Suicide and Life-Threatening BEHAVIOR
THE OFFICIAL JOURNAL OF THE AMERICAN ASSOCIATION OF SUICIDOLOGY


Suicide and Life-Threatening Behavior 47 (1) February 2017

DOI: 10.1111/sltb.12312

# A Machine Learning Approach to Identifying the Thought Markers of Suicidal Subjects: A Prospective Multicenter Trial

John P. Pestian, PhD, Michael Sorter, MD, Brian Connolly, PhD, Kevin Bretonnel Cohen, PhD, Cheryl McCullumsmith, MD, PhD, Jeffry T. Gee, MD, Louis-Philippe Morency, PhD, Stefan Scherer, PhD, and Lesley Rohlfs, MS, for the STM Research Group*

Death by suicide demonstrates profound personal suffering and societal failure. While basic sciences provide the opportunity to understand biological markers related to suicide, computer science provides opportunities to understand suicide thought markers. In this novel prospective, multimodal, multicenter, mixed demographic study, we used machine learning to measure and fuse two classes of suicidal thought markers: verbal and nonverbal. Machine learning algorithms were used with the subjects' words and vocal characteristics to classify 379 subjects recruited from two academic medical centers and a rural community hospital into one of three groups: suicidal, mentally ill but not suicidal, or controls. By combining linguistic and acoustic characteristics, subjects could be classified into one of the three groups with up to 85% accuracy. The results provide insight into how advanced technology can be used for suicide assessment and prevention.

Predicting when someone will commit suicide has been nearly impossible (American Psychiatric Association, 2003; Goldstein, Black, Nasrallah, & Winokur, 1991; Hughes, 1995; Large & Ryan, 2014; Olav Nielssen, 2012; Paris, 2006), but classifying the factors that contribute to suicide risk is possible with standardized, clinical tools when used by well-trained clinicians (Beck, Beck, & Kovacs, 1975; Beck, Kovacs, & Weissman, 1979; Bürk, Kurz, & Möller, 1985; Columbia-Suicide Severity Rating Scale, 2015;

John P. Pestian, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA; Michael Sorter, Division of Psychiatry, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA; Brian Connolly, Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA; Kevin Bretonnel Cohen, Computational Bioscience Program, University of Colorado School of Medicine, Denver, CO, USA; Cheryl McCullumsmith, Department of Psychiatry, College of Medicine, University of Cincinnati, Cincinnati, OH, USA; Jeffry T. Gee, Princeton Community Hospital, Princeton, WV, USA; Louis-Philippe Morency, Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA; Stefan Scherer, Institute for Creative Technologies, University of Southern California, Los Angeles, CA, USA.

*STM Research Group: Robert Faist, Leslie Korbee, Lisa McCord, Rachel McCourt, and Colleen Murphy.

Address correspondence to John P. Pestian, Cincinnati Children's Hospital Medical Center, University of Cincinnati, 3333 Burnet Ave., Cincinnati, OH 45229; E-mail: john.pestian@cchmc.org

Kovacs & Garrison, 1985; Müller & Dragicevic, 2003; Mundt, Greist, & Jefferson, 2013; Pokorny, 1983; Posner et al., 2008; Preston & Hansen, 2005; U.S. Food & Drug Administration, 2012; Winters, Myers, & Proud, 2002). Such tools can, however, be cumbersome and may not reliably translate into routine interactions between clinicians, caregivers, or educators. Here, we describe a novel approach using the subjects' linguistic and acoustic patterns to classify subjects automatically as either suicidal, mentally ill but not suicidal, or a control.

## BACKGROUND

Efforts to understand suicide risks can be roughly clustered into traits or states. *Trait analyses* focus on stable characteristics rooted in and measured using biological processes (Costanza et al., 2014; Le-Niculescu et al., 2013), whereas *state analyses* measure dynamic characteristics like verbal and nonverbal communication, termed "thought markers" (Pestian et al., 2015). Machine learning and natural language processing have successfully identified differences in retrospective suicide notes, newsgroups, and social media (Gomez, 2014; Huang, Goh, & Liew, 2007; Matykiewicz, Duch, & Pestian, 2009). Jashinsky et al. (2015) used multiple annotators to identify the risk of suicide from the keywords and phrases (interrater reliability = .79) in geographically based tweets. Thompson, Poulin, and Bryan (2014) and Desmet (2014) used text-based signals to identify suicide risk that ranged from 60% to 90%. Li, Ng, Chau, Wong, and Yip (2013) presented a framework using machine learning to identify individuals expressing suicidal thoughts in web forums; Zhang et al. (2015) used microblog data to build machine learning models that identified suicidal bloggers with approximately 90% accuracy. Pestian, Matykiewicz, and Grupp-Phelan (2008) demonstrated that machine learning algorithms could distinguish between notes written by people who died by suicide and simulated suicide notes written by age- and gender-matched controls better than mental health professionals could (71% vs. 79%; Pestian et al., 2008). In an international, shared task-setting that included multiple groups sharing the same task definition, data set, and scoring metric (Voorhees et al., 2005), 24 teams developed and tested computational algorithms to identify emotions in over 1,319 suicide notes written shortly before death. The results showed that the fusion of multiple methods outperform single methods (Pestian, Matykiewicz, & Linn-Gust, 2012).

Suicidal thought markers have also been studied prospectively. The Suicidal Adolescent Clinical Trial (Pestian et al., 2015), the single-site precursor to this study, which used machine learning to analyze interviews with 60 suicidal and control patients, classified patients into suicidal or control groups with greater than 90% accuracy (Pestian et al., 2015). Analysis of acoustic features such as pauses and vowel spacing yielded similar results (Scherer, Morency, Gratch, Pestian, & Playa Vista, 2015; Venek, Scherer, Morency, Rizzo, & Pestian, 2014). The study described herein is novel because it uses a multisite, multicultural setting to show that machine learning algorithms can be trained to automatically identify the suicidal subjects in a group of suicidal, mentally ill, and control subjects. Moreover, the inclusion of acoustic characteristics is most helpful when classifying between suicidal and mentally ill subjects.

## METHODS

### *Subject Enrollment*

Between October 2013 and March 2015, 379 subjects were enrolled from emergency departments (EDs) and inpatient and outpatient centers into a three-site, internal review board-approved prospective clinical trial. One hundred twenty-six subjects were enrolled at Cincinnati Children's Hospital Medical Center (CCHMC), a 600-bed urban level 1 academic medical center. One hundred twenty-eight subjects were enrolled at the University of Cincinnati Medical Center (UC), a 498-bed urban academic medical center. One hundred twenty-five subjects were enrolled at Princeton Community Hospital (PCH), a 267-bed Appalachian community hospital in southern West Virginia. The inclusion and exclusion criteria classified subjects into one of three groups: suicidal, mentally ill, or control.

Multi-gated inclusion criteria were used. For the first gate, all patients were reviewed using the electronic status board for a complaint of suicide, suicidal ideation, or psychiatric evaluation. For patients with mental illness and for control subjects, any complaint was accepted except those related to suicide. For those who passed the first gate, a second review of their electronic medical record (EMR) was conducted before they were approached for enrollment. Suicidal subjects were approached if they had come to the EDs or psychiatric units because of suicidal ideation or attempts within the previous 24 hours. Patients with mental illness were enrolled from the ED and outpatient mental health clinics if they had a definitive mental illness diagnosis but had not had prior suicidal attempts, active thoughts of suicide, or plans to die by suicide within the previous year as reported by the patient and EMR. Control subjects were patients who came to the ED with no history of mental health diagnoses or suicidal ideation, as reported by the patient and EMR (Figure 1).

Potential subjects were excluded if their native language was not English, if they had any serious medical injury or mental retardation that could prohibit consent, if they would be unavailable for a follow-up interview, or if they could not comply with study procedures.

Participation incentives were site-specific. CCHMC and UC subjects were paid $50 for the initial interview and $25 for the follow-up. PCH subjects were paid $25 for the initial interview and $25 for the follow-up interview.

### *Data Collection*

Data were collected and validated by trained mental health professionals. During enrollment, each subject completed standardized tools: Columbia-Suicide Severity Rating Scale, Young Mania Rating Scale, and Hamilton Rating Scale for Depression. Each subject also completed the ubiquitous questionnaire (UQ), a semistructured interview with five open-ended questions to stimulate conversation for language sampling: “Do you have hope?” “Do you have any fear?” “Do you have any secrets?” “Are you angry?” and “Does it hurt emotionally?” (Pestian, 2010; Pestian et al., 2015). Both subject and interviewer were video and audio recorded during the UQs. The results were transcribed with 98% accuracy based on completeness, accuracy of transcription, and adherence to the transcription guidelines.

### *Computational Analysis*

*Linguistic and Acoustic Feature Extraction.* Our analysis is based on two types of features: linguistic and acoustic. Feature extraction was performed automatically using the patient audio signals recorded during the interviews and their transcriptions. Linguistically, we followed related work on automatic identification and extraction of word instances (unigrams) and word-pair instances (bi-grams) from the transcriptions. A dictionary that includes all spoken words, word-pairs, and acoustic characteristics was created.

The selected vocal and prosodic characteristics include: vocal dynamics—fundamental frequency (f0; Drugman & Alwan, 2011) and square of the amplitude (A2); voice quality–harmonic richness factor (Childers & Lee, 1991), maximum dispersion quotient (Kane, 2012), peak slope (D’Alessandro & Sturmel, 2011), difference between the first and second harmonics (Hillenbrand, Cleveland, & Erickson, 1994), normalized amplitude quotient (Alku, Bäckström, & Vilkman, 2002),

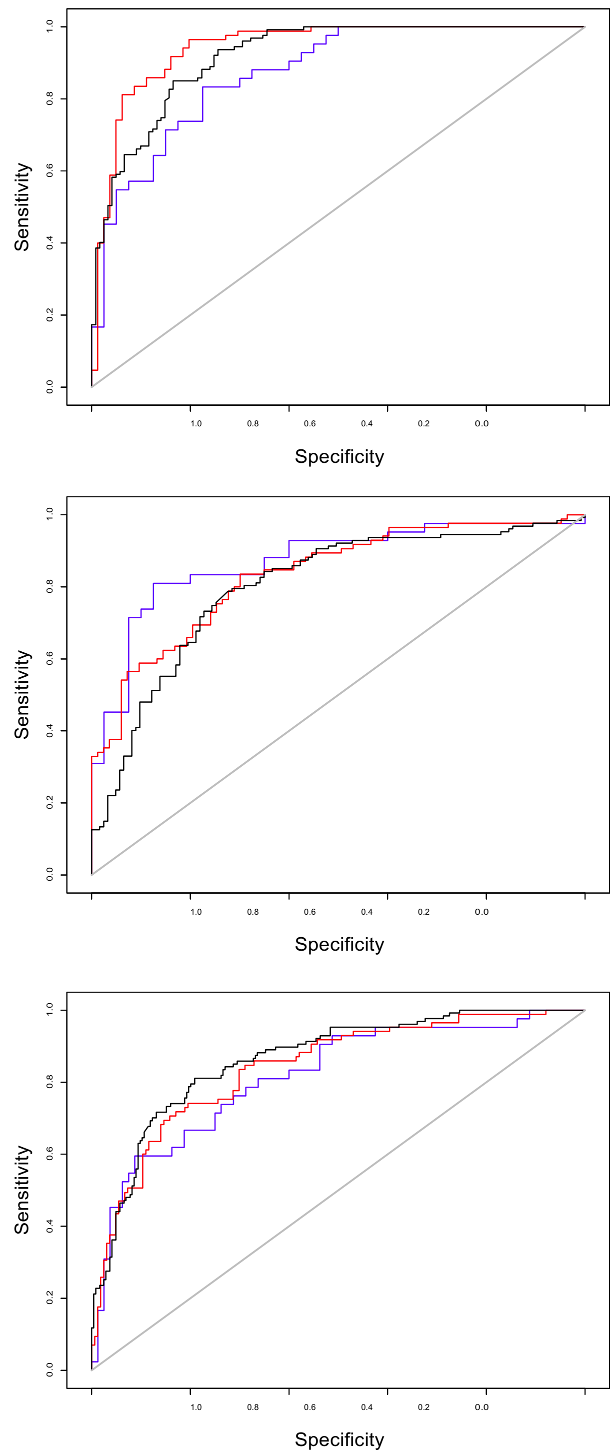


Figure 1. Receiver operator curve (ROC): suicide versus control (upper), suicide versus mentally ill (middle), and suicide versus mentally ill with control. The ROC curves for adolescents (blue), adults (red), and all subjects (black) generated where the nonsuicidal population is controls (top), mentally ill (middle), and mentally ill and controls, using linguistic and acoustic features. The gray line is the AROC curve for a baseline (random) classifier.

quasi-open quotient (Hacki, 1989), and parabolic spectral parameters (Alku, Strik, & Vilkman, 1997); the vocal tract resonance frequencies, as well as pause lengths characterized by formants F1 through F5 (Kwon, Chan, Hao, & Lee, 2003); and pause lengths that have been correlated with depression (Cummins et al., 2015). Once a subset of all features were extracted using the COVAREP software (Degottex, Kane, Drugman, Raitio, & Scherer, 2014), they were normalized by adjusting the measured values from the various features to a common zero-to-one scale (Dodge, 2006).

*Machine Learning Algorithms.* A goal of machine learning is to train a computational model from selected data that can then generalize to unseen (test) data. Machine learning can be roughly divided into three types: supervised learning, when the training data are already labelled; semi-supervised learning, when only part of the training set is labelled; and unsupervised learning, when the challenge is to learn structure in unlabelled data. Here, a supervised learning support vector machine (SVM) approach was used (Schölkopf & Smola, 1998).

Support vector machines are based on a computational learning theory called structural risk minimization, whose goal is to find a hypothesis with the lowest true error (Vapnik, 1999). The SVM constructs a hyperplane in a high-dimensional space, which can be used for classification, regression, or other tasks (Press, Teukolsky, Vetterling, & Flannery, 2007). The SVM is appropriate for this study's data because it can be used on multiple class problems, and because its connection to computational learning enables it to be a universal learner (Joachims, 1998) while in support of conceptual frameworks, such as spreading activation. Consequently, it tends to be fairly robust to overfitting (Sebastiani, 2002). The performance of the classifier was based on the area under the receiver operating curve, which was estimated using leave-one-out by subject cross-validation (Efron & Tibshirani, 1997; Molinaro, Simon, & Pfeiffer, 2005). Data were analyzed by members of the study team.

## RESULTS

### *Subjects*

A total of 955 patients were approached for enrollment. Of that group, 576 subjects did not meet the inclusion criteria ($n = 70$), refused to participate ($n = 436$), or were excluded for other reasons ($n = 70$). This resulted in 379 subjects enrolled, comprising 130 suicidal patients, 126 nonsuicidal patients with mental illness, and 123 controls. Eight subject interviews were incomplete and were excluded from the final analysis. A total of 371 subjects completed the study. The patient demographic characteristics within each study site are shown in Table 1.

### *Comparison of the Performance of Classification Algorithms*

Figure 1 and Table 2 show the performance of the machine learning algorithm in classifying subjects into suicidal and nonsuicidal subject groups. Classification performances are shown for adolescents, adults, and the combined adolescent and adult cohort. The table shows that the ROC threshold of 0.80 is met in all cases, except adults in the adult suicide versus mentally ill comparison. When these data are combined with adolescent data, however, the threshold is met. The table also shows that the signal from acoustic characteristics boosts the suicidal versus mentally ill classification.

Overall, the results show that machine learning algorithms can be trained to automatically identify the suicidal subjects in a group of suicidal, mentally ill, and control subjects. Moreover, the inclusion of acoustic characteristics is most helpful when classifying between suicidal and mentally ill subjects.

**TABLE 1**
*Averages and Standard Deviations (SD) of Demographic and Psychological Rating Scale Measures for Males and Females, According to Site*

| | Cincinnati Children's Hospital Medical Center | | | | Princeton Community Hospital | | | | University of Cincinnati Medical Center | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Male (*SD*) | Female (*SD*) | *p* value | Total (*SD*) | Male (*SD*) | Female (*SD*) | *p* value | Total (*SD*) | Male (*SD*) | Female (*SD*) | *p* value | Total (*SD*) |
| *n* | 39 | 83 | <0.0001 | 122 | 46 | 77 | 0.007 | 123 | 59 | 67 | 0.53 | 126 |
| Average age in years | 15.9 (1.1) | 15.5 (1.3) | 0.08 | 15.6 (1.3) | 41.0 (12.1) | 43.0 (12.3) | 0.38 | 42.2 (12.3) | 42.1 (12.9) | 43.1 (13.2) | 0.67 | 42.6 (13.2) |
| Average time of interview in minutes | 8.4 (3.3) | 7.2 (3.0) | 0.06 | 7.6 (3.2) | 6.7 (4.7) | 6.6 (4.0) | 0.90 | 6.7 (4.2) | 8.8 (4.6) | 10.8 (5.4) | 0.03 | 9.9 (5.2) |
| Patients approached | 74 | 154 | <0.0001 | 228 | 204 | 326 | <0.0001 | 530 | 94 | 103 | 0.57 | 197 |

| | All hospitals | | | |
|---|---|---|---|---|
| | Male (*SD*) | Female (*SD*) | *p* value | Total (*SD*) |
| *n* | 144 | 227 | <0.0001 | 371 |
| Average age in years | 34.6 (15.7) | 33.0 (16.7) | 0.35 | 33.6 (16.4) |
| Average time of interview in minutes | 8.1 (4.5) | 8.1 (4.5) | 1.0 | 8.1 (4.5) |
| Patients approached | 372 | 583 | <0.0001 | 955 |

**TABLE 2**
*The AROC for the Machine Learning Algorithm. The Nonsuicidal Group Comprises of Either Mentally Ill and Control Subjects. Classification Performances are Shown for Adolescents, Adults, and the Combined Adolescent and Adult Cohorts*

| | Suicidal versus Controls | | | Suicidal versus Mentally Ill | | | Suicidal versus Mentally Ill and Controls | | |
|---|---|---|---|---|---|---|---|---|---|
| | Adolescents ROC (*SD*) | Adults ROC (*SD*) | Adolescents + Adults ROC (*SD*) | Adolescents ROC (*SD*) | Adults ROC (*SD*) | Adolescents + Adults ROC (*SD*) | Adolescents ROC (*SD*) | Adults ROC (*SD*) | Adolescents + Adults ROC (*SD*) |
| Linguistics | 0.87 (0.04) | 0.91 (0.02) | 0.93 (0.02) | 0.82 (0.05) | 0.77 (0.04) | 0.79 (0.03) | 0.82 (0.04) | 0.84 (0.03) | 0.87 (0.02) |
| Acoustics | 0.74 (0.05) | 0.82 (0.03) | 0.79 (0.03) | 0.69 (0.06) | 0.74 (0.04) | 0.76 (0.03) | 0.74 (0.05) | 0.80 (0.03) | 0.76 (0.03) |
| Linguistics + Acoustics | 0.83 (0.05) | 0.93 (0.02) | 0.92 (0.02) | 0.80 (0.05) | 0.77 (0.04) | 0.82 (0.03) | 0.81 (0.04) | 0.84 (0.03) | 0.87 (0.02) |

## DISCUSSION

We anticipated that when computation methods are applied in multiple sites, their overall accuracy decreases. This decrease could be reduced by including a second measurement mode, such as acoustic data. Although there was a decrease from our initial study (Pestian et al., 2015), the decrease was not substantial. Moreover, the acoustic features did not play a substantial role in the initial study interview. Subsequent research, however, has shown that in some cases the acoustic features are statistically important during follow-up visits (Venek, Scherer, Morency, Rizzo, & Pestian, 2016; Venek et al., 2014). From the results of other studies, this was unexpected. Suggesting that additional research devoted to fusing nonverbal sentiment data with verbal sentiment data is still needed.

Clinicians may ask: Were there any differences in what the subjects said? Table 3 shows that the mentally ill and the control patients tended to laugh more during interviews, sigh less, and express less anger, less emotional pain, and more hope.

The study's sample was based on the subjects' self-reports, which means that some patients could have been disingenuous. But, using measures of authenticity, no differences were found between each of the groups (Newman, Pennebaker, Berry, & Richards, 2003).

## CONCLUSION

This study's methodology presents strong evidence for a useful objective tool that clinicians and others can use to determine suicidal intention. These computational approaches may provide novel opportunities for large-scale innovations in suicidal care. The methodology described here can be readily translated to such settings as school, shelters, youth clubs, juvenile justice centers, and community centers, where earlier identification may help to reduce suicide attempts and deaths.

**TABLE 3**
*Sentiments Expressed During the Interviews by Group*

| Sentiment expressed by subject | % of suicidal interviews | % of mentally ill interviews | % of controls | ANOVA *p* value |
|---|---|---|---|---|
| Laughed during interview | 39.4 | 69.1 | 71.9 | <0.0001 |
| Responded negatively to "Are you angry?" | 47.2 | 67.5 | 86.0 | <0.0001 |
| Laughed during response to anger question | 12.6 | 35.0 | 41.3 | <0.0001 |
| Sighed during the interview | 14.2 | 0.81 | 1.65 | 0.05 |
| Responded negatively to "Does it hurt emotionally?" | 40.9 | 66.7 | 70.2 | <0.0001 |
| Used the word "hope" in response to the question "Do you have hope?" | 48.0 | 72.4 | 75.2 | <0.0001 |

## REFERENCES


ALKU, P., BÄCKSTRÖM, T., & VILKMAN, E. (2002). Normalized amplitude quotient for parametrization of the glottal flow. *The Journal of the Acoustical Society of America*, *112*, 701–710.

ALKU, P., STRIK, H., & VILKMAN, E. (1997). Parabolic spectral parameter—A new method for quantification of the glottal flow. *Speech Communication*, *22*, 67–79.

American Psychiatric Association. (2003). Practice guideline for the assessment and treatment of patients with suicidal behaviors. *American Journal of Psychiatry*, *160*(Suppl), 1–60.

BECK, A. T., BECK, R., & KOVACS, M. (1975). Classification of suicidal behaviors: I. Quantifying intent and medical lethality. *The American Journal of Psychiatry*, *132*, 285–87.

BECK, A. T., KOVACS, M., & WEISSMAN, A. (1979). Assessment of suicidal intention: The Scale for Suicide Ideation. *Journal of Consulting and Clinical Psychology*, *47*, 343.

BÜRK, F., KURZ, A., & MÖLLER, H.-J. (1985). Suicide risk scales: Do they help to predict suicidal behaviour? *European Archives of Psychiatry and Neurological Sciences*, *235*, 153–157.

CHILDERS, D. G., & LEE, C. (1991). Vocal quality factors: Analysis, synthesis, and perception. *The Journal of the Acoustical Society of America*, *90*, 2394–2410.

Columbia-Suicide Severity Rating Scale. (2015) Training–Public health settings. Retrieved from http://www.cssrs.columbia.edu/training_cssrs.html

COSTANZA, A., D'ORTA, I., PERROUD, N., BURKHARDT, S., MALAFOSSE, A., MANGIN, P., ET AL. (2014). Neurobiology of suicide: Do biomarkers exist? *International Journal of Legal Medicine*, *128*, 73–82.

CUMMINS, N., SCHERER, S., KRAJEWSKI, J., SCHNIEDER, S., EPPS, J., & QUATIERI, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10–49.

D'ALESSANDRO, C., & STURMEL, N. (2011). Glottal closure instant and voice source analysis using time-scale lines of maximum amplitude. *Sadhana*, *36*, 601–622.

DEGOTTEX, G., KANE, J., DRUGMAN, T., RAITIO, T., & SCHERER, S. (2014). Covarep—A collaborative voice analysis repository for speech technologies. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing* (pp. 960–964), Florence, Italy.

DESMET, B. (2014, May). *Finding the online cry for help*, PhD diss., Universiteit Gent, Gent, Belgium.

DODGE, Y. (2006). *The Oxford dictionary of statistical terms: Entry for normalization*. New York: Oxford University Press.

DRUGMAN, T., & ALWAN, A. (2011). Joint robust voicing detection and pitch estimation based on residual harmonics. In *Interspeech* (pp. 1973–1976), Florence, Italy.

EFRON, B., & TIBSHIRANI, R. (1997). Improvements on cross-validation: The .632 + bootstrap method. *Journal of the American Statistical Association*, *92*, 548–560.

GOLDSTEIN, R. B., BLACK, D. W., NASRALLAH, A., & WINOKUR, G. (1991). The prediction of suicide: Sensitivity, specificity, and predictive value of a multivariate model applied to suicide among 1906 patients with affective disorders. *Archives of General Psychiatry*, *48*, 418–422.

GOMEZ, J. M. (2014). Language technologies for suicide prevention in social media. In *Workshop on Natural Language Processing in the 5th Information Systems Research Working Days*, (pp. 21-17), Quito, Ecuador.

HACKI, T. (1989). Klassifizierung von glottisdysfunktionen mit hilfe der elektroglottographie. *Folia Phoniatrica et Logopaedica*, *41*, 43–48.

HILLENBRAND, J., CLEVELAND, R. A., & ERICKSON, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, *37*, 769–778.

HUANG, Y. P., GOH, T., & LIEW, C. L. (2007). Hunting suicide notes in web 2.0-preliminary findings. In *Proceedings of the 9th IEEE International Symposium on Multimedia* (pp. 517–521). Taichung, Tiawan.

HUGHES, D. H. (1995). Can the clinician predict suicide? *Psychiatric Services*, *46*, 449–451.

JASHINSKY, J., BURTON, S. H., HANSON, C. L., WEST, J., GIRAUD-CARRIER, C., BARNES, M. D., ET AL. (2015). Tracking suicide risk factors through twitter in the US. *Crisis*, *35*, 51–59.

JOACHIMS, T. (1998). *Text categorization with support vector machines: Learning with many relevant features*. Berlin: Springer.

KANE, J. (2012). *Tools for analysing the voice*. PhD diss., Trinity College, Dublin, Ireland.

KOVACS, M., & GARRISON, B. (1985). Hopelessness and eventual suicide: A 10-year prospective study of patients hospitalized with suicidal ideation. *American Journal of Psychiatry*, *142*, 559–563.

KWON, O.-W., CHAN, K., HAO, J., & LEE, T.-W. (2003). Emotion recognition by speech signals. In. *Eurospeech* (pp. 125–128), Geneva, Switzerland.

LARGE, M., & RYAN, C. (2014). Suicide risk assessment: Myth and reality. *International Journal of Clinical Practice*, *68*, 679–681.

LE-NICULESCU, H., LEVEY, D., AYALEW, M., PALMER, L., GAVRIN, L., JAIN, N., ET AL. (2013). Discovery and validation of blood biomarkers for suicidality. *Molecular Psychiatry*, *18*, 1249–1264.

LI, T. M., NG, B. C., CHAU, M., WONG, P. W., & YIP, P. S. (2013). Collective intelligence for suicide surveillance in web forums. In *Intelligence and security informatics* (pp. 29–37). Berlin: Springer.

MATYKIEWICZ, P., DUCH, W., & PESTIAN, J. (2009). Clustering semantic spaces of suicide notes and newsgroups articles. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing* (pp. 179–184). Stroudsburg, PA: Association for Computational Linguistics.

MOLINARO, A., SIMON, R., & PFEIFFER, R. (2005). Prediction error estimation: A comparison of resampling methods. *Bioinformatics*, *21*, 3301–3307.

MÜLLER, M. J., & DRAGICEVIC, A. (2003). Standardized rater training for the Hamilton Depression Rating Scale (HAMD-17) in psychiatric novices. *Journal of Affective Disorders*, 77, 65–69.

MUNDT, J. C., GREIST, J. H., & JEFFERSON, J. W. (2013). Original research prediction of suicidal behavior in clinical research by. *Journal of Clinical Psychiatry*, *74*, 887–893.

NEWMAN, M. L., PENNEBAKER, J. W., BERRY, D. S., & RICHARDS, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*, 665–675.

OLAV NIELSSEN, M. (2012). Suicidal ideation and later suicide. *The American Journal of Psychiatry*, *169*, 662.

PARIS, J. (2006). Predicting and preventing suicide: Do we know enough to do either? *Harvard Review of Psychiatry*, *14*, 233–240.

PESTIAN, J. P. (2010). A conversation with Edwin Shneidman. *Suicide and Life-Threatening Behavior*, *40*, 516–523.

PESTIAN, J. P., GRUPP-PHELAN, J., BRETONNEL COHEN, K., MEYERS, G., RICHEY, L. A., MATYKIEWICZ, P., ET AL. (2015). A controlled trial using natural language processing to examine the language of suicidal adolescents in the emergency department. *Suicide and Life-Threatening Behavior*, *46*, 154–159.

PESTIAN, J. P., MATYKIEWICZ, P., & GRUPP-PHELAN, J. (2008). Using natural language processing to classify suicide notes. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing* (pp. 96–97). Stroudsburg, PA: Association for Computational Linguistics.

PESTIAN, J. P., MATYKIEWICZ, P., & LINN-GUST, M. (2012). What's in a note: Construction of a suicide note corpus. *Biomedical Informatics Insights*, *5*, 1.

POKORNY, A. D. (1983). Prediction of suicide in psychiatric patients: Report of a prospective study. *Archives of General Psychiatry*, *40*, 249–257.

POSNER, K., BRENT, D., LUCAS, C., GOULD, M., STANLEY, B., BROWN, G., ET AL. (2008). *Columbia-Suicide Severity Rating Scale (C-SSRS)*. New York, NY: New York State Psychiatric Institute.

PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T., & FLANNERY, B. (2007).

Section 16.5. Support vector machines. In *Numerical recipes: The art of scientific computing*, (3rd ed.). New York: Cambridge University Press.

PRESTON, E., & HANSEN, L. (2005). A systematic review of suicide rating scales in shizophrenia. *Crisis*, *26*, 170–180.

SCHERER, S., MORENCY, L.-P., GRATCH, J., PESTIAN, J., & PLAYA VISTA, C. (2015). Reduced vowel space is a robust indicator of psychological distress: A crosscorpus analysis. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*. (pp. 4789–4793), Brisbane, Australia.

SCHÖLKOPF, B., & SMOLA, A. (1998). Support vector machines. In *Encyclopedia of biostatistics* New York, NY: John Willey & Sons.

SEBASTIANI, F. (2002 March). Machine learning in automated text categorization. *ACM Computing Surveys*, *34*, 1–47.

THOMPSON, P., POULIN, C., & BRYAN, C. J. (2014). Predicting military and veteran suicide risk: Cultural aspects. In *Proceedings from the Workshop on Computational Linguistics and Clinical Psychology* (pp. 1–6) Baltimore, MD.

U.S. Food and Drug Administration. (2012). *Guidance for industry: Suicidal ideation and behavior: Prospective assessment of occurrence in clinical trials*. Silver Springs, MD: Author. Retrieved October, 5, 2012, from http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/htm.

VAPNIK, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*, 988–999.

VENEK, V., SCHERER, S., MORENCY, L.-P., RIZZO, A., & PESTIAN, J. (2014). Adolescent suicidal risk assessment in clinician-patient interaction: A study of verbal and acoustic behaviors. In *Proceedings of the IEEE Spoken Language Technology Workshop*, (pp. 277–282), Lake Tahoe, Nevada.

VENEK, V., SCHERER, S., MORENCY, L.-P., RIZZO, A., & PESTIAN, J. (2016). Adolescent suicidal risk assessment in clinician-patient interaction. In *IEEETransactions on Affective Computing*.

VOORHEES, E. M., HARMAN, D. K., ET AL. (2005). *TREC: Experiment and evaluation in information retrieval*. Cambridge: MIT Press.

WINTERS, N. C., MYERS, K., & PROUD, L. (2002). Ten-year review of rating scales. III: Scales assessing suicidality, cognitive style, and self-esteem. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*, 1150–1181.

ZHANG, L., HUANG, X., LIU, T., LI, A., CHEN, Z., & ZHU, T. (2015). Using linguistic features to estimate suicide probability of Chinese microblog users. In *Human Centered Computing* (pp. 549–559). Cham, Switzerland: Springer.

Manuscript Received: January 17, 2016
Revision Accepted: June 28, 2016

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article: **Appendix S1**.