computer-based intervention for screening and health promotion in the emergency department. *Ann Emerg Med* 37:284–291.
39. Kumar A, Clark S, Boudreaux ED, Camargo CA, Jr (2004) A multicenter study of depression among emergency department patients. *Acad Emerg Med* 11:1284–1289.
40. McCallum AK (2002) Mallet: A Machine Learning for Language Toolkit. mallet.cs.umass.edu.
41. Pennebaker JW, Boyd RL, Jordan K, Blackburn K (2015) The development and psychometric properties of LIWC2015 (University of Texas, Austin).
42. Kosinski M, Stillwell D, Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proc Natl Acad Sci USA* 110:5802–5805.
43. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300.