Social Sensing: Assessing Social Functioning of Patients Living with Schizophrenia using Mobile Phone Sensing

Weichen Wang, Shayan Mirjafari, Gabriella Harari[†], Dror Ben-Zeev[§], Rachel Brian[§], Tanzeem Choudhury[⋆], Marta Hauser[‖], John Kane[‖], Kizito Masaba, Subigya Nepal, Akane Sano[¶], Emily Scherer, Vincent Tseng[⋆], Rui Wang, Hongyi Wen[⋆], Jialing Wu, Andrew Campbell

Dartmouth College	[⋆]Cornell University	[‖]Northwell Health
Hanover, NH, USA	Ithaca, NY, USA	Glen Oaks, NY, USA
{wang, shayan, ruiwang, campbell}@cs.dartmouth.edu	{wt262, hw557, tanzeem.choudhury}@cornell.edu	{jkane2, mhauster}@northwell.edu
[¶]Rice University	[†]Stanford University	[§]University of Washington
Huston, TX, USA	Stanford, CA, USA	Seattle, WA, USA
akane.sano@rice.edu	gharari@stanford.edu	{dbenzeev, rbrian}@uw.edu

ABSTRACT
Impaired social functioning is a symptom of mental illness (e.g., depression, schizophrenia) and a wide range of other conditions (e.g., cognitive decline in the elderly, dementia). Today, assessing social functioning relies on subjective evaluations and self assessments. We propose a different approach and collect detailed social functioning measures and objective mobile sensing data from N=55 outpatients living with schizophrenia to study new methods of passively accessing social functioning. We identify a number of behavioral patterns from sensing data, and discuss important correlations between social function sub-scales and mobile sensing features. We show we can accurately predict the social functioning of outpatients in our study including the following sub-scales: prosocial activities (MAE = 7.79, r = 0.53), which indicates engagement in common social activities; interpersonal behavior (MAE = 3.39, r = 0.57), which represents the number of friends and quality of communications; and employment/occupation (MAE = 2.17, r = 0.62), which relates to engagement in productive employment or a structured program of daily activity. Our work on automatically inferring social functioning opens the way to new forms of assessment and intervention across a number of areas including mental health and aging in place.

Author Keywords
Social Functioning; Mobile Sensing; Social Sensing; Health

CCS Concepts
•**Applied computing → Life and medical sciences;**

INTRODUCTION
Social functioning describes how well an individual is able to

interact with their environment and fulfill key skills associated with the social roles they hold within the environment, such as in the domains of social activities, work, and family relationships [13, 72, 91, 88, 76, 75, 51]. As such, social functioning is a key factor of interest in many research areas aimed at understanding and assessing mental health. For example, social functioning has been a key factor in understanding the experiences of individuals suffering from schizophrenia-spectrum disorders [48, 40], depression [13, 101], and personality disorders [73]. Beyond the domain of mental disorders, deterioration in social functioning is also associated with the aging process because elderly people can also experience a reduced capacity and vulnerability to stress [102, 11, 59, 68] that can impact their ability to successfully interact with their environment and perform their social roles on a day-to-day basis. Given the significance of social functioning as a key factor in overall well-being, there is growing interest in developing passive assessments for monitoring social behaviors, measuring the results of therapy, and designing accurate interventions. However, the standard practice for assessing social functioning in clinical and non-clinical settings to date are typically based on self-reports (e.g., survey responses) or face-to-face evaluations (e.g., interviews). While informative, such approaches are often costly, labor-intensive, and suffer from limited ecological validity [92]. What is needed are new approaches to passively and continuously assessing social functioning in the context of daily life so that timely interventions can be designed. Inspired by this need and other emerging studies on how mobile technologies can be used to understand and assess mental health [18, 5, 96, 62, 80, 79, 9, 7, 99, 29], in this paper we ask: *(1) What behavioral patterns are associated with social functioning in daily life? (2) Can we predict social functioning through passive behavioral sensing assessments derived from people's smartphones?*

To answer these questions, we collected a rich and real-world dataset of social functioning measures and multimodal mobile sensing data from people living with schizophrenia, who were outpatients at a large psychiatric hospital in New York

City. Impairment in social functioning is a central feature of schizophrenia symptoms that is present in most patients [47, 31, 27]. The importance of social functioning in the assessment of schizophrenia has been acknowledged in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5 [4]); the standard classification system for mental disorders used by mental health professionals in the United States. People who have been clinically diagnosed with schizophrenia tend to vacillate between periods of remission and episodes of symptom exacerbation [87], which often triggers shifts in behavioral patterns. These symptom features require the collection of a high-quality longitudinal dataset of social functioning in daily life to capture the considerable variability and change in the behavioral patterns.

The mental health sensing study, CrossCheck, was first introduced in [8]. To date, researchers have leveraged the dataset for innovative analyses focused on assessing problematic symptoms that indicate relapse in schizophrenic patients. For example, Wang et al. [95] developed inference models based on sensing data to predict symptoms that were self-reported from patients (e.g., seeing things, hearing voices, feeling depressed) with a mean error of 7.6% of the score range. In [97], passive sensing data was used to predict Brief Psychiatry Rating Scale (BPRS) scores, which is a 7-item measure that describes symptoms directly associated with schizophrenia that are provided by clinicians (not the patients themselves) with ± 1.59 MAE. However, these studies were narrowly focused on monitoring the psychotic symptoms themselves (e.g., delusions, hallucinations, disorganized speech, catatonic behavior [17, 4]) to identify whether an individual was at an increasing risk of schizophrenic relapse [8]. But the findings from this previous work are less sensitive to understanding the general social functioning of these patients.

In this paper, we focus on understanding and assessing the everyday behavioral patterns of schizophrenic patients in order to predict their general social functioning. Differentiated from the work above that focuses on schizophrenic symptoms or risk of relapse, we specifically investigate the social functioning, a more generalized area of broader interest to the HCI community. We combine our year-long mobile phone sensing data collected in the wild with detailed social functioning measurements clinically administered every 3 months. Social functioning is measured by the *Social Functioning Scale (SFS)* [12] questionnaire, which rates social interactions, interpersonal relationships, and activities of independent living in 7 subareas: *social engagement/withdrawal*, *interpersonal behavior*, *prosocial activities*, *recreation*, *independence-competence*, *independence-performance* and *employment/occupation*. We find the impairments and disabilities in each detailed area of social functioning are associated with different behavioral signals captured by smartphones. Our approach is an instrumental step toward designing timely interventions that can be deployed in the context of people's natural lives, and can lead to behavioral insights that generalize to other populations with mental health issues or other social functioning difficulties.

The structure of the paper is as follows. We start by discussing the related work on social functioning and then detail our study,

dataset, models, methods, results and insights. We also discuss the privacy and ethical implications surrounding the collection of sensitive data from populations at risk – we make some concluding remarks on this important topic in the discussion section. Our results show mobile phones can accurately predict social functioning impairment among the people living with schizophrenia in our study. Our results on sensing social functioning open the way to new forms of assessment and intervention across a number of areas important to the HCI community, such as, mental health, aging in place, cognitive decline and dementia.

Privacy, Ethics and Disclosure

While it is clear that clinicians and patients could benefit from passive sensing, there are risks associated with collecting sensitive data. Considering the broader concern about ethics and privacy in this field, the misuse of such technologies could cause serious privacy issues. In what follows, we discuss how such sensitive sensed data and inferences should be protected.

First, researchers must make sure that participants knowingly consent to give up some of their privacy during the study. We explicitly stated the risks of participation in our consent forms, and discussed this in detail while enrolling participants. During enrollment, participants also had the opportunity to ask questions and were given supportive materials including visuals outlining all sensors and information collected by the app. Prior to obtaining consent, we also tested participants' knowledge of the type of data collected, whether the study was confidential, not anonymous, and ways in which we protect their data. Participants were administered a competency screener to verify that they understood the details and were able to provide informed consent. They were not allowed to participate if they could not pass this test.

We have taken great effort to protect the privacy of our participants. We took extra precautions with this vulnerable population when developing the study, software and refined the data we collected down to only the essential identifiable sensors and data captured for this study. Almost all the data captured by the app was unidentifiable. For example, we developed unique processing (conversation detecting algorithms) on the device itself to capture the presence of socialization, omitting the need to capture any actual spoken word. The voice detection algorithm only detects whether the person is around speech, and does not collect the content of conversations, only the frequency and duration of conversations. We also captured light sensing, accelerometer, and the number of calls and texts a participant sends during the course of a day; each unidentifiable, and if disclosed, would have little consequence to the participant. We did not collect other types of sensitive information (e.g., web-browsing, keyboard strokes, social media content, friends/family's phone numbers, pictures, videos). In addition, all the sensed data is decoupled from demographic data and is associated with a random study ID.

RELATED WORK

HCI researchers have examined ways of using passively collected sensing data to infer individual and community behaviors and mental states in different contexts [81, 50, 53, 71].

For example, Alharbi et al. [3] designed WillSense, a wearable device to collect fine-grain eating behaviors in the wild among people with obese, overweight, and normal BMI. Li et al. [63] utilized IDSense tags for unobtrusive human object interaction detection that enables inferring daily activities at home. Sun et al. [89] used MoveMeant, a location awareness app to monitor local community ties and support local community building. HCI researchers have also leveraged naturalistic content from online communities to address broad topics including self-disclosure [66, 104], behavioral prediction [22, 23], community norms [21] and mental illness [35, 38].

Mental health is one of the common domains underlying these efforts. In recent years, there has been increasingly influential work on mobile sensing for mental health [18, 5, 96, 62, 80, 79, 9, 7]. In [77], the authors used an early mobile sensing platform device [26] to show how conversation and physical activity can be used to infer mental and social well-being among elderly adults living in a continuing care retirement community. The StudentLife study [96] demonstrated that depression, stress, loneliness, and flourishing are associated with passive sensing behaviors (e.g., conversation, sleep, activity, co-location). Seab et al. [80] investigated that location features extracted from GPS data including circadian movement, normalized entropy, and location variance, and phone usage features including usage duration and usage frequency are associated with depressive symptoms. The findings from their initial study were replicated [79] using the StudentLife dataset [96]. Canzian et al [18] developed an extended set of mobility features over [80, 79] to show how location data is related to depression [86, 61, 60]. Wang et al. [98] proposed a set of symptom features derived from phone and wearable sensors that proxy the DSM-5 defined depression symptoms specifically designed for college students. Sarker et al. [81] propose a pattern mining method to detect significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. Researchers also develop remote monitoring tools for clinicians or self-care interventions for mental health. Schroeder et al. [83] contribute a conversational mobile web app to help people with complex disorders maintain positive relationships, and control their emotions. The intersection between ubiquitous computing and sensing, social media and emerging technologies offers promising avenues for novel human-centered designs in enhancing mental wellbeing [94].

Social functioning impairment in schizophrenia [17] is studied widely by several authors. The studies range from understanding human social cognition which is highly affected by schizophrenia to studying how it is impaired by the illness. The work in [12] discusses new methods for quantifying social functioning impairment in schizophrenia using the SFS survey to assess social functioning. Other researchers [6] show that social cognitive impairments can be mitigated with the help of family members of patients with schizophrenia by being conscious of their expressed emotion. HCI researchers have studied social functioning impairment across a wider range of conditions; for example, aging, cognitive decline and dementia [102, 11, 59, 68, 102, 65, 19], as well as hearing impairment [56] and post-traumatic stress disorder [37]. HCI researchers have designed or evaluated innovative systems that offer so-cial support and improve social interaction. Adams et al. [1] examine the health applications of staccato social support in mobile environments designed for brief, rapid social sharing and interaction. Wiley et al. [102] design Message Center, a home-based communication solution for enhancing elder communication. Laput et al. [65] propose the StoryCubes system that creates mutual understanding and appreciation between independent living residents through the experience of telling and listening to stories.

DATASET

The CrossCheck study [95] is a randomized control trial (RCT) [20] conducted in collaboration with a large psychiatric hospital, Zucker Hillside Hospital, in Long Island, New York[1]. The whole study lasted for 4 years and recruited 150 participants for 12 months using rolling enrollment. The participants are randomized into either the smartphone arm (n=75) or treatment-as-usual arm (n=75). Potential study candidates are identified according to the hospital's electronic medical records [8, 95]. Participants receive Samsung Galaxy S5 Android phone equipped with the CrossCheck app and receive a tutorial on how to use the phone. They are asked to keep the phone turned on, to carry it with them as they go about their day, charge it close to where they sleep at night and also answer EMA every 2-3 days. The research staff check the daily report of sensed data and would call noncompliant patients to assist and get them back on track[2]. Due to these efforts, the sensing data shows good compliance except for the cases when the participants had relapses and were hospitalized with the phones being off. We exclude these days during the analysis (thus all the aggregated values discussed in the paper are normalized by the number of valid data during a time frame). In this paper, 55 participants in the smartphone arm who were complying with the study and social functioning assessment are involved in the analysis. Table 1 shows demographics of 55 participants. The demographics in our sample is also reflective in the ratio reported in other research [15].

Table 1: Demographics of participants

race	count	percent
Asian	1	1.82%
Native Hawaiian or Other Pacific Islander	3	5.45%
Black or African American	17	30.91%
White	19	34.55%
More than one race	13	23.64%
Unknown or Not Reported	2	3.64%

Mobile Sensing Data

The CrossCheck app collected a wide range of behavioral passive sensing data from the phone. The details of system design

---

[1]This study was approved by the Committee for Protection of Human Subjects at Dartmouth College and Institutional Review Board at North Shore-Long Island Jewish Health System.

[2]To ensure the acquired data has a broad coverage of behaviors, participants' personal phone numbers are migrated to the new phone and they are provided with an unlimited data plan for data uploading. In addition, to encourage the adherence, participants are compensated $20 every 3 months during the assessment based on participation and the length they were in the study. We also give them the phone after the completion of the study.

have been discussed in the prior work [95, 97]. More specifically, it captured physical activities [42], locations [46], ambient sound levels [45], voice/noise labels [77], number of calls and text messages [43], application usage, screen lock/unlock, and ambient light intensity [44]. We compute features from the passive sensing data on a daily basis, which describe participant's behaviors (e.g., duration of different physical activities in a day, conversation duration and frequency, different types of places visited, app usage).

Social Functioning Assessment

A clinical assessor administers each participant's symptom severity, depression, and social functioning in person every 3 months during the year-long study. As discussed earlier we use the SFS survey [12] to measure the social functioning of outpatients in our study. The questions of SFS are informed by the Disability Assessment Schedule [55] and previously successful psychosocial intervention programs. SFS is shown to be valid, reliable, and sensitive to assessing a range of social functioning impairment [12]. As discussed SFS consists of 7 sub-scales: 1) *social engagement/withdrawal*, which measures time spent alone, initiation of conversation and social avoidance; 2) *interpersonal behavior*, which captures number of friends/having a romantic partner and quality of communication; 3) *prosocial activities*, which assesses the engagement in a range of social activities (e.g., sport); 4) *recreation*, which gauges the engagement in a range of hobbies, interests, and pastimes; 5) *independence-competence*, which inquires about the ability to perform skills necessary for independent living; 6) *independence-performance*, which rates the performance of skills necessary for independent living; and finally 7) *employment/occupation*, which relates to the engagement in productive employment or a structured program of daily activity. A higher sub-scale score indicates better social functioning.

Table 2: Statistics of Social Functioning Scale

SFS sub-scale (scale range)	response range	overall mean(std)	within-person range*
social engagement (0-15)	2-15	11(2.5)	3(2-5)
interpersonal behavior (0-30)	4-29	20.8(4.9)	5(3-8)
prosocial activities (0-66)	0-61	20(12.2)	12(8-17)
recreation (0-48)	0-39	20.6(8.5)	10(5-16)
independence-competence (0-39)	19-39	36.7(3.4)	3(1-5)
independence-performance (0-39)	7-39	28.3(6.7)	7(5-10)
employment (0-10)	0-10	5.3(3.4)	2(1-4)

* Values in median(LQ-UQ) where LQ and UQ means the lower quartile (25%) and upper quartile (75%) respectively.

Table 2 shows the range of the sub-scales, the range of the scored responses from our study participants, the overall mean and standard deviation and the median, lower and upper quartile of the ranges of within-person changes among the multiple assessments during the year. Many participants experienced varying levels of social functioning during the study, which consents to other longitudinal studies in social functioning [93, 57, 30, 2]. We also observe that a participant may score higher in certain sub-scales and lower in others. For example, one participant has higher scores in social engagement and interpersonal behavior but lower scores in prosocial activities.

MODEL AND METHOD

Behavioral Features

We incorporate features that describe the six aspects of daily behaviors from the smartphone passive sensing data. Specifically, we compute the features of 8 categories listed in Table 3, including *physical activities*, *mobility*, *sleep patterns*, *ambient environmental context*, *face-to-face conversations*, *smartphone-based communications*, *smartphone usage* and *semantic location*. Features are computed on a daily basis and broken down into four epochs of the day: *morning (6am-12pm)*, *afternoon (12pm-6pm)*, *evening (6pm-12am)* and *night (12am-6am)*, that allow us to model people's behaviors during different parts of the day. Note that we compute new features for social functioning that has not been well-studied before. For example, *smartphone-based communications* features now include the usage of apps for voice or video calls (e.g., Skype and Hangouts); the time on different types of apps, e.g., social networking (Facebook, etc.), game, browser, entertainment (music/audio/video, etc.), and engagement (finance/tools/education/business, etc.) is also well investigated. Importantly, we consider *semantic location* features. SFS contains questions about activities at different types of places. For example, the "Social Engagement/Withdrawal" section asks "do you leave the house"; the "Independence-Performance" section asks about "buying everyday items from stores"; and the "prosocial" section asks about "visiting art gallery/museum". Motivated by the questions, we aim to assign semantics to places where participants visit. Specifically, we consider the following places: home, food, travel, art & entertainment, nightlife, education, park & outdoors, library, shop, gym, medical and residence. We compute the time spent at these places during a certain time frame.

Table 3: Features computed from mobile sensing data

category	details*
physical activities	duration on foot / in vehicle / on bicycle / sedentary
mobility	# of locations visited, distance travelled
sleep patterns [25]	sleep duration, sleep start time, and sleep end time
ambient env. context	amplitude of ambient sound, level of ambient light
f2f conversations	# of conversations, duration of conversations
smartphone-based communications	# in-coming phone calls, # of out-going phone calls, # of in-coming SMS, # of out-going SMS, duration of usage of apps for voice or video calls
smartphone usage	# of lock/unlocks, unlocked duration, duration of using different genres of apps: social networking, game, browser, entertainment and engagement.
semantic location	duration @ home, food, travel, art&entertainment, nightlife, education, parks&outdoors, library, shop, gym, medical and residence.

* We have 115 behavioral features in total. The length of each feature is 160, corresponding to the number of SFS responses adopted in our dataset. The unit could be in seconds (e.g., duration of various activities, conversations, etc.), an integer of counts (number of places visited, number of lock/unlocks, etc.), or ambient light level/ambient sound amplitude directly from sensors.

To label locations with semantics, we first cluster the raw GPS coordinates using density-based spatial clustering of applications with noise (DBSCAN) [36]. The centroid of each cluster is considered a significant location, where a participant dwells for a significant amount of time. We first label a significant

location as the home where a participant spends most of the time and typically frequents between 2 am to 6 am. We use the Foursquare API [39] to label the other significant locations. It takes a GPS coordinate and a radius as input and returns a list of location entities. Each location entity is associated with a name, coordinates, and categories (e.g., food, art&entertainment). We test the sensitivity based on different radiuses: 50m, 30m and 20m. Overall, 50m outperforms the rest in predicting the social functioning scores. Therefore, we use the semantic features from a radius of 50m in the analysis. A location may be associated with multiple different categories. We compute the dwell duration at a location for all associated location categories. For instance, if the API returns "food" and "art&entertainment" categories for a given significant location, we include the dwell duration at this significant location to both "food" and "art&entertainment" places. This is an approximation given that there is an error in location coordinates. We do not simply select a single returned location entity closest to the significant location coordinate because as mentioned GPS data is noisy [34] and the same building could be associated with different semantics.

Association Analysis
In order to understand how the passive sensing data is associated with SFS, we use principal components analysis (PCA) [82] and generalized estimating equations (GEE) [64, 52] to explore the associations between the principal components of behavioral features computed from passive sensing data and SFS sub-scale scores. Specifically, we look at the sensing data within 90 days prior to the SFS response. In what follows, we discuss our association analysis method in detail.

Principal Components Analysis. We compute the mean for each daily feature (e.g., mean of daily conversation duration in the 90-day time frame) per participant. After computing the features, we organize the features in a matrix M, where a column represents one of the 115 features, and a row represents features that are associated with one of the 160 SFS responses. We apply PCA on M to find a small number of principal components (PCs) that explain most of the variance in the feature matrix M. Using PC scores in our analysis presents several advantages over using behavioral features. First, selecting a smaller number of PCs significantly reduces the feature dimensionality. Second, each PC can be interpreted as a behavioral pattern. For example, if a PC has large positive weight in the
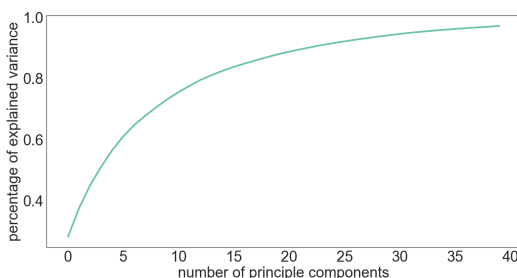

Figure 1: The cumulative variability in the data explained by the top-n principal components. We include the top 23 principal components that explain 90% of the variance in our analysis.

component for phone unlock duration and phone call duration features, and large negative weight for still duration, we would interpret this PC represents a high phone usage and high sedentary behavioral pattern. The absolute value of the PC score represents how significant this pattern is to a row in M (i.e., a participant's behavioral patterns within the 90 days prior to the SFS response). Figure 1 shows the cumulative variability in the data explained by the top-n principal components. We select the top 23 PCs in M that explain 90% of the variance for our analysis.

Bivariate GEE. We apply bivariate GEE [64, 52][3] to investigate how the selected 23 PCs are related to SFS scores by regressing the PC scores to SFS sub-scale scores. The SFS data is longitudinal: we receive multiple responses from one participant at different time in the year long RCT study. The responses from the same subject are dependent. We apply GEE to the combinations of 23 PCs and 7 SFS scores. To determine the false discovery rate (FDR), we perform the two-stage Benjamini-Hochberg procedure (TSBH) [10][4].

Predicting Social Functioning Scale
We experiment with different methods to find the best predictive model for social functioning scale. Specifically, we experiment with different time windows, different dimensionality reduction methods and different machine learning models. In what follows, we discuss our method in detail.

Prediction Time Window. Social Functioning Scale asks a respondent's experience over the last 3 months. Therefore, it is intuitive to use the sensing data within 90 days prior to the response to predict the SFS scores. However, we suspect participants' responses may subject to recall bias such that the responses may be more accurate to describe the participants' recent behaviors. Therefore, we also examine the predicting performance using shorter time windows. We assess the prediction performance with 90-day, 60-day, 30-day, and 15-day time windows.

Feature Transformation and Dimensionality Reduction. We experiment three different approaches for feature transformation and dimensionality reduction: PCA [103], truncated SVD [49], and Kernel PCA [82] with radial basis function kernel. We assess the prediction performance of using different number of principal components explaining 70%, 80%, 90%, and 95% of the variance for different time windows.

---

[3]The GEE method is designed to analyze longitudinal data. It is an extension of Generalized Linear models (GLM) [70] to model the regression and within-subject correlation separately. The GEE approach fits marginal models, therefore is commonly used to estimate population-averaged effects. Suppose Y_{it} represents the response from participant i at the tth measurement, the bivariate GEE model is usually described as $g(\mu_{it}) = \beta_0 + \beta x_{it}$, where $\mu_{it} = E(Y_{it})$ denotes the expectation of the response for subject i at time t, x_{it} is the value of feature for participant i at time t from sensing data, g indicates a link function, β_0 is the intercept, and β denotes the association between the sensing feature and the response. The p-value associated with β shows the probability of the coefficient being zero.

[4]The TSBH first examines the distribution of p-values to estimate the fraction of the null hypotheses that are actually true. It then uses this information to decide when a p-value is low enough to be called a discovery. This presents a rigorous approach to the analysis.

Evaluation of Machine Learning Models. We evaluate the prediction performance of different time windows, dimensionality reduction methods, and prediction models. We evaluate the prediction performance of linear regression (Lasso [90]/Ridge [54]/ElasticNet [105]), linear support vector regression (LinearSVR) [33], support vector regression with the radial basis function kernel (SVR_rbf) [85], extremely randomized trees (ET, Extra Trees) [41], random forest (RF) [14], and XGBoost [24].

We first apply 5-fold cross-validation to evaluate the prediction performance of a time window, dimensionality reduction method, and predictive model combination. We first iteratively split the data into 80% training and 20% hold-out testing. We then use the training fold to tune model hyper-parameters (e.g., regularization strength, learning rate, depth of regression trees, number of estimators) which maximizes prediction performance using 5-fold cross-validation. We train the predictive model with the best hyper-parameter using all the training data and test on the test set. We report test mean absolute error (MAE), root-mean-square error (RMSE) and Pearson correlation between predicted values and the SFS ground truth.

During the training process, in some folds, the data from the same subject could be split to both training and testing groups while the data of other subjects might only exist in either training or testing group. Therefore, we consider these mixed settings as where some of the patients may have been assessed multiple times while others may not. For example, there are some patients who were hospitalized and had been assessed, however, clinicians must still keep track of their status even after they are discharged. Nevertheless, there are also other patients who do not have any SFS records in the system, and clinicians need to get an initial sense of their status. The 5-fold cross-validation training process reports averaged performance metrics of models considering these circumstances.

To better understand the robustness of the system, we further perform the leave-one-subject-out (LOSO) validation. In the LOSO, models are trained on the data from other study participants with one subject's data left out. This emulates a new patient without any historical SFS records. In addition, we perform a customized leave-one-subject-out (CLOSO), which uses the first data point (in chronological order) of one specific patient combined with all the data points of other patients to predict the responses of that patient in latter assessments. This corresponds to a situation (which happens mostly in real practice) that patients have established their first medical record in the hospital and will rely on the app to track the social functioning afterwards.

RESULTS
In what follows, we present results for association analysis and prediction of social functioning.

Association Analysis
We investigate the correlation between the PC scores associated with 23 PCs and the SFS sub-scale scores using GEE and FDR as discussed earlier. We label each principal component by investigating its loadings (i.e., the weights associated with features). We rely on the highly ranked behavioral features

(i.e., features with larger absolute weight in the component) to understand the typical behaviors that make up each principal component. Our results show that 10 of the 23 behavioral patterns (i.e., principal components) are associated with various social functioning sub-scales. We list the 10 behavioral patterns and their corresponding features in Table 4, where the features in each pattern are ordered by the absolute value of the weight.

Behavioral Patterns. The behavioral patterns are listed in Table 4 in detail. It shows the behavioral characteristics of each pattern indexed by the n-th component of PCA (e.g., Pattern 1 where 1 is the first component of the PCA). The listed patterns are associated with social functioning sub-scales. Specifically, Pattern 1 describes people who often talk on the phone; Pattern 3 describes people who travel longer distance during evening and night (often using vehicle); Pattern 4 describes people who communicate with others using SMS rather than face to face social interactions captured by the audio sensor periods; Pattern 7 suggests people who are more active on foot, spend more time in brighter places in the mornings and afternoons periods, and have less exposure to face to face conversation interaction – more socially isolated; pattern 9 describes people who do more exercise on bikes; Pattern 11 is associated with people who like to Spend time in educational settings and sleep late and be gamingless; Pattern 14 describes individuals who show up frequently in libraries, show up less in residential areas (e.g., places which may be the homes of their friends) but not in the residential area where they live; Pattern 17 describes people who tend to rarely visit the gym, go to bed early, rarely make long phone calls during the night, spend less time at home, use more communication apps (e.g., Hangout, Skype, Messenger), and use less social networking apps (e.g., Facebook; pattern 21 suggests people who are mostly engaged in playing games on the phones.

Associations between behavioral patterns and social functioning sub-scales. Table 5 shows the behavioral patterns that are associated with the 7 social functioning sub-scales. A higher sub-scale score indicates better abilities and performance in a social functioning domain. A positive association between a pattern and a sub-scale suggests people whose behavior matches such pattern are more likely to have a higher sub-scale score, thus higher ability. Conversely, a negative association suggests people whose behavior matches such pattern are more likely to have a lower score (i.e., more likely to have impairment). In what follows, we describe our findings.

Social engagement sub-scale measures time spent alone, initiation of conversation, and social avoidance tendencies. It asks questions about the time one gets up, the hours one spends alone, the frequency one starts a conversation at home, the frequency one leaves home, and how one reacts to the presence of strangers. We find the social engagement sub-scale is negatively associated with pattern 7 (being physically active outside during the day without others around) and positively associated with pattern 17 (spending less time in the gym and going to bed early).

Interpersonal behavior sub-scale asks about the number of friends, the quality of the relationship, and communications

Table 4: Characteristics of patterns associated with social functioning indexed by n-th component of PCA

pattern	features
1 (Using the phone frequently for phone calls)	high unlock numbers, high call-in-and-out numbers especially during the evening, less sedentary duration
3 (Traveling and socializing with others at night)	longer distance traveled during night and evening, longer conversation duration, longer duration in vehicle during night, more frequently in nightlife areas, getting up late, less time playing games, less usage of engagement apps (i.e., finance/tools/education/business, etc.,) on the phone
4 (Texting and chatting on the phone without others around)	high SMS in-and-out numbers, less environmental voice duration, more time using social networking apps (i.e., Facebook, Instagram, etc.).
7 (Being physically active outside during the day without others around)	more active on foot, in brighter ambient light places in the morning and afternoon periods, less exposure to conversation
9 (Bike riding and spending time in different places at night)	more active on bike, visiting more places during the night period, less frequently in nightlife areas, less time playing games, less time using entertainment apps (e.g., music/audio/video, etc.).
11 (Spending time in educational settings and sleeping late and gaming less)	more frequently being in functional areas especially in educational places and library, less time at home, shorter sleep duration, going to bed late, longer unlock duration (i.e, heavy phone user), playing less games but often using browser app and music/video apps
13 (Sleeping and waking earlier)	early to bed, early to rise, less active on foot especially during the afternoon and evening periods, more usage of engagement apps (finance/tools/education/business etc.,), less frequenting educational areas
14 (Spending time in the library and visiting few places in the evening)	showing up more frequently in libraries, showing up less in residential area but not in the residential area where they live, visiting fewer places during the evening period, light phone usage, not often playing games, not using browser, engagement apps and direct communication apps (i.e., Hangout, Skype, etc.,), however, using social networking apps (e.g., Facebook)
17 (Spending less time in the gym and going to bed early)	less gym visiting, going to bed early, shorter call duration during the night period, less time at home, more phone use of direct communication apps, less usage of social networking apps
21 (Mostly using gaming apps and going to sleep late)	using mobile phones as solely as gaming devices, seldom using other apps like social apps, communication apps and browser, going to bed late

Table 5: Patterns associated with 7 sub-areas of social functioning

SFS sub-scale	associated behavioral pattern
social engagement	*pattern 7 (-)*, pattern 17 (+)
interpersonal behavior	pattern 1 (+), pattern 14 (-)
prosocial activities	pattern 1 (+)
recreation	pattern 11 (-), pattern 21 (-), pattern 3 (+)
independence-competence	pattern 9 (+), pattern 4 (-)
independence-performance	**pattern 1 (+)**
employment	**pattern 3 (+)**, pattern 13 (+)

(-) negative association, (+) positive association, all associations with $p < 0.05$. $FDR < 0.1$ in **bold** and $FDR < 0.05$ in ***bold italic***

the person has with friends (e.g., "do people discuss their problems with you?", "how often are you able to carry out a sensible or natural conversation?"). We find higher interpersonal behavior scores are positively associated with pattern 1 (using the phone frequently for phone calls) and negatively associated with pattern 14 (spending time in the library and visiting few places in the evenings).

Prosocial activities sub-scale contains 22 questions measuring detailed social activities in everyday life. The questions ask how often people go to the movies/theatre/concert, watch/play indoor/outdoor sports, visit art gallery/museum/exhibition/fair, visit relatives/friends, go to pub/bar, eat out in restaurant, etc. We find that pattern 1 (using the phone frequently for phone calls) is positively associated with higher scores in the prosocial activity sub-scale.

Recreation sub-scale asks people's engagement in a range of hobbies and interests, such as singing, playing instruments, reading, gardening, watching television, cooking, and hiking. Pattern 11 (spending time in educational settings and sleeping later and less gaming) and pattern 21 (merely using gaming apps and going to sleep late) are negatively associated with

recreation score whereas pattern 3 (traveling and socializing with others at night) is positively associated with recreation. The finding suggests participants who are engaged in educational settings or mobile games tend to have fewer hobbies and interests, while those who socialize with other people during the night are more likely to enjoy various recreations.

Independence-competence sub-scale measures people's ability to perform the skills necessary for living independently. Such abilities include taking public transportation, budgeting, handling money correctly, cleaning, weekly purchasing, taking care of personal appearance. These abilities are usually internal and hard to be captured by passive sensing. However, we find two patterns that indicate the strength and weakness in this scale. Stronger pattern 9 (bike riding and spending time in different places) indicates a higher independence competence score whereas weaker pattern 4 (texting and chatting on the phone without others around) indicates a lower independence competence score.

Independence-performance sub-scale examines the performance of skills necessary for independent living. Unlike the Independence-Competence sub-scale, the independence-performance sub-scale asks how often people perform activities necessary in independent living rather than their self-evaluated abilities. We find that pattern 1 (talker over phone) is positively associated with the performance in activities. Although calling and texting are often associated with interpersonal motives [58], no prior work shows the correlations between using calling on phones and the performance of skills necessary for independent living. There is a need to better understand the reason for such connections.

Employment or occupation sub-scale measures engagement in employment or structured daily activity program. The score ranges from 0 to 10 where 10 indicates either full time gainful

earnings, full time employment, or are homemaker managing most household affairs for self and 0 indicates unable to make attempts to find a job [12]. Pattern 3 (traveling and socializing with others at night) and pattern 13 (Sleeping and waking earlier) are positively associated with engagement in employment or occupation. Pattern 3 describes people who are likely to enjoy the nightlife whereas pattern 13 portrays who are early to bed, early to rise and having fewer activities on foot during the afternoon and evening. They might depict two styles of employers: some are night owls who like to have fun after work until late, and some prefer early hours. It is not surprising since traveling in a vehicle (regular commute) in the evening and night is very likely to be a sign of employment. On the other hand, existing literature suggests the link between work schedules and chronotype, in which the authors found the unemployed are less likely to be morning types [74].

Predicting Social Functioning Scale
We present the best combinations of time window, feature transformation, and predictive model to predict the SFS sub-scale scores in Table 6. The performance is reported using mean absolute error (MAE), root-mean-square error (RMSE), Pearson correlation, the association between the predicted and the true values grouped by 5 folds in cross validation. We also compare the MAE and RMSE of our model with the baseline model, in which every test case is predicted using the mean of the scores from the training data. Overall, the tree-ensemble models (RF, ET & XGBoost) outperform linear regression and SVR models. The mean absolute error is around 10% of the possible range of each sub-scale. The predicted SFS scores are positively associated with their responses. All the predictive models are performing better than the baseline models.

To further understand the performance under various scenarios, we present the accuracy from LOSO and CLOSO. Table 6 shows the performance of the best models. In general, we find that the pure LOSO process performs poorly. This indicates that our dataset is still not sufficient enough in order to train a model purely on the data from other people. This is somehow expected; social functioning sub-scales consist of many items that measure a specific type of functioning from multiple behaviors. Those behaviors could be highly personalized, even among people with similar scores. Behaviors, unfamiliar to the model from an unseen user, prevent the system to be trained well. Therefore, we still need to leverage knowledge from the population and a small amount of subjective behavioral patterns; in this case, we adopt the first chronological response of the participant. This additional data point serves as an initial anchor; the models then predict the following variances based on the knowledge from the population. The performance of CLOSO is significantly better than LOSO's and close to 5-fold cross validation's. We calculate the mean absolute error for each participant during the LOSO and CLOSO, and thus compute the distribution of the per-user errors. The per-user absolute errors from CLOSO are less than 15% of the total ranges for most of the sub-scales.

Previous studies in sensing schizophrenia show that the variations or trajectory of sensed behaviors are informative in predicting the dynamic symptoms. However, including these

features did not show any significant improvement in our prediction. We assume these features are unimportant to prediction social function where it asks about the recent/past aggregated activities. Therefore, we choose a simple approach; that is, we adopt only the average of sensed behaviors in d days (where d = 90, 60, 30, 15) before the date of the assessments as predictors in our model. Considering the different characters in sub-scales, SFS measures behavior in various periods for different sub-scales. In particular, participants are asked "how often you have done the following *over the past 3 months*" in prosocial activities, independence-competence, and independence-performance, while for other sub-scales the period is not explicitly indicated. As shown in Table 6, we find different optimum window for each sub-scale. The model selects an optimal window of a 90-day time frame for prosocial activities, independence-competence, and independence-performance, which matches the way social functioning is measured. This verifies that the mobile sensing techniques can capture the daily behaviors for automatic assessment. For sub-scale with questions mostly focusing on current status without specifying a range of time (e.g., social engagement/withdraw and interpersonal communication), we find shorter time frames work better.

The model performs worst in predicting the recreation sub-scale. The reason might be due to the disability of mobile phone sensors in capturing detailed items for recreation, such as swimming, gardening, cooking, etc. In addition, the authors of SFS [12] report that the reliability (α coefficient [28] = 0.69) of the recreation sub-scale is significantly lower than the other sub-scales (0.82-0.87). This indicates that the people could not remember the frequency of various recreational activities over the past 3 months. In future studies, extra mobile sensors can be leveraged for better capturing recreational activities.

Patient Case Studies
Here we highlight one anecdotal case study for showing the effect of this model. The patient in Figure 2 is a 50 year-old white female diagnosed with schizophrenia. She has a high school education, is unemployed and lives independently. She was working as a secretary in her last job. She was enrolled in Mar. 2015 and did not experience any relapses during the study. At the beginning of the study, she reported that usually in the morning, she *had breakfast, drank coffee, watched news channel, did shower and dress, checked schedule and phoned with friends*; in the afternoon, she did *job searching, going for work appointment, workout, spending time with husband*; and during the evening, she *watched TV*, then *went to sleep*. In the first 6 months, she was confident in her capability of employment and was looking for jobs every day. However, her daily occupation significantly changed according to her assessment in Sept. 2015. During that time, she usually *slept until 2-3 pm* and *watched TV with husband* during the afternoon. She was still confident that she was capable of some sort of employment, but she rarely did any job hunting. Her last assessment in Mar. 2016 showed that she *watched TV* almost all the time in the morning as well as in the afternoon. She reported *dinner, cleaning up and sleep* as the major occupation in the evening. She thought she *would have some difficulty in employment* and never looked for a job during that time.

Table 6: Prediction Summary using leave-one-subject-out and customized leave-one-subject-out

sub-scale	model		5-fold cross validation			leave-one-subject-out		customized leave-one-subject-out	
	window/regressor/feature	Pearson r-value	test MAE/RMSE	vs. baseline MAE/RMSE	Pearson r-value	per-user error mean(std)	Pearson r-value	per-user error mean(std)
social engagement	15 / RF / kernel pca 80%	0.44	1.75/2.28	-0.12/-0.28	0.22	2.06 (1.01)	0.46	2.00 (1.51)
interpersonal behavior	60 / ET / orig. features	0.57	3.39/4.26	-0.56/-0.88	0.38	4.76 (2.77)	0.44	3.87 (2.27)
prosocial activities	90 / XGB / orig. features	0.53	7.79/10.35	-2.11/-1.91	0.23	11.19 (6.81)	0.39	9.10 (5.23)
recreation	15 / XGB / orig. features	0.39	6.62/8.15	-0.72/-0.64	0.16	10.15 (4.57)	0.31	6.75 (3.94)
independence competence	90 / ET / kernel pca 95%	0.51	2.06/2.91	-0.31/-0.49	0.17	2.35 (2.35)	0.37	2.59 (1.80)
independence performance	90 / ET / orig. features	0.47	4.49/5.78	-0.80/-0.81	0.23	5.65 (3.61)	0.47	4.87 (3.43)
employment	90/ ET / orig. features	0.62	2.17/2.67	-0.76/-0.74	0.38	2.75 (1.41)	0.60	2.35 (1.35)

Figure 2 shows how the CLOSO model predicts her social functioning. The first assessment is included in the training set together with the knowledge from the population.
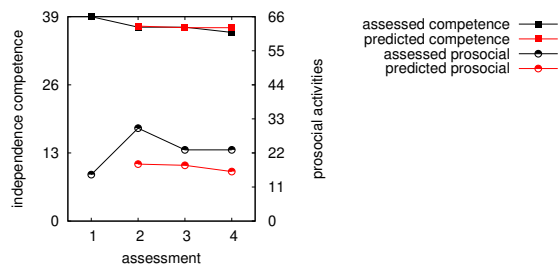
(a) independence competence & prosocial activities
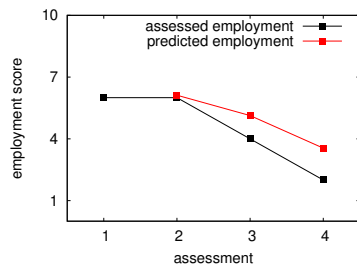
(b) employment

Figure 2: Case study: the dynamics of a participant's assessed and predicted social functioning using CLOSO. (a) the model identifies the strengths and weaknesses between the different components in social functioning. (b) the model tracks the dynamics of social functioning.

From the predicted values we know: (a) this participant kept high independence competence yet had weakness in prosocial activities during the study and (b) experienced deterioration in her employment sub-scale around Sept. 2015. Although the CLOSO still generates predictions skewed to the first input, the clinicians should be aware of the changes indicated by the model. This user case shows the importance and effectiveness of the proposed system. Limited by the scale of the dataset, we admit the model could be further improved with data from more patients. However, the system is capable of detecting the strengths and weaknesses between the different components in social functioning, as well as tracking the dynamics.

DISCUSSION
We use a number of computational approaches to better understand behavioral patterns associated with various aspects of

social functioning. We demonstrate how we can design features and train a model using mobile sensing data from phones of outpatients with schizophrenia to predict social functioning. To the best of our knowledge, we are the first to use mobile sensing in wild settings to study social functioning. Importantly, our approach in assessing social function is applicable to other areas of interest to the HCI community including, for example, cognitive decline in the elderly and dementia. Furthermore, we believe our work opens the way for the development of new forms of assessment and intervention across these areas.

Implications for Social Computing and Health Research
The computational methods discussed in the paper augment current methods to improve understanding of social functioning developed by HCI community; for example, face-to-face evaluation [100] (which can be labor-intensive) and extracting behavior and language from social networks, such as Facebook, Twitter and Reddit [16, 84] (which can only captures online behavior and not continuous assessment in the wild afforded by our approach). In the future, an approach that combined mobile sensing and social media may prove to be even more effective.

Our approach highlights the importance of using multi-modal data in assessing various components of social functioning. In addition to features that have proved effective in prior research [95, 97], we show new features (e.g., communications through mobile phone applications like Skype, Hangout, semantic locations) are important signals for social functioning. Specifically, this work demonstrates how researchers can use the classical PCA technique for interpretable and generalizable behavioral patterns associated with different outcomes. Our approach also shows the steps of combining machine learning techniques (e.g., feature transformation, dimensionality reduction, testing various time windows) to improve the accuracy of assessment models for accessing everyday behavioral patterns in order to predict their general social functioning. This work also proposes three scenarios for understanding performance in different potential clinical settings. It suggests researchers think about the subsequent design and development of possible intervention or support tools. This detailed investigation of how the models might be applied in real-world clinical settings, and understanding the extent to which they may or may not work in various contexts.

Implications for HCI and Design

We believe our approach opens up new forms of social functioning assessment and intervention in the wild to improve peoples' lives. Specifically, we imagine a "sensing core" (that can accurately assess social function in a passive low burden manner) integrated into existing and future health care systems. For populations living serious mental health issues as discussed in this paper, our approach could be a game-changer: many people living with schizophrenia for example live on the edges of our society (like many of the people in our study), many do not have access to good healthcare services or even a stable place to live – fewer than 40% of people living with schizophrenia have contact with mental health services for treatment [32]. In the future, we imagine phone-based mental health assistants will combine passive assessment and social functioning intervention (e.g., akin to cognitive processing therapy) to inform and help people. There are many different design options to consider in the future from standalone systems to ones integrated with care support systems.

We also believe automated tools can be designed to be integrated into domain expert systems using our methods. For example, utilizing PCA of sensor data, clinicians can understand the patterns associated with higher and lower social functions of populations of clients in specific locations, and can be provided with more accurate interventions and suggestions for their target communities by the system – for example vulnerable communities, such as those with mental health issues, aging in place, homeless, etc. We imagine that various providers could integrate core social functioning sensing and intervention technology into their systems – for example, the core could trigger intervention and an early warning sign to clinicians when an outpatient has a significant deterioration of social functions. Furthermore patients, or other at need groups, such as elders, can also benefit from accurate assessment, self-reflection and intervention tools. Using our approach, individuals with social functioning impairment are able to learn from the behavioral patterns of groups with higher social functioning, and potentially avoid patterns of groups with lower social functioning. For example self-tracking of social functioning requires advances in how such data could be presented to the user in an informative and validated manner. Such challenges are at the core of HCI design. How do you present data and hints in an effective manner? How do you take into account that no design or model fits all?. Our approach advances how future systems can be designed. Many mobile intervention systems (e.g., [83, 67]) include predefined conversational therapies, however, they need considerable user input to provide suggestions. Our methods can be integrated into intervening HCI designs [69, 1, 78] to automatically find out personalized strategies according to individuals' current status and improve social functions.

Implications for Sensing Populations at Risk

The participants in our study live extremely challenging lives and we were aware that many had not used a smartphone before and importantly did not fully understand the types of data being collected. Because of this we spent a considerable amount of time and effort educating participants about smartphones and the data being collected. Our team included computer scientists, psychiatrists, clinical psychologists, IRB specialists and the head of psychiatric research at Zucker Hillside Hospital, which specializes in the treatment of schizophrenia. Collectively the team had considerable experience working with vulnerable populations and mobile technology. For example, our mobile UI (e.g., self-reports, turning on/off sensor streams, etc) was designed by a clinical psychologist who has specialized in the development of mobile technology for communities with severe mental illness for two decades. It is important that participants fully understand the implications and risks associated with consenting to a mental health study in the wild. As part of the education and consent process, we screened out people incapable of understanding mobile technology, the data being collected and risks associated with participation.

CONCLUSION

We recognize the limitations of our work. (1) We only have 160 social functioning reports to train our models on. All patients live in a large dense city and the models may not generalize to other locations, such as, patients living in rural communities. (2) Smartphone sensing, though captures a variety of personal behaviors, is still limited to capture the fine-grained items in the questionnaire. For example, according to Table 6, the recreation sub-scale has the worst performance. It might be because mobile phone sensors are not able to capture detailed items in recreation such as "swimming", "gardening", etc. In the future, we could study additional sensing modalities (e.g., wearables, in-house sensors) for more accurate predictions. Due to privacy requirement we only collect the metadata and do not look at the exact content in social media, SMS and audios). In the future, we could study privacy-preserving NLP technologies to boost performance. (3) Some of the analysis and discussions imply some assumptions that may not be generalizable. For example, most of the employees in our study have full-time jobs and are working at regular hours as cashiers, construction workers, etc. Consequently, our interpretation of employment sub-scale is based on regular job hours. However, the results and interpretations may not be generalizable when applied to another dataset, where people could work on non-traditional hours. Researchers should pay attention to this when they generalize the results in this paper. (4) The performance from LOSO reveals that it is still challenging to accurately predict a new patient's social functioning without any historical SFS records. This is inevitable at this stage - we have made the best effort to enroll the participants and collect the year-long data, but the dataset is still small. (5) The real-world impact of the prediction errors is still unclear. For example, from the data science perspective, MAE of 6.6 for recreation is 13.7% error and the lower the MAE, the better the prediction. However, from a strictly clinical perspective, the impact is not yet clear due to the lack of substantial research on the subject. That is why we hope we could open new directions to be precisely studied in future work.

REFERENCES

[1] Phil Adams, Eric PS Baumer, and Geri Gay. 2014. Staccato social support in mobile health applications. In Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 653–662.

[2] Jean Addington and Donald Addington. 2000. Neurocognitive and social functioning in schizophrenia: a 2.5 year follow-up study. Schizophrenia research 44, 1 (2000), 47–56.

[3] Rawan Alharbi, Angela Pfammatter, Bonnie Spring, and Nabil Alshurafa. 2017. WillSense: Adherence Barriers for Passive Sensing Systems That Track Eating Behavior. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, 2329–2336.

[4] American Psychiatric Association and others. 2013. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

[5] Min Aung, Faisal Alquaddoomi, Cheng-Kang Hsieh, Mashfiqui Rabbi, Longqi Yang, JP Pollak, Deborah Estrin, and Tanzeem Choudhury. 2016. Leveraging Multi-Modal Sensing for Mobile Health: a Case Review in Chronic Pain. IEEE Journal of Selected Topics in Signal Processing 10, 5 (2016), 1–13.

[6] Christine Barrowclough and Nicholas Tarrier. 1990. Social functioning in schizophrenic patients. Social Psychiatry and Psychiatric Epidemiology 25, 3 (1990), 125–129.

[7] Dror Ben-Zeev, Christopher J Brenner, Mark Begale, Jennifer Duffecy, David C Mohr, and Kim T Mueser. 2014. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophrenia bulletin (2014), sbu033.

[8] Dror Ben-Zeev, Rachel Brian, Rui Wang, Weichen Wang, Andrew T Campbell, Min SH Aung, Michael Merrill, Vincent WS Tseng, Tanzeem Choudhury, Marta Hauser, and others. 2017. CrossCheck: Integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. Psychiatric rehabilitation journal 40, 3 (2017), 266.

[9] Dror Ben-Zeev, Rui Wang, Saeed Abdullah, Rachel Brian, Emily A Scherer, Lisa A Mistler, Marta Hauser, John M Kane, Andrew Campbell, and Tanzeem Choudhury. 2015. Mobile behavioral sensing for outpatients and inpatients with schizophrenia. Psychiatric services 67, 5 (2015), 558–561.

[10] Yoav Benjamini, Abba M Krieger, and Daniel Yekutieli. 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93, 3 (2006), 491–507.

[11] Frank Bentley and Santosh Basapur. 2012. StoryPlace. Me: the path from studying elder communication to a public location-based video service. In CHI' 12 Extended Abstracts on Human Factors in Computing Systems. ACM, 777–792.

[12] Max Birchwood, JO Smith, Ray Cochrane, Sheila Wetton, and SONJA Copestake. 1990. The Social Functioning Scale The development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients. The British Journal of Psychiatry 157, 6 (1990), 853–859.

[13] M Bosc. 2000. Assessment of social functioning in depression. Comprehensive Psychiatry 41, 1 (2000), 63–69.

[14] Leo Breiman. 2001. Random forests. Machine learning 45, 1 (2001), 5–32.

[15] Michaeline Bresnahan, Melissa D Begg, Alan Brown, Catherine Schaefer, Nancy Sohler, Beverly Insel, Leah Vella, and Ezra Susser. 2007. Race and risk of schizophrenia in a US birth cohort: another example of health disparity? International journal of epidemiology 36, 4 (2007), 751–758.

[16] Moira Burke, Cameron Marlow, and Thomas Lento. 2010. Social network activity and social well-being. In Proceedings of the SIGCHI conference on human factors in computing systems. ACM, 1909–1912.

[17] T Burns and D Patrick. 2007. Social functioning as an outcome measure in schizophrenia studies. Acta Psychiatrica Scandinavica 116, 6 (2007), 403–418.

[18] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, 1293–1304.

[19] Romina Carrasco. 2017. Designing virtual avatars to empower social participation among older adults. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, 259–262.

[20] Thomas C Chalmers, Harry Smith Jr, Bradley Blackburn, Bernard Silverman, Biruta Schroeder, Dinah Reitman, and Alexander Ambroz. 1981. A method for assessing the quality of a randomized control trial. Controlled clinical trials 2, 1 (1981), 31–49.

[21] Stevie Chancellor, Andrea Hu, and Munmun De Choudhury. 2018. Norms matter: contrasting social support around behavior change in online weight loss communities. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 666.

[22] Stevie Chancellor, Yannis Kalantidis, Jessica A Pater, Munmun De Choudhury, and David A Shamma. 2017. Multimodal classification of moderated online pro-eating disorder content. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, 3213–3226.

[23] Stevie Chancellor, Zhiyuan Jerry Lin, and Munmun De Choudhury. 2016. This post will just get taken down: characterizing removed pro-eating disorder social media content. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 1157–1162.

[24] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 785–794.

[25] Zhenyu Chen, Mu Lin, Fanglin Chen, Nicholas D Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tanzeem Choudhury, and Andrew T Campbell. 2013. Unobtrusive sleep monitoring using smartphones. In Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 145–152.

[26] Tanzeem Choudhury, Sunny Consolvo, Beverly Harrison, Jeffrey Hightower, Anthony LaMarca, Louis LeGrand, Ali Rahimi, Adam Rea, G Borriello, Bruce Hemingway, and others. 2008. The mobile sensing platform: An embedded activity recognition system. IEEE Pervasive Computing 7, 2 (2008).

[27] Shannon M Couture, David L Penn, and David L Roberts. 2006. The functional significance of social cognition in schizophrenia: a review. Schizophrenia bulletin 32, suppl_1 (2006), S44–S63.

[28] Lee J Cronbach. 1951. Coefficient alpha and the internal structure of tests. psychometrika 16, 3 (1951), 297–334.

[29] Alex W DaSilva, Jeremy F Huckins, Rui Wang, Weichen Wang, Dylan D Wagner, and Andrew T Campbell. 2019. Correlates of Stress in the College Environment Uncovered by the Application of Penalized Generalized Estimating Equations to Mobile Sensing Data. JMIR mHealth and uHealth 7, 3 (2019), e12084.

[30] Faith Dickerson, John J Boronow, Norman Ringel, and Frederick Parente. 1999. Social functioning and neurocognitive deficits in outpatients with schizophrenia: a 2-year follow-up1. Schizophrenia research 37, 1 (1999), 13–20.

[31] David Dodell-Feder, Laura M Tully, and Christine I Hooker. 2015. Social impairment in schizophrenia: new approaches for treating a persistent problem. Current opinion in psychiatry 28, 3 (2015), 236.

[32] Roisin Doyle, Niall Turner, Felicity Fanning, Daria Brennan, Laoise Renwick, Elizabeth Lawlor, and Mary Clarke. 2014. First-episode psychosis and disengagement from treatment: a systematic review. Psychiatric Services 65, 5 (2014), 603–611.

[33] Harris Drucker, Christopher JC Burges, Linda Kaufman, Alex J Smola, and Vladimir Vapnik. 1997. Support vector regression machines. In Advances in neural information processing systems. 155–161.

[34] Ahmed El-Rabbany. 2002. Introduction to GPS: the global positioning system. Artech house.

[35] Sindhu Kiranmai Ernala, Michael L Birnbaum, Kristin A Candan, Asra F Rizvi, William A Sterling, John M Kane, and Munmun De Choudhury. 2019. Methodological gaps in predicting mental health states from social media: Triangulating diagnostic signals. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, 134.

[36] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, and others. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In Kdd, Vol. 96. 226–231.

[37] Jessica L Feuston, Charlotte G Marshall-Fricker, and Anne Marie Piper. 2017. The social lives of individuals with traumatic brain injury. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, 182–194.

[38] Jessica L Feuston and Anne Marie Piper. 2019. Everyday Experiences: Small Stories and Mental Illness on Instagram. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, 265.

[39] Foursquare Place API. 2018. Foursquare Place API. `https://developer.foursquare.com/places-api`. (2018).

[40] Daniel Fulford, Tim Campellone, and David E Gard. 2018. Social motivation in schizophrenia: How research on basic reward processes informs and limits our understanding. Clinical Psychology Review 63 (2018), 12–24.

[41] Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. Machine learning 63, 1 (2006), 3–42.

[42] Google Activity Recognition Api. 2017. Google Activity Recognition Api. `https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi`. (2017).

[43] Google Call Log Api. 2019. Google Call Log Api. `https://developer.android.com/reference/android/provider/CallLog.Calls`. (2019).

[44] Google Environment Sensors Api. 2019. Google Environment Sensors Api. `https://developer.android.com/guide/topics/sensors/sensors_environment.html`. (2019).

[45] Google Microphone Api. 2019. Google Microphone Api. `https://developer.android.com/reference/android/media/AudioRecord`. (2019).

[46] Google Position Sensors Api. 2019. Google Position Sensors Api. `https://developer.android.com/guide/topics/sensors/sensors_position.html`. (2019).

[47] Michael Foster Green. 1996. What are the functional consequences of neurocognitive deficits in schizophrenia? The American journal of psychiatry 153, 3 (1996), 321.

[48] Michael F Green, William P Horan, Junghee Lee, Amanda McCleery, L Felice Reddy, and Jonathan K Wynn. 2017. Social disconnection in schizophrenia and the general community. Schizophrenia bulletin 44, 2 (2017), 242–249.

[49] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. 2011. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM review 53, 2 (2011), 217–288.

[50] Teng Han, Xiang Xiao, Lanfei Shi, John Canny, and Jingtao Wang. 2015. Balancing Accuracy and Fun: Designing Camera Based Mobile Games for Implicit Heart Rate Monitoring. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 847–856. DOI: http://dx.doi.org/10.1145/2702123.2702502

[51] Gabriella M Harari, Sandrine R Müller, Clemens Stachl, Rui Wang, Weichen Wang, Markus Bühner, Peter J Rentfrow, Andrew T Campbell, and Samuel D Gosling. 2019. Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. Journal of personality and social psychology (2019).

[52] James W Hardin and Joseph M Hilbe. 2002. Generalized estimating equations. Chapman and Hall/CRC.

[53] Javier Hernandez, Pablo Paredes, Asta Roseway, and Mary Czerwinski. 2014. Under Pressure: Sensing Stress of Computer Users. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14). ACM, New York, NY, USA, 51–60. DOI: http://dx.doi.org/10.1145/2556288.2557165

[54] Arthur E Hoerl and Robert W Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12, 1 (1970), 55–67.

[55] Nan Holmes, Amitta Shah, and Lorna Wing. 1982. The Disability Assessment Schedule: a brief screening device for use with the mentally retarded. Psychological Medicine 12, 4 (1982), 879–890.

[56] Hsiang-Ting Hong, Tzu-Yu Su, Po-Hsun Lee, Ping-Chun Hsieh, and Mian-Jhong Chiu. 2017. VisualLink: Strengthening the Connection between Hearing-impaired Elderly and their Family. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, 67–73.

[57] Joon Hwan Jang, Na Young Shin, Geumsook Shim, Hye Yoon Park, Euitae Kim, Go-Eun Jang, Soo Jin Kwon, Ji-Won Hur, Suk Kyoon An, and Jun Soo Kwon.

2011. Longitudinal patterns of social functioning and conversion to psychosis in subjects at ultra-high risk. Australian & New Zealand Journal of Psychiatry 45, 9 (2011), 763–770.

[58] Borae Jin and Namkee Park. 2010. In-person contact begets calling and texting: Interpersonal motives for cell phone use, face-to-face interaction, and loneliness. Cyberpsychology, Behavior, and Social Networking 13, 6 (2010), 611–618.

[59] Rosalie A Kane. 1987. Assessing social function in the elderly. Clinics in geriatric medicine 3, 1 (1987), 87–98.

[60] Kurt Kroenke and Robert L Spitzer. 2002. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 32, 9 (2002), 509–515.

[61] Kurt Kroenke, Robert L Spitzer, and Janet BW Williams. 2001. The PHQ-9. Journal of general internal medicine 16, 9 (2001), 606–613.

[62] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. 2010. A survey of mobile phone sensing. Communications Magazine, IEEE 48, 9 (2010), 140–150.

[63] Hanchuan Li, Can Ye, and Alanson P Sample. 2015. IDSense: A human object interaction detection system based on passive UHF RFID. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2555–2564.

[64] Kung-Yee Liang and Scott L Zeger. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 1 (1986), 13–22.

[65] Micah Linnemeier, Yi-Ying Lin, Gierad Laput, and Ramachandra Vijjapurapu. 2012. StoryCubes: connecting elders in independent living through storytelling. In CHI '12 Extended Abstracts on Human Factors in Computing Systems. ACM, 1321–1326.

[66] Lydia Manikonda and Munmun De Choudhury. 2017. Modeling and understanding visual attributes of mental health disclosures in social media. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, 170–181.

[67] Mark Matthews, Stephen Voida, Saeed Abdullah, Gavin Doherty, Tanzeem Choudhury, Sangha Im, and Geri Gay. 2015. In situ design for mental illness: Considering the pathology of bipolar disorder in mhealth design. In Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services. ACM, 86–97.

[68] M Pilar Matud and M Concepción García. 2019. Psychological distress and social functioning in elderly Spanish people: a gender analysis. International journal of environmental research and public health 16, 3 (2019), 341.

[69] Claire McCallum, John Rooksby, Parvin Asadzadeh, Cindy M Gray, and Matthew Chalmers. 2019. An N-of-1 Evaluation Framework for Behaviour Change Applications. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, LBW1810.

[70] Peter McCullagh and John A Nelder. 1989. Generalized linear models. Vol. 37. CRC press.

[71] Shayan Mirjafari, Kizito Masaba, Ted Grover, Weichen Wang, Pino Audia, Andrew T Campbell, Nitesh V Chawla, Vedant Das Swain, Munmun De Choudhury, Anind K Dey, and others. 2019. Differentiating Higher and Lower Job Performers in the Workplace Using Mobile Sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 3, 2 (2019), 37.

[72] Kim Tornvall Ed Mueser and Nicholas Ed Tarrier. 1998. Handbook of social functioning in schizophrenia. Allyn & Bacon.

[73] Thomas F Oltmanns, Alison H Melley, and Eric Turkheimer. 2002. Impaired social function and symptoms of personality disorders assessed by peer and self-report in a nonclinical population. Journal of Personality Disorders 16, 5 (2002), 437–452.

[74] Sarah-Jane Paine, Philippa H Gander, and Noemie Travier. 2006. The epidemiology of morningness/eveningness: influence of age, gender, ethnicity, and socioeconomic factors in adults (30-49 years). Journal of biological rhythms 21, 1 (2006), 68–76.

[75] David L Penn, Patrick W Corrigan, Richard P Bentall, J Racenstein, and Leonard Newman. 1997. Social cognition in schizophrenia. Psychological bulletin 121, 1 (1997), 114.

[76] Deborah Perlick, Peter Stastny, Steven Mattis, and Jeanne Teresi. 1992. Contribution of family, cognitive and clinical dimensions to long-term outcome in schizophrenia. Schizophrenia Research 6, 3 (1992), 257–265.

[77] Mashfiqui Rabbi, Shahid Ali, Tanzeem Choudhury, and Ethan Berke. 2011. Passive and in-situ assessment of mental and physical well-being using mobile sensors. In Proceedings of the 13th international conference on Ubiquitous computing. ACM, 385–394.

[78] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2015. Pass the ball: enforced turn-taking in activity tracking. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2417–2426.

[79] Sohrab Saeb, Emily G Lattie, Stephen M Schueller, Konrad P Kording, and David C Mohr. 2016. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 4 (2016), e2537.

[80] Sohrab Saeb, Mi Zhang, Christopher J Karr, Stephen M Schueller, Marya E Corden, Konrad P Kording, and David C Mohr. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. Journal of medical Internet research 17, 7 (2015), e175.

[81] Hillol Sarker, Matthew Tyburski, Md Mahbubur Rahman, Karen Hovsepian, Moushumi Sharmin, David H Epstein, Kenzie L Preston, C Debra Furr-Holden, Adam Milam, Inbal Nahum-Shani, and others. 2016. Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. ACM, 4489–4501.

[82] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. 1997. Kernel principal component analysis. In International Conference on Artificial Neural Networks. Springer, 583–588.

[83] Jessica Schroeder, Chelsey Wilkes, Kael Rowan, Arturo Toledo, Ann Paradiso, Mary Czerwinski, Gloria Mark, and Marsha M Linehan. 2018. Pocket skills: A conversational mobile web app to support dialectical behavioral therapy. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 398.

[84] N Sadat Shami, Jeffrey Nichols, and Jilin Chen. 2014. Social media participation and performance at work: a longitudinal study. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 115–118.

[85] Alex J Smola and Bernhard Schölkopf. 2004. A tutorial on support vector regression. Statistics and computing 14, 3 (2004), 199–222.

[86] Robert L Spitzer, Kurt Kroenke, Janet BW Williams, Patient Health Questionnaire Primary Care Study Group, and others. 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Jama 282, 18 (1999), 1737–1744.

[87] Gregory P Strauss, Martin Harrow, Linda S Grossman, and Cherise Rosen. 2008. Periods of recovery in deficit syndrome schizophrenia: a 20-year multi–follow-up longitudinal study. Schizophrenia bulletin 36, 4 (2008), 788–799.

[88] Greer Sullivan, Stephen R Marder, Robert P Liberman, Clyde P Donahoe, and Jim Mintz. 1990. Social skills and relapse history in outpatient schizophrenics. Psychiatry 53, 4 (1990), 340–345.

[89] Emily Sun and Mor Naaman. 2018. A Multi-site Investigation of Community Awareness Through Passive Location Sharing. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 581.

[90] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) (1996), 267–288.

[91] Allen Y Tien and William W Eaton. 1992. Psychopathologic precursors and sociodemographic risk factors for the schizophrenia syndrome. Archives of General Psychiatry 49, 1 (1992), 37–46.

[92] Timothy J Trull and Ulrich Ebner-Priemer. 2013. Ambulatory assessment. Annual review of clinical psychology 9 (2013), 151–176.

[93] Eva Velthorst, Anne-Kathrin J Fett, Avraham Reichenberg, Greg Perlman, Jim van Os, Evelyn J Bromet, and Roman Kotov. 2016. The 20-year longitudinal trajectories of social functioning in individuals with psychotic disorders. American Journal of Psychiatry 174, 11 (2016), 1075–1085.

[94] Greg Wadley, Rafael Calvo, John Torous, and Mary Czerwinski. 2018. 3rd symposium on computing and mental health: understanding, engaging, and delighting users. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, Sym04.

[95] Rui Wang, Min S. H. Aung, Saeed Abdullah, Rachel Brian, Andrew T. Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Michael Merrill, Emily A. Scherer, Vincent W. S. Tseng, and Dror Ben-Zeev. 2016. CrossCheck: Toward Passive Sensing and Detection of Mental Health Changes in People with Schizophrenia. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16). ACM, New York, NY, USA, 886–897. DOI: `http://dx.doi.org/10.1145/2971648.2971740`

[96] Rui Wang, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. 2014. StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14). ACM, New York, NY, USA, 3–14. DOI: `http://dx.doi.org/10.1145/2632048.2632054`

[97] Rui Wang, Weichen Wang, Min SH Aung, Dror Ben-Zeev, Rachel Brian, Andrew T Campbell, Tanzeem Choudhury, Marta Hauser, John Kane, Emily A Scherer, and others. 2017. Predicting Symptom Trajectories of Schizophrenia using Mobile Sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, 3 (2017), 110.

[98] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F Huckins, William M Kelley, Todd F Heatherton, and Andrew T Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 1 (2018), 43.

[99] Weichen Wang, Gabriella M Harari, Rui Wang, Sandrine R Müller, Shayan Mirjafari, Kizito Masaba, and Andrew T Campbell. 2018a. Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, 3 (2018), 141.

[100] Xiying Wang, Tiffany Knearem, Fanlu Gui, Srishti Gupta, Haining Zhu, Michael Williams, and John M Carroll. 2018b. A Safety Net: How Older Adults Build and Maintain Interpersonal Relationships. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, LBW015.

[101] Myrna M Weissman. 1975. The assessment of social adjustment: a review of techniques. Archives of general psychiatry 32, 3 (1975), 357–365.

[102] Jennifer Wiley, Ja-young Sung, and Gregory Abowd. 2006. The message center: enhancing elder communication. In CHI'06 Extended Abstracts on Human Factors in Computing Systems. ACM, 1523–1528.

[103] Svante Wold, Kim Esbensen, and Paul Geladi. 1987. Principal component analysis. Chemometrics and intelligent laboratory systems 2, 1-3 (1987), 37–52.

[104] Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut. 2019. The Channel Matters: Self-disclosure, Reciprocity and Social Support in Online Cancer Support Groups. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, 31.

[105] Hui Zou and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 2 (2005), 301–320.