Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Special Issue Article

Prediction of Loneliness in Older Adults Using Natural Language Processing: Exploring Sex Differences in Speech

Varsha D. Badal, Ph.D., Sarah A. Graham, Ph.D., Colin A. Depp, Ph.D., Kaoru Shinkawa, M.E., Yasunori Yamada, Ph.D., Lawrence A. Palinkas, Ph.D., Ho-Cheol Kim, Ph.D., Dilip V. Jeste, M.D., Ellen E. Lee, M.D.

ARTICLE INFO

ABSTRACT

Objective: The growing pandemic of loneliness has great relevance to aging populations, though assessments are limited by self-report approaches. This paper explores the use of artificial intelligence (AI) technology to evaluate interviews on loneliness, notably, employing natural language processing (NLP) to quantify sentiment and features that indicate loneliness in transcribed speech text of older adults. *Design:* Participants completed semi-structured qualitative interviews regarding the experience of loneliness and a quantitative self-report scale (University of California Los Angeles or UCLA Loneliness scale) to assess loneliness. Lonely and non-lonely participants (based on qualitative and quantitative assessments) were compared. *Setting:* Independent living sector of a senior housing community in San Diego County. *Participants:* Eighty English-speaking older adults with age range 66−94 (mean 83 years). *Measurements:* Interviews were audiotaped and manually transcribed. Transcripts were examined using NLP approaches to quantify sentiment and expressed emotions. *Results:* Lonely individuals (by qualitative assessments) had longer responses with greater expression of sadness to direct questions about loneliness. Women were more likely to endorse feeling lonely during the qualitative interview. Men used more fearful and joyful words in their responses. Using linguistic features, machine learning models could predict qualitative loneliness with 94% precision (sensitivity = 0.90, specificity = 1.00) and quantitative loneliness with 76% precision (sensitivity = 0.57,

Prediction of Loneliness in Older Adults Using Natural Language Processing

specificity = 0.89). **Conclusions:** *AI (e.g., NLP and machine learning approaches) can provide unique insights into how linguistic features of transcribed speech data may reflect loneliness. Eventually linguistic features could be used to assess loneliness of individuals, despite limitations of commercially developed natural language understanding programs.* (Am J Geriatr Psychiatry 2020; ■■:■■−■■)

INTRODUCTION

The loneliness pandemic has been associated with serious physical and mental health consequences, rivaling smoking and obesity.[1−5] Loneliness also has economic consequences like lost productivity, greater healthcare utilization, and indirect costs (estimated to be over $3 billion annually). These cost estimates included the increased risk of cognitive decline and development of dementia among lonely individuals, while controlling for demographic factors, social isolation, and mood symptoms.[6−9] Older individuals are at particularly high risk for loneliness due to loss of partners and friends, as well as declining physical health and mobility.[10] While rates of loneliness have been previously found to be fairly stable,[2] prevalence of loneliness among older adults may rise due to the rapidly growing older population, increased loneliness with aging,[3] increased social isolation,[4,5] and potential contribution of physical distancing measures related to the COVID-19 pandemic.[11]

Qualitative analysis of interviews is an important approach to understanding the experience of loneliness, especially for vulnerable populations like older adults. While several reports have examined qualitative experiences of loneliness among immigrant populations,[12] medically ill persons,[13,14] and people at highest risk for loneliness,[15] there are few qualitative studies of independently living older adults. Our recent qualitative study of residents of senior housing communities found that despite living in a communal setting with services designed to reduce social isolation, many older adults reported feeling lonely.[16] While loneliness and social isolation may be interrelated, loneliness is a distinct construct − some people feel "lonely in a crowd" while others are content with few social connections.[17] Furthermore, findings from qualitative (e.g., open-ended or semi-structured interviews) and quantitative (e.g., based on the University of California Loneliness Scale - version 3 or UCLA-3) assessments of loneliness reflect discrepancies that

warrant further investigation. For example, sex differences in loneliness appear to be driven by assessment type. In response to direct questions about loneliness (e.g., "Do you feel lonely?"), women may be more likely to report feeling lonely.[18] However men and women have similar scores on the commonly used UCLA-3 loneliness scale (which does not explicitly use the term "lonely").[10] Better understanding of sex differences in reporting loneliness can refine assessment measures and guide interventions for loneliness.

Due to the time- and effort-intensive nature of data analyses, most qualitative studies have been limited to small scales, e.g., experiences of 20−30 individuals, which may not capture a breadth of perspectives and to focusing on overall themes, which may be subject to the rater's biases. Such qualitative studies focus on commonly expressed viewpoints, rather that the sociodemographic and clinical features that may distinguish individuals with other opinions. Similarly, nuanced features such as word choice, expressed emotions, and sentence structure are not easily assessed by the human eye. Unstructured speech data are a unique window into an individual's experience of loneliness. Emerging data science strategies like automated speech-to-text (transcription), natural language processing (NLP) and machine learning (ML) can be used to gain novel insights from unstructured speech data and scale up qualitative analyses.

NLP[19,20] refers to a variety of techniques (including but not limited to parts-of-speech tagging, named entity recognition and parsing) that process, analyze and manipulate text to get insights and information from unstructured text data. Natural language understanding (NLU) is a subset of NLP which is more aligned with comprehension of the analyzed text and enables tasks such as reasoning, translation, summarization, question-answering, sentiment and emotion analysis.

Some recent investigations using NLP tools for psychiatric applications include predicting psychiatric

readmission,[21] suicidality[22,23] or mental health crises[24]; diagnosing mental illnesses[25]; and predicting treatment outcomes in patients with depression.[26] These applications used a variety of NLP tools including rule-based systems (systems that use explicitly stated If/Then/Else rules), artificial neural networks (ANN, models inspired by neurons that use weighted sums of inputs and activation functions), and deep neural networks (multilayer ANN with each layer representing more advanced representation). NLP and NLU techniques can enable quantification of abstract fuzzy constructs such as loneliness on dimensions of sentiment and the embedded emotions, though their use has been limited in psychiatry. To our knowledge, specific text features of lonely individuals and their sex differences have not been previously examined in older adults.

In this study, we conducted semi-structured qualitative interviews about loneliness and completed quantitative loneliness assessments with residents of a continuing care senior housing community. The interviews were analyzed using NLP to identify differences in transcribed speech patterns in lonely versus non-lonely individuals (based on qualitative and quantitative assessments). For this proof-of-concept study, we explored how NLP analytic methods could assess whether individuals reported feeling lonely in response to a direct question about loneliness (e.g., "Do you feel lonely?"). We explored how responses of lonely individuals differed in length, sentiment, and emotion from non-lonely individuals (using qualitative and quantitative measures of loneliness). We also explored sex differences in the response features. Lastly, we investigated the possibility of automated prediction of loneliness (through ML models) using only text features.

RESEARCH DESIGN AND METHODS

Participants and Procedures

Study procedures and subjects have been described previously.[16,27] Briefly, subjects were recruited from the independent living sector of a senior housing community in San Diego County. This continuing care senior housing community has 278 independent residential units and offers all three levels of care: independent living, assisted living, and memory care. All subjects provided a written informed consent for study participation.

Selection criteria for enrollment were: 1) English-speaking individuals ≥65 years, 2) Ability to complete study assessments and engage in a qualitative interview, and 3) No known diagnosis of dementia or any other disabling illness. This study protocol was approved by the University of California San Diego Human Research Protections Program and the administrators of the housing community. Participants were recruited through short presentations using Human Research Protections Program-approved script and flyers.

Sociodemographic and Clinical Measures

Trained study staff gathered sociodemographic data including age, sex assigned at birth, racial background, and marital status. They administered scales to assess emotional support (Emotional Support Scale), anxiety (Brief Symptom Inventory − Anxiety subscale),[28] and depression (Patient Health Questionnaire, 9-item).[29]

Quantitative Loneliness Measure

The UCLA Loneliness Scale (Version 3) or UCLA-3 is the most commonly used measure of loneliness, with strong test-retest reliability, high internal consistency, and validity.[30] While the word "lonely" is never used explicitly in the 20-item scale, subjects are asked to report the frequency of specific experiences (e.g., "How often do you feel in tune with others around you?") on a 4-point Likert scale (1 = "I never feel this way" to 4 = "I often feel this way.") The cutoffs for loneliness severity on the UCLA-3 scale were adapted from Doryab (2019)[31] and include: total score less than or equal to 40 as Not lonely, total score greater than 40 as Lonely.

Qualitative Interviews

Trained study staff conducted semi-structured interviews with participants between April 2018 and August 2019. The interview format followed a predetermined list of broad, research-driven probes developed by study investigators[16]; however, the interview was intended to be conducted in a conversational way. The first question inquired directly about

loneliness: (Q1) "Do you ever feel lonely, and if so, how often?" If the participant endorsed feeling lonely, the follow-up question was: (Q2) "What does loneliness feel like to you? What is your general mood during that time?" If the participant denied feeling lonely, the follow-up question was: (Q3) "Why do you think others may feel lonely?" Interviewers were trained in qualitative methods according to research techniques outlined by Patton.[32] Each interview was audio-taped and transcribed (maximum length of 90 minutes).

Analytic Procedures

In order to create the dataset, we targeted the responses to primary questions from the interview to gain insights into loneliness. We identified the location of the first loneliness question in the transcript and analyzed the sentiment and emotional content of the responses to the loneliness question (Q1) using IBM Watson NLU iv program[33] depicted in Figure 1.

We manually established ground truth for *interview-based* or qualitative assessment by interpreting the response text to Q1 (as acknowledging versus denying loneliness) and labeling the dataset (lonely versus not lonely). Each Q1 response was independently coded by two trained raters (EEL, SAG) to reflect qualitative loneliness ("yes" versus "no"). Kappa was 0.90, indicating a high degree of concordance among the raters.[34] Disagreements in qualitative loneliness classification were adjudicated by a third author (VDB). We also used UCLA-3 scores to establish the ground truth for *quantitative assessment*.

FIGURE 1. Processing pipeline for the qualitative interview data. API: application programming interface; NLU: natural language understanding; Q1: Question 1 ("Do you ever feel lonely, and if so, how often?"); Q2: Question 2 ("What does loneliness feel like to you? What is your general mood during that time?"); Q3: Question 3 ("Why do you think others may feel lonely?"); TF-IDF: term frequency − inverse document frequency.
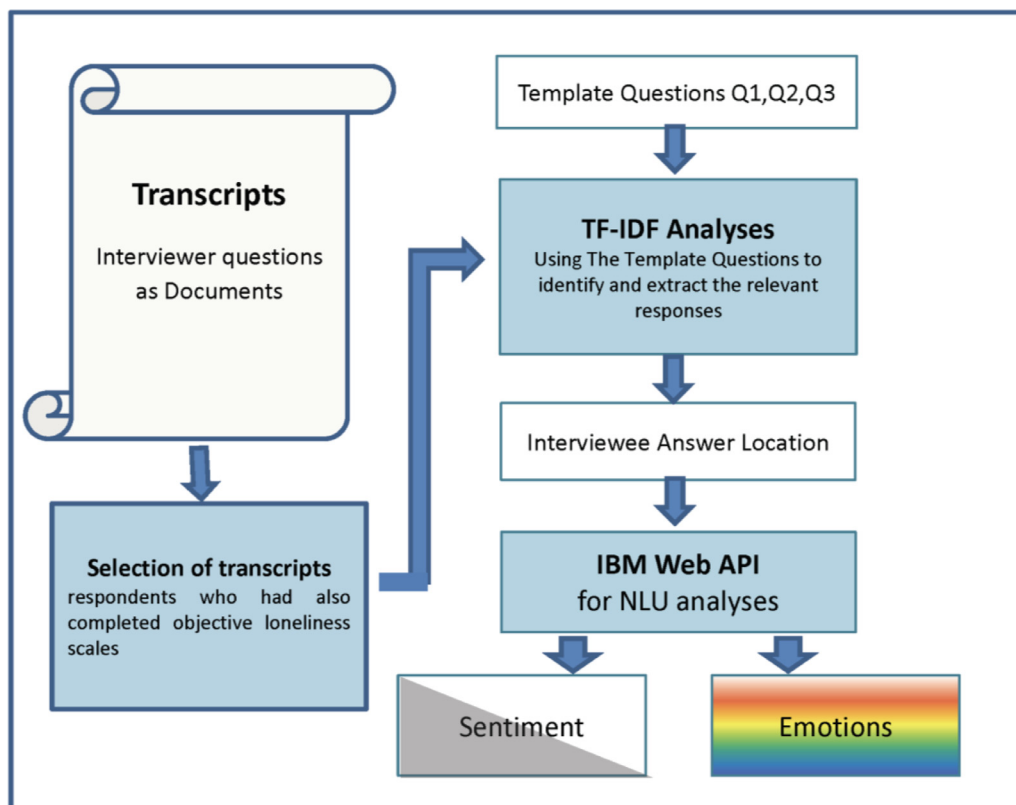
We used ML models to predict both classifications of loneliness.

Text processing

Due to semi-structured nature of the interview and unconstrained responses from interviewee, we identified location of relevant questions (and subsequently, the responses) using term frequency − inverse document frequency (TF-IDF) techniques,[35,36] that are commonly used in document retrieval and data mining.[37] The TF-IDF scores serve as features in ML classification (described later). In the transcripts, each question starts on a new line preceded by the "Q:" characters. Each question is analogous to a "document" and the transcript to a "corpus" in TF-IDF terminology. The procedure is repeated for each transcript.

First, the corpus (or collection of documents) is converted into vectors that capture both frequency of words (henceforth referred to as "terms") and uniqueness of the terms contained in the document. Queries, or specific spans of text, are also vectorized and compared with documents to identify matches. TF-IDF "searches" for sections of text within each transcript that best match the query, thus extracting specific sections of text from transcripts. Further details regarding TF-IDF are available in the *Supplemental materials* (Appendix A). Once the location of question was identified, we extracted the following lines (marked with "A:" in the transcribed interview text) as the answer provided by the subject. The number of characters (including spaces and punctuations) constituted the length. As the length of responses varied greatly, from a few characters up to thousands of characters in length, the results were presented using a log scale (logarithm to base 10) for the histogram (e.g., 10 characters would be log(10) = 1, 100 characters would be log(100) = 2.)

IBM NLU tools

The *IBM Cloud* contains a suite of advanced data and artificial intelligence (AI) tools that are widely available for users [https://www.ibm.com/cloud]. IBM NLU iv (IBM, Watson NLU) was used for sentiment and emotion analysis of the text data. These tools were selected for their robustness and applicability for the research question. Other tools (reasoning, translation, summarization, and question-answering) attempt to solve more complex AI tasks, and the current state of art is not suitable to general application. Most systems for these problems are exploratory and work in very limited domains and scopes. Reasoning and translation were not relevant to the task. Usage details are publicly available[38] and details of these tools are discussed in the supplemental material (Appendix B).

Sentiment (positive and negative) is represented as a number [continuous range between −1.0 and 1.0], indicating speaker is in (total) disagreement or (total) agreement with the current context of conversation. *Emotion* is a five-tuple (sadness, joy, fear, disgust, anger) containing values [continuous range between 0.0 and 1.0], in proportion to the strength for each dimension of emotion. Complex emotions can be comprised from these basic dimensions.[39]

Once the response to Q1 is extracted, we used the IBM NLU tool to evaluate its sentiment and emotions. Supplemental Figure 1 depicts IBM NLU IV output of sentiment and emotion analysis based on a sample response to Q1.[40]

We compared lonely versus non-lonely individuals (by both qualitative and quantitative assessments) by length, sentiment, and emotional content using Mann-Whitney U tests (for continuous variables) and Fisher's exact test or Spearman's correlation (for categorical variables). For all analyses, unadjusted two-tailed p-values were considered significant at p less than 0.05. Significance was defined as Type I error alpha = 0.05 (two-tailed) for all analyses. The effect sizes presented include Cohen's d (parametric) and Cliff's delta (nonparametric). Cliff's delta was computed using available software.[41] The statistical analyses were conducted using the IBM SPSS Version 25 (IBM Corp., Armonk, NY) and R.

ML models

Features for the ML models included sentiment and emotions (joy, fear, anger, disgust, sadness) obtained from NLU analysis of response to Q1, TF-IDF score of top matching document to Q1, as well as presence of Q2 and Q3. Of note, presence of follow-up questions Q2 and Q3 depended on the interviewee's response to Q1. We used these nine features to classify interviewees into: qualitative loneliness categories [True, False] and quantitative loneliness categories [True, False].We assessed ML performance using Orange3,[42] a data-mining toolbox with random

80-20 training-testing data split. We selected a broad range of ML models in order to accommodate different types of data. ML methods included: support vector machine (SVM with variety of kernels: linear, polynomial, and radial basis function), k-Nearest Neighbors (kNN), Tree, AdaBoost, ANN (activation functions included tanh, rectified linear unit and logistic), random forest and a stacking of aforementioned methods.[43,44] We ranked the features for the two classification tasks using three popular methods (GINI, ANOVA, and chi-squared scores).[45−47] These methods are described in greater detail in Appendix C. Ensemble techniques are a common approach where several ML models are used, especially to assess novel domains and applications, and achieve better performance than would be possible by committing to any single one.[43,44,48−50]

We used Orange3 visual programming tool that provides sophisticated widgets for ML applications. The Orange3 processing code for all ML models used in the study are provided as a separate file and described in the Supplemental materials (Appendix D). Orange3 is available for public download from (https://orange.biolab.si).

RESULTS

Ninety-seven unique interviews were completed and manually transcribed. Seventeen of these transcripts were removed from the analyses (four lacked baseline data and thirteen lacked UCLA-3 data), resulting in eighty transcripts (sum total of 1,021,969 words and target document Q1 length of 10 words.) Distribution of transcript lengths are depicted in Supplemental Figure 4.

Description of the Study Sample

Mean age of interviewees was 83.0 years (SD = 6.9 years, range 66−94 years) (Table 1). Men were older than the women. Education, racial background, marital status, proportion with qualitative and quantitative loneliness, mean UCLA-3 scores, instrumental support, negative interactions, anxiety and depression were similar by sex. Women reported greater emotional support.

Overall incidence of loneliness by qualitative assessment was 45%. Of the 30 people with UCLA-3 scores above the lonely cutoff (37.5% of respondents), 11 (36.7%) did not report feeling lonely in response to Q1. Examples of specific responses to Q1 and the qualitative ratings are shown in Supplemental Table 1. The Kappa score of agreement between the qualitative and quantitative assessments of loneliness was 0.28.

Response Analyses

Location of answer corresponding to Q1 in the transcripts was identified correctly for all 80 interviewees. The presence of Q2 and Q3 TF-IDF scores were related to manually-scored Q1 responses

TABLE 1. Sociodemographic and Clinical Data of the Interviewees by Sex

	Women			Men			t or X²	df	p	Cohen's d
	N	Mean	SD	N	Mean	SD				
Age at Visit (years)	51	81.6	7.1	29	85.5	5.7	−2.51	78	0.01	−0.85
Education (years)	51	15.4	2.4	29	16.3	2.1	−1.76	78	0.08	−0.59
Race (% Caucasian)		90.2			93.1		0.20	1	0.66	
Marital Status (% not single)		37.3			51.7		1.58	1	0.21	
Qualitative Lonely (% yes)		52.9			31.0		3.58	1	0.06	
Quantitatively Lonely (% yes)		33.3			44.8		1.04	1	0.31	
UCLA-3 Score	51	36.5	9.4	29	38.7	11.2	−0.92	78	0.36	−0.30
Emotional Support (ESS-E)	51	2.8	0.4	29	2.5	0.5	2.15	78	0.04	0.69
Instrumental Support (ESS-I)	51	1.9	0.8	29	1.8	0.8	0.88	78	0.38	0.29
Negative social interactions (ESS-NI)	51	0.7	0.8	29	0.8	0.7	−1.06	78	0.29	−0.35
Anxiety (BSIAS)	51	1.6	2.6	28	1.4	1.5	0.33	77	0.75	0.12
Depression (PHQ-9)	48	2.8	3.6	27	3.0	3.6	−0.31	73	0.76	−0.11

BSIAS: Brief Symptom Inventory Anxiety Scale; ESS-E: Emotional Support Scale − Emotional Support score; ESS-I: Emotional Support Scale − Instrumental Support score; ESS-NI: Emotional Support Scale − Negative Interaction Score; PHQ-9: Patient Health Questionnaire 9-item; UCLA-3: UCLA Loneliness Scale (Version 3).

(qualitative rating) (Supplemental Fig. 2). Length of Q1 responses varied greatly (word count M = 69.2 SD = 168.2; character count M = 331.2, SD = 802.5). Q1 responses were longer in respondents who were lonely by qualitative assessment (Mann-Whitney U = 426, p <0.001, Cliff's delta = 0.46) and also by quantitative assessment (UCLA-3 score >40) (Mann-Whitney U = 581.0, p = 0.047, Cliff's delta = 0.23) (Fig. 2).

We mapped the distribution of emotions expressed in the responses to Q1. The respondents who acknowledged feeling lonely were more likely to express sadness in their responses (Mann-Whitney U = 543.0, p = 0.008, Cliff's delta = 0.31). Expression of sentiment and other emotions (disgust, anger, joy, fear) did not differ between lonely versus non-lonely groups (Fig. 3).

Sex Differences in Reported Loneliness

Discrepancies between qualitative and quantitative loneliness assessments differed by sex. Women were more likely than men to be lonely by qualitative but not quantitative assessments (endorsing loneliness in the interview and having UCLA-3 scores ≤40) (76.4% of women versus 46.1% of men). Men were more likely than women to be lonely by quantitative but not qualitative assessments (having UCLA-3 scores >40 and not endorsing loneliness in the interview.) Women were more likely to acknowledge feeling lonely in interviews, when they were quantitatively lonely compared to men (Fisher's exact p <0.001). Fourteen (27%) women reported feeling lonely during the qualitative interview despite having UCLA-3 scores less than or equal to 40, compared to only three (10%) men. On the other hand, four (8%) women did not acknowledge feeling lonely on the qualitative interview despite having UCLA-3 scores greater than 40, compared to seven (24%) men (Fisher's exact test =0.02, p <0.05).

While there were no differences in response length by sex in the overall sample (Mann-Whitney U = 686.5, p = 0.30), quantitatively lonely men had longer responses compared to lonely women (Mann-Whitney U = 66.5, p = 0.03, Cliff's delta = −0.4). Men expressed more fear in their Q1 responses compared to women (overall sample: Mann-Whitney U = 559,

FIGURE 2. Distribution of length of response to Question 1 by quantitatively assessed loneliness. Q1: Question 1 ("Do you ever feel lonely and if so, how often?").
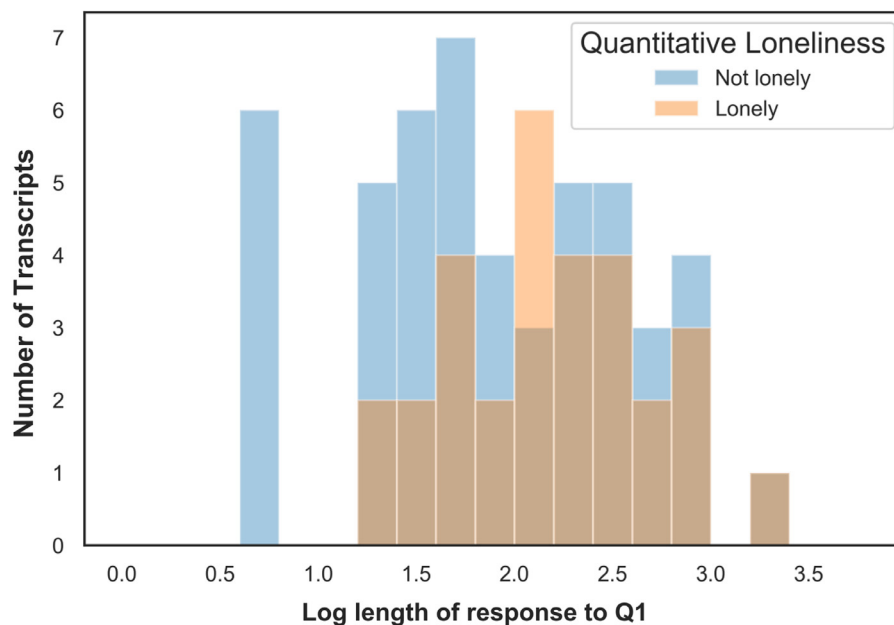
FIGURE 3. Emotional composition of response to Question 1 ("Do you ever feel lonely and if so, how often?") by (A) Qualitative loneliness and (B) Quantitative assessment of loneliness (UCLA-3 Score). Dashed lines in the middle of distribution indicate median (second quartile) and dotted lines indicates first and third quartiles in the distribution.
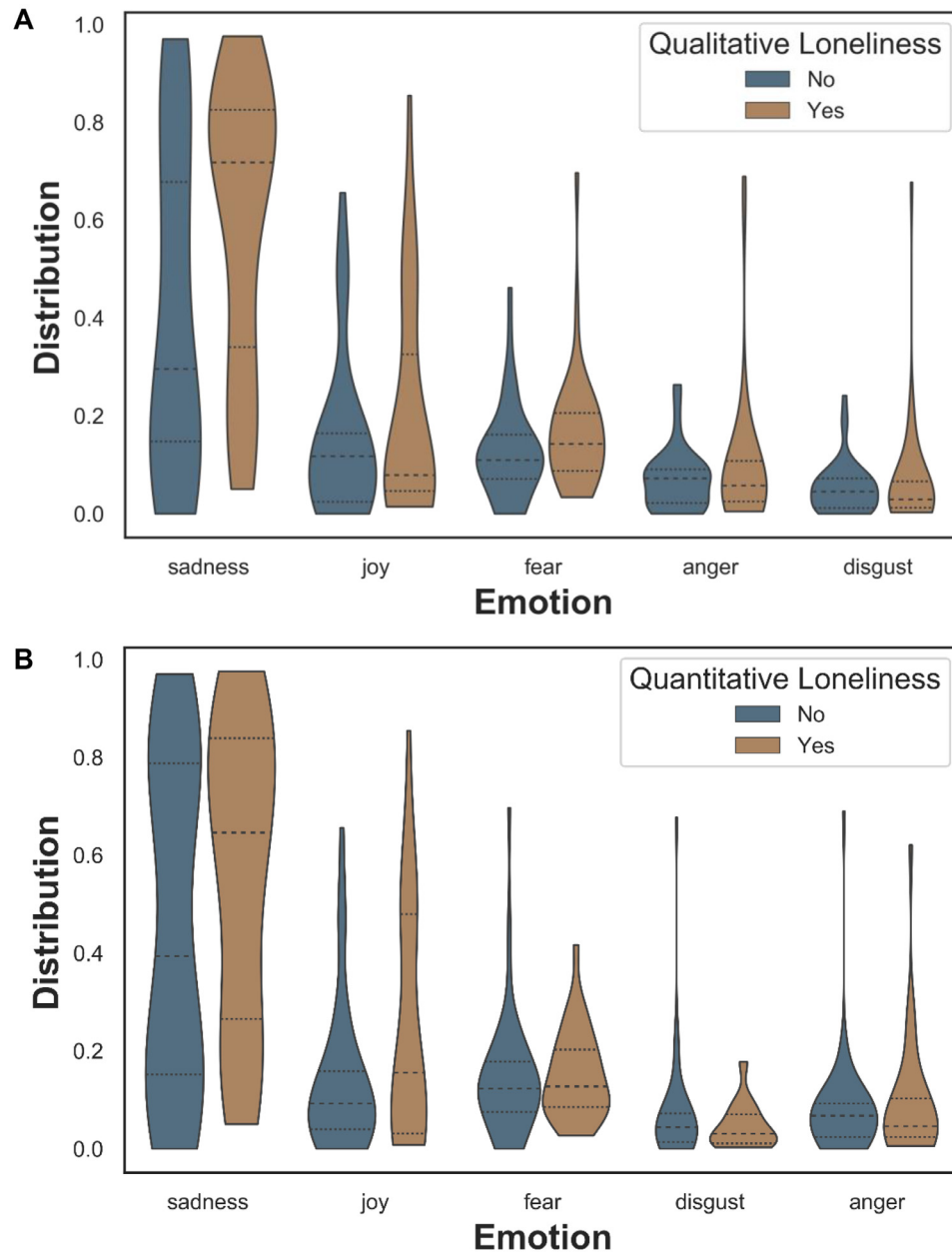
p = 0.04, Cliff's delta = −0.24). Lonely men expressed more joy than women (quantitatively lonely subsample: Mann-Whitney U = 70.0, p = 0.05, Cliff's delta = −0.37) (Fig. 4).

ML Models to Predict Loneliness

Qualitative loneliness (based on manually scored responses to Q1) on test data was best predicted by

FIGURE 4. Distribution of emotions (sadness, joy, fear, disgust, anger) in response to Question 1 ("Do you ever feel lonely and if so, how often?") by sex. Dashed lines in the middle of distribution indicate median (second quartile) and dotted lines indicate first and third quartiles in the distribution.
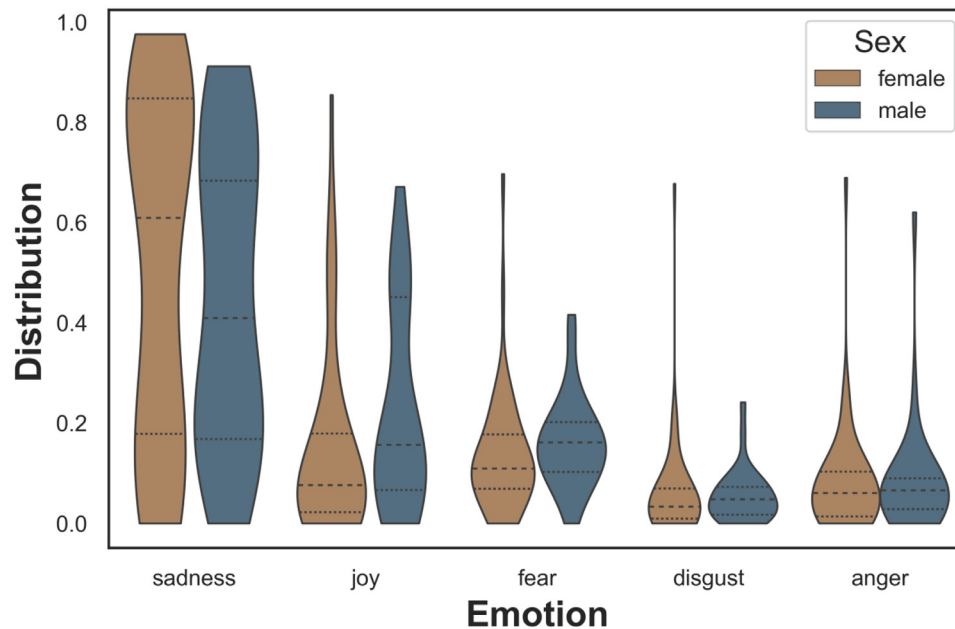

the kNN model (F1 score of 0.94 on test data) (Table 2, ROC curves in Supplemental Fig. 3A.)

Quantitative loneliness (based on UCLA-3 scores) was also best predicted by the ANN tanh model (F1 score of 0.74) (Table 3; ROC curves in Supplemental Fig. 3B).

Cross-validation using 5-Fold analysis on all data yielded: an F1 score of 0.86 for qualitative loneliness using ANN tanh model and 0.75 for quantitative loneliness (UCLA-3) using random forest model respectively.[42] The high F1 scores and area under the curve suggest data is well separated with a little overlap. Relative to other ML methods, the tanh activation function allows for faster learning for feature values close to 0 owing to its slope being maximum at 0. The Orange3 software provides readily available implementations of several ML models, that require simple configuration and connections using a visual programming tool. All the ML models used in the study were from Orange3 tool and how they were configured in a pipeline is provided as a separate file (nlp5_cutoff40.ows) and described in Supplemental materials (Appendix D).

TABLE 2. Performance of Machine Learning (ML) Models (80−20 Split) in Predicting Qualitative Loneliness (Lonely versus Not Lonely)

ML Model	AUC	F1[a]	Precision[a]	Recall[a]
kNN	**0.96**	**0.94**	**0.94**	**0.93**
Stack	0.91	0.87	0.90	0.87
SVM linear	0.95	0.87	0.90	0.87
ANN tanh	0.93	0.87	0.90	0.87
ANN ReLu	0.93	0.87	0.90	0.87
ANN Logistic	0.95	0.87	0.90	0.87
SVM RBF	0.91	0.81	0.82	0.81
SVM Polynomial	0.88	0.81	0.87	0.81
Random Forest	0.91	0.81	0.87	0.81
AdaBoost	0.80	0.74	0.85	0.75
Tree	0.71	0.69	0.74	0.68

Notes: Qualitative loneliness was manually determined based on responses to Question 1. Input features included: five emotions (joy, fear, anger, disgust, sadness), Question 1 TF-IDF score, Question 2 TF-IDF score, Question 3 TF-IDF score, and sentiment for Question 1. Results depicted reflect the best of 10 runs. Stack includes (SVM Polynomial, KNN, Tree, AdaBoost, ANN ReLu, random forest). AUC: area under curve (performance measure); kNN: K-nearest neighbor (algorithm), k = 9; ReLu: rectified linear unit (activation function); RBF: radial basis function (kernel function); SVM: support vector machine (algorithm); tanh: hyperbolic tangent (activation function); TF-IDF: term frequency − inverse document frequency.
Bold values indicate the best performing model.

[a] The performance measures shown are average over classes and computed as documented Orange.[42]

Prediction of Loneliness in Older Adults Using Natural Language Processing

TABLE 3. Performance of Machine Learning (ML) Models (80−20 Split) in Predicting Quantitative Loneliness (Lonely Versus Not Lonely)

ML Model	AUC	F1[a]	Precision[a]	Recall[a]
ANN tanh	**0.79**	**0.74**	**0.76**	**0.75**
Tree	0.69	0.68	0.69	0.68
Random Forest	0.59	0.62	0.62	0.62
AdaBoost	0.61	0.62	0.62	0.62
kNN	0.60	0.55	0.55	0.56
Stack	0.58	0.53	0.54	0.56
ANN Logistic	0.60	0.53	0.54	0.56
SVM RBF	0.65	0.53	0.77	0.62
SVM Polynomial	0.69	0.53	0.77	0.62
SVM Linear	0.53	0.53	0.77	0.62
ANN ReLu	0.65	0.44	0.44	0.50

Notes: Quantitative loneliness was determined by total score on the UCLA Loneliness Scale (version 3): ≤ 40 = No/Low Loneliness and >40 as Lonely. Input features included: five emotions (joy, fear, anger, disgust, sadness), Question 1 TF-IDF score, Question 2 TF-IDF score, Question 3 TF-IDF score, and sentiment for Question 1. Results depicted reflect the best of 10 runs. Stack includes (SVM polynomial, KNN, Tree, AdaBoost, ANN ReLu, random forest). AUC: area under curve (performance measure); kNN: K-nearest neighbour (algorithm); ReLu: rectified linear unit (activation function); RBF: radial basis function (kernel function); SVM: support vector machine (algorithm); tanh: hyperbolic tangent (activation function); TF-IDF: term frequency − inverse document frequency.
Bold values indicate the best performing model.
[a] The performance measures shown are average over classes and computed as documented Orange.[42]

Feature Ranking for Classification Tasks

Presence of Q3 in interview (indirectly) captures the expression of loneliness by interviewee and the choice of alternative questions by interviewer, making it the highest-ranking feature in both classification tasks (for both qualitative and quantitative loneliness) (Supplemental Tables 2 and 3). IBM sentiment (i.e., verbal agreement to Q1) ranks highly in qualitative loneliness classification, but not in the quantitative loneliness classification. Expressed emotions in the Q1 responses ranked comparably with the top feature (Q3) for quantitatively assessed loneliness, but not as highly for qualitative loneliness.

DISCUSSION

This study demonstrated the feasibility of using NLP analyses to examine transcribed speech data regarding loneliness. This work was a useful first step in understanding how to derive meaning from a large sample size of transcribed speech data, by traditional qualitative methods. We found that qualitatively lonely individuals had longer responses to direct questions about loneliness. Women were more likely to endorse loneliness during interviews when they were quantitatively lonely. Men were more likely to express fearful sentiment in their Q1 responses. ML models based on language features could predict the presence of loneliness (by both qualitative and qualitative assessments) with reasonable precision. ML models could predict qualitative loneliness with sensitivity (proportion of positives that were correctly identified) = 0.90, and specificity (proportion of negatives that were correctly identified) = 1.00. Quantitative loneliness could be predicted with sensitivity = 0.57 and specificity = 0.89.

To our knowledge, this is one of the first published NLP studies with both qualitative and quantitative assessments of loneliness among older adults. The agreement of qualitative and quantitative assessments was fair, and male sex appeared to underlie the discrepancies between self-reported and scale-based loneliness. Other studies reported discrepancies between responses to direct questions about loneliness compared to scores on the UCLA-3 among younger male participants, attributed to stigma of acknowledging loneliness.[51] Our findings were similar to these previous studies, with a larger proportion of older men who did not endorsing loneliness on interview despite having "lonely" UCLA-3 scores. Interpretation of participants' responses using NLP should account for key sociodemographic factors such as age and sex. Further investigation into understanding these responses on a deeper semantic and structural level is needed.

The exploratory analyses of sex differences also raised interesting foci for future investigations. Interestingly, male and female respondents had similar mean anxiety scores, depression scores, and measures of instrumental support and negative social interactions. Only emotional support scores differed − with women reporting more emotional supports, though this difference was not reflected by the loneliness assessments. Studies using the DeJong Giervald Loneliness scales specifically assess emotional loneliness (missing an intimate relationship) and social loneliness (missing a wider social network), and have reported that men are less emotionally lonely but more socially lonely than women.[52,53] Such nuances in the definition of loneliness may be important for future studies of sex differences in loneliness. While this study was limited to differences by sex assigned at birth, these differences may also reflect societal

gender stereotypes rather than the effects of biological sex. Such nuances in the definition of loneliness and gender roles may be important for future studies of sex/gender differences in loneliness.

There was an increased use of words of fearful sentiment in responses of men, both in the overall group as well as the subset of lonely individuals, though the effect sizes were small to medium. This finding contrasts a census-based Swedish study of older adults that reported lower levels of fear and loneliness among men, compared to women in response to direct loneliness questions.[54] However, this study sample had key sex differences (younger men, higher proportion of men living independently and with someone) that may have contributed to increased loneliness and fear in women. Also, it is unclear how personal experiences of loneliness relate to linguistic expressions of fear. Lastly, these sex-based findings must be considered in the context of the sample characteristics (older age and lower emotional support in the male participants). Due to a general lack of standardization and calibration in NLU tools, we must limit the claim to being of theoretical interest. While these findings require further exploration with a nuanced emotional analyses of text data and a larger sample size to understand how the emotional content of these responses may differ by sex and loneliness, this is an important first step to understand how individuals may respond when asked about loneliness.

The use of NLP methods to analyze subjective states like loneliness will require further study and refinement to understand the complex results and nature of loneliness. However, this proof-of-concept study demonstrates the value of incorporating a large number of perspectives in qualitative analyses. NLP and ML techniques can be scaled up to handle hundreds or thousands of interviews and can provide consistent ratings that may not be possible with human raters. The current study extends earlier qualitative work based on the traditional coding of 30 interviews.[16] The manual coding method, while time-consuming and labor-intensive, allowed for specific and sensitive interpretation of the respondents' risks for and experience of loneliness as well as their coping strategies. These results highlighted the importance of wisdom components (spirituality, emotional regulation, compassion) for preventing and coping with loneliness. However, the traditional approach could not capture perspectives of the full cohort, as was possible in this study, and was vulnerable to human error and bias, thus requiring parallel analyses by two independent reviewers.

In order to further extend and complement traditional qualitative approaches,[16] the current study's NLP approach can handle large datasets using semi-automated approaches, thus enabling future replication and subgroup analyses by sex. The NLP methods were able to quantify the expressed sentiment and emotion of the responses using a consistent algorithm. Through quantifying the text into specific features, the NLP methods were able to link transcript features to qualitative (interview-based) and quantitative (UCLA-3 score-based) loneliness and model the outcomes using ML. The current study illustrates how NLP methods provide an additional data-stream to combine with quantitative measures and create synergy with "higher order" themes identified by traditional qualitative methods.[55–57]

The current study identified kNN and ANN (with tanh activation) as the top-performing ML models for qualitative and quantitative loneliness classification tasks respectively. The outperformance of kNN model (F1 score 0.94) for qualitative classification suggests that samples of each class appear as clusters, possibly around an "archetype" for the class. There is little overlap between the two classes and the features used do indeed represent the inputs well.

Best performance on quantitative loneliness was achieved using ANN (with tanh activation). A relatively weaker performance compared to the one achieved for qualitative loneliness implies that such (UCLA-3) assessments capture information not readily available in the interviews and/or are sparsely represented in the features used for classification. Further, the classification boundary is a complex one (required the use of non-linear classification). ANN models, in general, outperform SVM models in a number of cases. Performance of SVM models rely on the structure of features and appropriate choice of kernels (algorithms). ANN are trained using "Backpropagation" or backward propagation of error, an efficient method to train the model. The function tanh has higher derivatives and it is 0 centered, which provides advantages to learning. While challenging to fully interpret the model, this finding reflects the complex, non-linear nature of how transcribed speech data reflects quantitative loneliness.

ML models had greater precision in predicting qualitative loneliness (kNN model F1 score of 0.94) based on linguistic features alone, compared to

quantitative loneliness (ANN tanh F1 score of 0.74). However, the most predictive feature was the presence or absence of follow-up questions from the standardized battery. Thus, analysis of interviews could be automated, especially when the interviews are well-designed. In comparison, a lower F1 score for quantitative assessment of loneliness from the same set of features may indicate how linguistic features may be more reflective of qualitative rather than quantitative loneliness. Thus, to better predict quantitative loneliness, other features and participant characteristics (e.g., baseline response length, specific fearful words used, highest achieved education, neuroticism) may need to be considered in future models. While it is not possible to draw definitive conclusions from the best-fit ML models, the current study demonstrated the feasibility of using ML models for "fuzzy" psychological constructs such as loneliness.

The study had several limitations. First, data were cross-sectional; thus, causal inference is not possible. Longitudinal studies are needed to understand the quality and trajectory of loneliness over time. Next, the sample size was too small to fully understand the potential of NLP in diagnosing loneliness. However, this proof-of-concept study serves to demonstrate how NLP of unstructured text data can be used in a deeply phenotyped sample. Second, the sample size was limited to residents within a San Diego housing community and thus, these findings may not be generalizable to other populations. Third, the qualitative and quantitative loneliness assessments may differ in timescale of loneliness. Q1 refers to "ever" feeling lonely, while the UCLA-3 does not inquire about a specific time period. Future studies should examine loneliness as a transient trait as well as a persistent trait. This sample included men who were on average older and had less emotional support than the women included in the study, which may confound the sex-based results. The current study focused on the potential of ML and NLP analyses to examine novel speech data. However, a thorough examination of all the ML models for this data was beyond the scope of the current paper. Future work should examine the nuances of the ML models in handling transcribed speech data. The analyses were not corrected for multiple comparisons due to their exploratory nature. Finally, NLU software methods and tools were developed to analyze conversational text data and were not developed specifically for clinical uses. For example, the five emotions used for the IBM NLU iv tool may not be best suited for understanding loneliness.

CONCLUSIONS

This proof-of-concept study demonstrates how text features can be used to predict loneliness. NLP and ML are effective and novel tools to analyze linguistic features of interview data for psychological constructs like loneliness. State-of-the-art sentiment and emotion analysis can provide insights into composition of a complex emotion (e.g., loneliness). Understanding sex differences in how older individuals discuss loneliness will be instrumental in detecting loneliness through text data. Future studies will need larger samples of diverse individuals, combined with other sensor data-streams (e.g., voice recordings, social interactions, GPS data, physical activity or sleep measures) to personalize the findings. Nuanced linguistic data will be key in developing future AI tools to detect loneliness among individuals based on their speech alone, enabling remote diagnosis of loneliness. Eventually, complex AI systems could intervene in real-time to help individuals to reduce their loneliness by adopting in positive cognitions, managing social anxiety, and engaging in meaningful social activities.

AUTHOR CONTRIBUTIONS

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.jagp.2020.09.009.

REFERENCES

1. Lubben J, Gironda M, Sabbath E, et al. *Social isolation presents a grand challenge for social work* (Grand Challenges for Social Work Initiative Working Paper No. 7), 2015, Cleveland, OH: American Academy of Social Work and Social Welfare.
2. Hawkley LC, Wroblewski K, Kaiser T, et al: Are U.S. older adults getting lonelier? Age, period, and cohort differences. Psychol Aging 2019; 34:1144–1157
3. Dahlberg L, Andersson L, McKee KJ, et al: Predictors of loneliness among older women and men in Sweden: a national longitudinal study. Aging Ment Health 2015; 19:409–417
4. United States Census Bureau: Changing American Households. Washington, D.C.: United States Census Bureau, 2011
5. Pew Research Center: Social Isolation and New Technology. Pew Internet & American Life Project, 2009
6. Holwerda TJ, Deeg DJ, Beekman AT, et al: Feelings of loneliness, but not social isolation, predict dementia onset: results from the Amsterdam Study of the Elderly (AMSTEL). J Neurol Neurosurg Psychiatry 2014; 85:135–142
7. Wilson RS, Krueger KR, Arnold SE, et al: Loneliness and risk of Alzheimer disease. Arch Gen Psychiatry 2007; 64:234–240
8. Boss L, Kang DH, Branson S: Loneliness and cognitive function in the older adult: a systematic review. Int Psychogeriatr 2015; 27:541–553
9. McDaid D, Bauer A, Park A-L: Making the economic case for investing in actions to prevent and/or tackle loneliness: a systematic review. A Briefing paper. Personal Social Services Research Unit (PSSRU). London: London School of Economics and Political Science, 2017
10. Lee EE, Depp C, Palmer BW, et al: High prevalence and adverse health effects of loneliness in community-dwelling adults across the lifespan: role of wisdom as a protective factor. Int Psychogeriatr 2019; 31:1447–1462
11. Luchetti M, Lee JH, Aschwanden D, et al: The trajectory of loneliness in response to COVID-19. Am Psychol 2020;doi:10.1037/amp0000690
12. Wong A, Chau AKC, Fang Y, et al: Illuminating the psychological experience of elderly loneliness from a societal perspective: a qualitative study of alienation between older people and society. Int J Environ Res Public Health 2017; 14:824–842
13. Sjöberg M, Edberg AK, Rasmussen BH, et al: Being acknowledged by others and bracketing negative thoughts and feelings: frail older people's narrations of how existential loneliness is eased. Int J Older People Nurs 2019; 14:e12213
14. Drageset J, Eide GE, Dysvik E, et al: Loneliness, loss, and social support among cognitively intact older people with cancer, living in nursing homes—a mixed-methods study. Clin Interv Aging 2015; 10:1529–1536
15. Neves BB, Sanders A, Kokanović R: "It's the worst bloody feeling in the world": experiences of loneliness and social isolation among older people living in care homes. J Aging Stud 2019; 49:74–84
16. Morlett Paredes A, Lee EE, Chik L, et al: Qualitative study of loneliness in a senior housing community: the importance of wisdom and other coping strategies. Aging Ment Health 2020: 1–8; doi:10.1080/13607863.2019.1699022
17. Hawkley LC, Cacioppo JT: Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010; 40:218–227
18. Pinquart M, Sorensen S: Gender differences in self-concept and psychological well-being in old age: a meta-analysis. J Gerontol Series B Psychol Sci Soc Sci 2001; 56:P195–P213
19. Manning CD, Manning CD, Schütze H: Foundations of Statistical Natural Language Processing 1999
20. Bird S, Klein E, Loper E: Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc., 2009
21. Rumshisky A, Ghassemi M, Naumann T, et al: Predicting early psychiatric readmission with natural language processing of narrative discharge summaries. Transl Psychiatry 2016; 6, e921-e921
22. Cook BL, Progovac AM, Chen P, et al: Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. Comput Math Methods Med 2016; 2016;doi:10.1155/2016/8708434
23. Fernandes AC, Dutta R, Velupillai S, et al: Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. Sci Rep 2018; 8:1–10
24. Kolliakou A, Bakolis I, Chandran D, et al: Mental health-related conversations on social media and crisis episodes: a time-series regression analysis. Sci Rep 2020; 10:1342
25. Tran T, Kavuluru R: Predicting mental conditions based on "history of present illness" in psychiatric notes with deep neural networks. J Biomed Inform 2017; 75:S138–S148

26. Perlis R, Iosifescu D, Castro V, et al: Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. Psychol Med 2012; 42:41–50

27. Jeste DV, Glorioso D, Lee EE, et al: Study of independent living residents of a continuing care senior housing community: socio-demographic and clinical associations of cognitive, physical, and mental health. Am J Geriatr Psychiatry 2019; 27:895–907

28. Derogatis LR, Melisaratos N: The brief symptom inventory: an introductory report. Psychol Med 1983; 13:595–605

29. Kroenke K, Spitzer RL, Williams JB: The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001; 16:606–613

30. Russell DW: UCLA loneliness scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996; 66:20–40

31. Doryab A, Villalba DK, Chikersal P, et al: Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and fitbit data. JMIR Mhealth Uhealth 2019; 7:e13209

32. Patton M: Qualitative Research and Evaluation Methods. 3d ed. Thousand Oaks, CA: Sage, 2002

33. IBM WNLU: Available at: https://www.ibm.com/cloud/watson-natural-language-understanding?lnk=STW_US_STESCH&lnk2=trial_WatNatLangUnd&pexp=def&psrc=none&mhsrc=ibmsearch_a&mhq=NLU. Accessed February 24, 2020

34. Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174

35. Joachims T: *A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization*, Pittsburgh, PA, School of Computer Science, Carnegie Mellon University, 1996

36. Aizawa A: An information-theoretic perspective of TF-IDF measures. Inf Process Manage 2003; 39:45–65

37. Rajaraman A, Ullman JD: Mining of Massive Datasets. Cambridge University Press, 2011

38. IBM Cloud API Docs, Natural Language Understanding, Available at: https://cloud.ibm.com/apidocs/natural-language-understanding/natural-language-understanding. Accessed March 4, 2020

39. Plutchik R: Emotions and life: perspectives from psychology, biology, and evolution. *American Psychological Association, Washington, DC,* 2003

40. IBM Watson: Sentiment and context analysis, Available at: https://www.pubnub.com/docs/blocks-catalog/group-sentiment-analysis. Accessed December 19, 2019

41. Ernst N: Cliff's delta, python. Available at: https://github.com/neilernst/cliffsDelta, Accessed August 25, 2020

42. Demsar J CT, Erjavec A, Gorup C, et al: Orange: data mining toolbox in Python. J Mach Learn Res 2013; 14:2349–2353

43. Opitz D, Maclin R: Popular ensemble methods: an empirical study. J Artif Intell Res 1999; 11:169–198

44. Polikar R: Ensemble based systems in decision making. IEEE Circuits Syst Mag 2006; 6:21–45

45. Bolón-Canedo V, Sánchez-Maroño N, Alonso-Betanzos A: A Critical Review of Feature Selection Methods, in Feature Selection for High-Dimensional Data. Cham: Springer International Publishing, 2015:29–60

46. Remeseiro B, Bolon-Canedo V: A review of feature selection methods in medical applications. Comput Biol Med 2019; 112:103375

47. Saeys Y, Inza I, Larrañaga P: A review of feature selection techniques in bioinformatics. Bioinformatics 2007; 23:2507–2517

48. Rokach L: Ensemble-based classifiers. Artif Intell Rev 2010; 33:1–39

49. Sollich P, Krogh A: Learning with ensembles: how overfitting can be useful, 1996

50. Adeva JJG, Beresi U, Calvo R: Accuracy and diversity in ensembles of text categorisers. CLEI Electron J 2005; 9:1–12

51. Nicolaisen M, Thorsen K: Who are lonely? Loneliness in different age groups (18-81 years old), using two measures of loneliness. Int J Aging Human Dev 2014; 78:229–257

52. De Jong Gierveld J, Van Tilburg T: The De Jong Gierveld short scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. Eur J Ageing 2010; 7:121–130

53. Dykstra PA, de Jong Gierveld J: Gender and marital-history differences in emotional and social loneliness among Dutch older adults. Can J Aging Revue Canadienne Vieillissement 2004; 23:141–155

54. Jakobsson U, Hallberg IR: Loneliness, fear, and quality of life among elderly in Sweden: a gender perspective. Aging Clin Exp Res 2005; 17:494–501

55. Guetterman TC, Chang T, DeJonckheere M, et al: Augmenting qualitative text analysis with natural language processing: methodological study. J Med Internet Res 2018; 20, e231-e231

56. Leeson W, Resnick A, Alexander D, et al: Natural Language Processing (NLP) in qualitative public health research: a proof of concept study. Int J Qual Methods 2019; 18:160940691988702

57. Crowston K, Allen EE, Heckman R: Using natural language processing technology for qualitative data analysis. Int J Soc Res Methodol 2012; 15:523–543