Objective Measurement of Sociability and Activity: Mobile Sensing in the Community

Ethan M. Berke, MD, MPH[1,2]

Tanzeem Choudhury, PhD[3]

Shabid Ali, BS[2]

Mashfiqui Rabbi, MS[3]

[1]Center for Population Health, The Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, New Hampshire

[2]Department of Community and Family Medicine, Dartmouth Medical School, Lebanon, New Hampshire

[3]Department of Computer Science, Dartmouth College, Hanover, New Hampshire

ABSTRACT

PURPOSE Automated systems able to infer detailed measures of a person's social interactions and physical activities in their natural environments could lead to better understanding of factors influencing well-being. We assessed the feasibility of a wireless mobile device in measuring sociability and physical activity in older adults, and compared results with those of traditional questionnaires.

METHODS This pilot observational study was conducted among a convenience sample of 8 men and women aged 65 years or older in a continuing care retirement community. Participants wore a waist-mounted device containing sensors that continuously capture data pertaining to behavior and environment (accelerometer, microphone, barometer, and sensors for temperature, humidity, and light). The sensors measured time spent walking level, up or down an elevation, and stationary (sitting or standing), and time spent speaking with 1 or more other people. The participants also completed 4 questionnaires: the 36-Item Short Form Health Survey (SF-36), the Yale Physical Activity Survey (YPAS), the Center for Epidemiologic Studies–Depression (CES-D) scale, and the Friendship Scale.

RESULTS Men spent 21.3% of their time walking and 64.4% stationary. Women spent 20.7% of their time walking and 62.0% stationary. Sensed physical activity was correlated with aggregate YPAS scores ($r^2 = 0.79$, $P = .02$). Sensed time speaking was positively correlated with the mental component score of the SF-36 ($r^2 = 0.86$, $P = .03$), and social interaction as assessed with the Friendship Scale ($r^2 = 0.97$, $P = .002$), and showed a trend toward association with CES-D score ($r^2 = -0.75$, $P = .08$). In adjusted models, sensed time speaking was associated with SF-36 mental component score ($P = .08$), social interaction measured with the Friendship Scale ($P = .045$), and CES-D score ($P = .04$).

CONCLUSIONS Mobile sensing of sociability and activity is well correlated with traditional measures and less prone to biases associated with questionnaires that rely on recall. Using mobile devices to collect data from and monitor older adult patients has the potential to improve detection of changes in their health.

INTRODUCTION

An important goal of community health programs is to improve the overall quality of life by promoting cognitive, physical, and social/emotional well-being.[1,2] Everyday behaviors are often reflective of physical and physiologic health states, and can be predictive of future health problems. The standard practice for collecting behavioral data in the health sciences relies on observational data collected in laboratory settings or through periodic surveys or self-reports. These proxy measures have several major limitations, however: (1) the time and resource requirements are too great to simultaneously gather data from a large number of individuals; (2) the measurements are prone to considerable bias, and the manual and sporadic recording of information often fails to capture the finer details of behavior that may be important; and (3) the effort required

Conflicts of interest: authors report none.

CORRESPONDING AUTHOR

Ethan Berke, MD, MPH
The Dartmouth Institute for Health Policy and Clinical Practice
HB 7251, Room 3014
35 Centerra Parkway
Lebanon, NH 03766
Ethan.Berke@tdi.dartmouth.edu

of the end user is too high to be suitable for continuous long-term monitoring.

Automatic sensing of physical health at the level of the individual is an active research area.[3-8] Progress in sensing social and cognitive well-being has been limited, however, especially when it comes to measuring social interactions (through shared activities and conversations) and relationships (eg, friendly, formal). Current work pertaining to mobile sensing in health has predominantly focused on physical activity, using accelerometers present in many devices, including some activity-specific devices available commercially.[9,10] Others use the accelerometer to sense changes in gait.[11] Use of sensing in devices to measure behavior has been rarer,[12] with only a few examples of mobile phones used to encourage behavior change based on sensed information reported.[8,13,14]

Mobile sensing, if proven valid, feasible, and effective, could be of benefit to clinicians, patients, and researchers. Automated systems that infer highly detailed measures of people's social and behavioral dynamics in their natural environments over extended periods of time may improve care in several ways. First, they may enable greater access to care by dramatically increasing the number of individuals whose health and well-being can be monitored simultaneously through automation, allowing for larger panels of patients to be monitored by a primary care or medical home team. Second, they may enable improved quality of care as a result of detailed analysis of how individuals interact with each other and how well they perform a given task, which can lead to better understanding of behavioral factors that influence social and cognitive well-being and thereby allow clinicians to better select appropriate interventions. Third, they may enable reduced burden and improved effectiveness of care by lowering the effort needed for early diagnosis, behavioral interventions, and self-monitoring to improve social and cognitive well-being through automatic tracking and detailed analysis of behavior.

We developed an automated behavioral monitoring paradigm for sensing, recognizing, and presenting a range of physical, social, and mental indicators of well-being in natural everyday settings in older adults, and tested it against existing measures. The metrics we developed may be useful for quantifying social wellness from the behavioral indicators and for better understanding how mobile technology can advance health assessment and interventions. We describe this platform and its deployment in a real-world setting of older adults to assess its validity with established health instruments and its feasibility in this population.

METHODS

Participants

All participants in this pilot study resided in a local continuing care community of approximately 400 residents, with the majority fully independent. The facility also provides assisted living and nursing home care, although in our study, all participants were independent. Men and women were eligible for the study if they were aged 65 years or older, spoke English, were not wheelchair bound, were not institutionalized, and had a Mini-Mental State Examination[15] score of 24 or higher, indicating normal cognitive function. We set a sample size of 8 as a goal based on the available number of devices at the time of the study. The convenience sample included 4 men and 4 women, with 2 couples and 4 single residents. Recruitment was carried out through use of posters placed in residents' mailboxes and elsewhere, and word of mouth. With this approach, we were contacted by older adults who were recruited sequentially until our predetermined sample size of 8 participants was met. We did not try to pick specific ages as long as participants were aged 65 years or older, but we did attempt to recruit 2 couples to better test audio sensing as couples would be more likely to interact with each other.

The Dartmouth Committee for the Protection of Human Subjects approved the study protocol. All participants completed a written informed consent process before starting the protocol.

Questionnaires

At the start of the protocol in August 2009, all participants completed a battery of questionnaires to assess their behavior and activity. The Yale Physical Activity Survey (YPAS) is a researcher-administered, written questionnaire that queries participants on various types of activity, including recreational and work activity. Results are reported in kilocalories of energy expended as well as minutes.[16] The 36-Item Short Form Health Survey (SF-36) is a commonly used scale of physical and mental well-being that can be reported in sum or in component parts. Scores range from 0 to 100, with higher scores indicating better physical and mental health.[17] The Friendship Scale is a self-administered questionnaire with 6 questions scored from 0 to 4. Total scores range from 0 (completely socially isolated) to 24 (highly socially connected).[18] The Center for Epidemiologic Studies–Depression (CES-D) scale is a validated, self-administered 20-item questionnaire of depressive symptoms, with a range of 0 to 60. Scores of 16 or higher are usually regarded as being associated with marked depressive symptoms.[19] At the end of the study period, participants again completed these questionnaires, as well as a brief

usability questionnaire about the device and trust of sensing devices, and a focus group.

Sensing Device

The sensing device was approximately 2 inches long by 2 inches wide by one-half inch deep. The device contained an accelerometer (sampled at 256 Hz); a microphone (sampled at 16 kHz); sensors for barometric pressure, temperature, humidity, visible light, and infrared light; and a battery and processor. Each data point was time-stamped. The devices were equipped with a clip so the participant could wear it comfortably around the waist.

Each participant wore the device for 10 days. Participants were asked to wear it for 8 to 10 hours per day, morning through evening, according to an agreed-on schedule. They received no special instruction on changing their daily routine, and were allowed to travel outside the facility with the device. At the start of the protocol, participants completed a series of activities (walking, ascending and descending stairs, and conversing with each other) that lasted less than 10 minutes to calibrate the device. The calibration process involved updating the classifiers with the labeled examples and verifying the performance was consistent with the previous results. Devices were collected daily, at which point data were extracted and batteries were recharged.

The sensing device did not record raw audio, except during the hour-long orientation period at the beginning of device deployment. We therefore did not have any information about the content of a conversation, to protect participants' privacy. We were not interested in the actual words spoken, but rather in the presence and style of interactions. The style of interactions were characterized by various paralinguistic aspects of speech, such as speaking rate, pitch, loudness, turn duration, and turn frequency. With the information collected, we had no ability to reconstruct actual words or phrases that in turn could raise serious privacy concerns. We computed the amount of time spoken for each individual using a 2-state hidden Markov model that classified speech vs nonspeech by examining features shown to be useful for differentiating speech from other noise (noninitial maximum autocorrelation peak, total number of autocorrelation peaks, and relative spectral entropy of sound detected).

On the basis of the motion information collected using the triaxial accelerometer and barometer (which provided data on change in elevation associated with using stairs, elevators, or ramps), we inferred the amount of time an individual spent (1) walking on flat surfaces, (2) walking up or down an elevation, including stairs, (3) being stationary (sitting or standing), and (4) doing other, unclassified activities, in 0.25-second increments. We also recorded environmental information (light, temperature, and humidity) that we did not use in the current analyses. Additional technical details about the sensors and sensor-processing algorithms can be found elsewhere.[20,21]

Analysis

We scored questionnaires using their established protocols. Extracted sensor data were amalgamated into a number of measures and then classified into physical behaviors (activity) or social behaviors (sociability) automatically by computer. Total time spent doing physical activity was a weighted score of time spent walking level, running, walking up elevations, and walking down elevations, using a predictive regression protocol to find the best approximation of the YPAS summary value. To get a single physical activity score from the sensed activities, we had to weight the different activities based on their relative effect on physical well-being. For example, walking up elevation should get a higher score than walking on a flat surface. Considering "unclassified" activity may seem unnecessary, but it may be relevant as this category represents activities that are not walking level, up, or down. Consequently, a higher percentage of unclassified activity implies lower percentage of other activities. Using this reasoning as a base, we refined the model weights by using the YPAS activity score as a reference; we then performed multivariate regression analysis of the (objectively measured) activity percentages and estimated weights to improve the relationship with the YPAS scores. The weights were in agreement with our a priori expectations (higher weight for walking up compared with walking level, negative weight for "unclassified") with the added advantage of using a data-driven approach to selecting the absolute weight values. Time spent stationary was also recorded. To assess sociability, we measured fraction of time spent speaking in conversations with at least 1 other person.

After computing descriptive statistics and determining distributions for normality, we computed unadjusted correlation coefficients between the YPAS and SF-36 physical component score (PCS) and sensor measures of physical activity. We then computed correlations between SF-36 mental component score (MCS), CES-D, and Friendship Scale with sensor measures for mental well-being using time spent speaking. Data from the questionnaires completed at the start and end of the study were compared using a *t* test. We used robust multivariate linear regression models to account for heteroskedastic errors and presence of outliers.

RESULTS

Participant Characteristics

The 8 participants ranged in age from 80.7 to 92.0 years; on average, men were 2 years younger than women (84.3 vs 86.4). Overall, the participants had a mean age of 85.3 years (SD = 4.1) and a mean weight of 68.6 kg (SD = 14.6). One-half were women and one-quarter owned a pet. Five participants were married. Three used some type of assistive device.

Questionnaire and Sensor Data

Table 1 shows participants' scores on questionnaires and time spent in various activities as measured from the sensors. There was no statistically significant difference between scores on the questionnaires between the start and end of the study period (data not shown); therefore, we used the values at the start of the study in subsequent analyses.

Men and women did not differ significantly with respect to the SF-36 summary score, SF-36 MCS, SF-36 PCS, YPAS (either in total hours or in kilocalories per week), CES-D, or Friendship Scale. Sensor results indicated that men spent 21.3% of their time walking (up, down, or level) and 64.4% stationary (sitting or standing); women spent 20.7% of their time walking and 62.0% stationary.

Table 2 shows pairwise correlations between the subjective questionnaire data and the objective sensor data for behavior and physical activity. Using data from the start-of-study questionnaires in the statistical comparison, all relationships (except for those for the SF-36 PCS) trended in the expected direction, although only some were statistically significant. The correlation for physical activity was not as strong as the correlation for behavior, particularly with the SF-36 (PCS: $r^2 = -0.29$, $P = .49$ vs MCS: $r^2 = 0.86$, $P = .03$). The YPAS results showed little correlation with either the SF-36 summary score ($r^2 = -0.19$, $P = .68$) or the SF-36 PCS ($r^2 = 0.14$, $P = .76$) (data not shown).

In both unadjusted models and adjusted robust regression models, there were significant associations between sensed behaviors and the results from the questionnaires, particularly with adjusted models of mood, mental health, and sociability (Table 3). Model fit was excellent in all cases, with r^2 consistently greater than 0.9 in adjusted models.

The models also generally showed stronger associations between questionnaire and sensor data for behavior than for physical activity, as described above. There were still strong statistical associations between energy expenditure as assessed from the YPAS in kilocalories per week and the weighted physical activity score from the sensors in an adjusted model, however (β coefficient = 95.7, $P = .01$). Such association did not

---

Table 1. Participant Questionnaire and Sensor Data (N = 8)

Assessment	Mean (SD)	Range
Questionnaire[a]		
SF-36 summary score	77.9 (16)	55-97
SF-36 PCS	68.8 (22.5)	35-92
SF-36 MCS	88.0 (10.9)	64-100
CES-D	7.2 (8.2)	0.5-25.0
YPAS, hr	22.6 (14.4)	1.0-42.5
YPAS, kcal/wk	5,980.5 (4,123.5)	180-13,155
Friendship Scale	20.8 (3.2)	14-24
Sensor		
Speaking[b]	20.7 (6.1)	9.6-29.1
Stationary	63.2 (5.7)	54.6-73.4
Walking level	7.1 (5.6)	2.7-18.4
Walking up	5.8 (4.9)	0.7-17.5
Walking down	8.1 (5.1)	2.1-17.5

CES-D = Center for Epidemiologic Studies–Depression; MCS = mental component score; PCS = physical component score; SF-36 = 36-Item Short Form Health Survey; YPAS = Yale Physical Activity Survey.

[a] Scores at the start and end of the study did not differ significantly. Scores shown are from the start of the study.

[b] For this activity, n = 7.

Table 2. Correlations Between Questionnaire and Sensor Data

Questionnaire	Speaking[a] r^2 (P Value)	Physical Activity Score[b] r^2 (P Value)	Walking Level[a] r^2 (P Value)	Stationary[a] r^2 (P Value)	Walking Up[a] r^2 (P Value)	Walking Down[a] r^2 (P Value)
SF-36 MCS	0.86 (.03)	–	–	–	–	–
CES-D	−0.75 (.08)	–	–	–	–	–
Friendship Scale	0.97 (.002)	–	–	–	–	–
YPAS hours	–	0.79 (.02)	−0.50 (.21)	−0.68 (.07)	0.80 (.02)	0.72 (.43)
YPAS kcal/week	–	0.82 (.01)	−0.52 (.18)	−0.60 (.12)	0.82 (.01)	0.65 (.08)
SF-36 PCS	–	−0.29 (.49)	0.21 (.62)	0.47 (.24)	−0.56 (.15)	0.18 (.66)

CES-D = Center for Epidemiologic Studies–Depression; MCS = mental component score; PCS = Physical Component Score; SF-36 = 36-Item Short Form Health Survey; YPAS = Yale Physical Activity Survey.

[a] Fraction of time (speaking) or percentage of time (walking, stationary) spent in this activity.

[b] Weighted using the following weights: 0.6 for stationary, 3 for walking level, 5 for walking up, 1 for walking down, and −1.5 for unclassified.

Table 3. Associations Between Questionnaire and Sensor Data

Questionnaire	Sensor Measure	Unadjusted β Coefficient (95% CI)	P Value	Adjusted β Coefficient (95% CI)	P Value
Behavior					
SF-36 MCS	Speaking	1.29 (−0.46 to 2.62)	.055	1.14 (−0.36 to 2.65)[a]	.08
CES-D	Speaking	−1.14 (−2.42 to 0.13)	.07	−1.23 (−2.38 to −0.77)[a]	.04
Friendship Scale	Speaking	0.58 (0.44 to 0.72)	<.001	0.49 (0.027 to 0.95)[a]	.045
Physical activity					
SF-36 PCS	Physical activity score[c]	−0.32 (−1.27 to 0.63)	.44	0.71 (−0.60 to 2.02)[b]	.14
	Stationary	1.67 (−2.06 to 5.4)	.32	−0.38 (−3.06 to 2.30)[a]	.71
	Walking level	0.76 (−2.51 to 4.03)	.59	0.97 (−2.02 to 3.96)[a]	.42
	Walking up	−2.31 (−4.48 to −0.13)	.04	1.68 (−7.00 to 10.36)[a]	.62
	Walking down	0.73 (−2.36 to 3.81)	.59	−0.37 (−2.32 to 1.59)[a]	.63
YPAS total hours	Physical activity score[c]	0.57 (0.36 to 0.78)	.001	0.31 (−0.17 to 0.79)[b]	.11
	Stationary	−1.57 (−2.68 to −0.47)	.01	−1.30 (−2.52 to −0.07)[a]	.04
	Walking level	−1.18 (−2.46 to 0.091)	.06	−0.40 (−3.77 to 2.97)[a]	.76
	Walking up	2.12 (1.06 to 3.18)	.003	5.06 (−0.31 to 10.43)[a]	.06
	Walking down	1.85 (0.48 to 3.22)	.02	1.98 (0.80 to 3.16)[a]	.01
YPAS kcal/week	Physical activity score[c]	152.3 (97.8 to 206.8)	<.001	95.7 (50.0 to 141.4)[b]	.01
	Stationary	−358.5 (−700.8 to −16.2)	.04	−230.5 (−551.9 to 90.9)[a]	.12
	Walking level	−319.5 (−660.9 to 22.0)	.06	−51.3 (−791.3 to 688.7)[a]	.86
	Walking up	567.9 (269.8 to 865.9)	.003	1,148.4 (263.9 to 2,032.9)[a]	.02
	Walking down	431.9 (37.7 to 826.1)	.04	374.8 (31.3 to 718.3)[c]	.04

CES-D = Center for Epidemiologic Studies–Depression; CI = confidence interval; MCS = mental component score; PCS = physical component score; SF-36 = 36-Item Short Form Health Survey; YPAS = Yale Physical Activity Survey.

Notes: Questionnaire is the dependent variable, and sensor measure is the independent variable. Speaking, stationary, walking flat, walking up, and walking down are percentages of time spent in those activities.

[a] Adjusted for age and sex.

[b] Adjusted for age, sex, marital status, and pet ownership.

[c] Weighted using the following weights: 0.6 for stationary, 3 for walking level, 5 for walking up, 1 for walking down, and −1.5 for unclassified.

---

hold true for YPAS total hours of activity (β coefficient = 0.31, P = .11), although selected sensor measures (time stationary, time walking up, and time walking down) were significantly or marginally associated when examined separately, as shown in Table 3. All unadjusted models for these 3 measures showed significant associations.

Usability Assessment

In the focus group and on the usability questionnaire completed at the end of the study, participants generally expressed frustration with completion of the written questionnaires used to assess behavior and activity, noting their complexity, length, and relevance as primary concerns. On the other hand, most participants noted the sensing device was comfortable, nonobtrusive, and easy to use. There were no complaints about technical challenges from the participants, although a research assistant was on site to assist with morning deployment and evening collection of the sensors, which may have influenced this observation. As the device gave no direct, real-time feedback to users, there was little required interaction with the device.

DISCUSSION

We found that data from mobile sensors for behavior correlated highly with the results obtained with established questionnaires, including measures of depressive symptoms, in older adults. Study participants found the device easy to use, comfortable to wear, and less inconvenient than written questionnaires. This quantitative robustness combined with qualitative acceptance of the technology makes automated inference of behavior using sensing potentially feasible and valid in older populations. The questionnaires measured a broad cognitive, physical, emotional, and social construct of aging; the significant associations we found support further research and development of sensor machine algorithms.

In our comparison with established questionnaires, we found excellent concordance with the CES-D, the mental health dimension of the SF-36, and the Friendship Scale, even after model adjustment. Interestingly, there was less agreement between sensor and questionnaire results for measures of physical activity, a domain currently more commonly studied in mobile devices.[22-24] This finding could be due to errors in the processing of sensed physical activity data (despite

basic calibration at the start of the protocol), inadequate statistical power in our small sample, or inaccurate responses on the questionnaires, as participants expressed frustration with the difficulty of recalling some activities' frequency and intensity. Additional study with more directly observed behavior and a larger sample size is needed to better understand this discrepancy. It is possible that the error lies in the current standards for questionnaires, and not in the sensing of behavior. Indeed, the sensor algorithms have been tested extensively in previous work and have been shown to be very reliable,[21] and unlike the sensing device, questionnaires are subject to temporal recall[25] and social desirability bias.

To our knowledge, there has been little study of the use of mobile sensing to infer sociability and behavior through computer analysis of sensor streams without manual input, with most investigations instead using a mobile device as a questionnaire, or either a diary in a momentary sampling scheme[25,26] or a more traditional diary.[27] A growing body of evidence supports the use of acoustical properties to detect changes in emotional health.[28] Detection of vocal affect, or the signals in speech that are associated with emotion, holds promise as a dimension beyond actual content of a conversation or who the speaker is.[29] Prosodic voice features have been shown to be associated with variations in mood.[30,31] Others have studied the glottal waveform, measured as airflow through the glottis, and the change in its pattern during periods of stress.[32] As technology advances, models utilizing a number of physiologic voice features will become more sensitive and specific to various mental health states without the need for understanding the context of a conversation, thus preserving privacy while enabling a robust method of objectively measuring behavior.

Combinations of sensors have the potential to greatly improve our sensitivity and accuracy in detecting behaviors. Many of these sensors are readily available in mobile phones, particularly the microphone and accelerometer used here, and have enough computational power to do all the real-time processing that was carried out as part of this study. By combining the voice-processing metrics with measures of movement, we believe we can improve on already relatively accurate measurement of behavior. In addition, these measures could monitor subtle changes in well-being as a result of medical or behavioral therapies. When combined with remote or Web-based communication methods integrated into next-generation electronic health records, clinicians could have a rich, objective source of information pertaining to treatment response, collected in a natural environment.

Should our results prove replicable in larger studies, the potential applications for this technology are intriguing. In the research environment, use of mobile phone technology could complement or replace traditional questionnaires as a way to better to collect objective participant data. Specifically, the leveraging of voicing data could advance work on understanding social isolation and subclinical depression in older adults. These data could be used in comparative effectiveness research to understand the potential benefit of new treatments, or be used to improve early detection of these important behavioral issues. Clinically, primary care physicians could use mobile sensing to monitor the effect of treatment, or to identify at-risk populations (such as older adults living alone) and detect changes in their behavior earlier than would occur at office visits. These data could potentially link to electronic health records and be part of a system that warns clinicians of changes in a patient's behavior before it is identified by family or caregivers.

Our study has a number of limitations. As our sample was small, it is difficult to draw conclusions from nonsignificant results as they are likely underpowered. Similarly, this study was conducted among older adults having an average age of 85 years; additional work should be performed in other populations to see if findings are generalizable. Although we did collect a small amount of observed behavior data, we did not perform continuous direct observation. Doing so might have made comparisons with actual behavior easier, but it would have likely biased observations. A repeated-measures design was not used as conditions during the study would not be expected to change over the short time period, and there may be regression to the mean within participants. We did not use exact testing, such as the Fisher exact test, as our data were continuous and not categorical. The sensors had a limited battery life (10 hours/day), and it is possible that we missed some behaviors occurring only after the sensors were removed for the day. Finally, these technologies, although feasible in a small setting, are not developed to the point of large-scale, unmonitored deployments. With rapid progress in mobile software and hardware, however, we believe these technologies will become more widely available for large-scale use. Future work should address issues that would make sensing technology a valuable tool in population-based research.

Despite these limitations, this pilot study demonstrates the power and potential of utilizing commonly available sensors with sophisticated processing techniques to improve the detection of specific physical and behavioral activities. As more people are carrying sensors as part of everyday mobile devices, the potential to detect health problems and monitor treatment could become more efficient and effective.

References

1. Camicioli R, Moore MM, Sexton G, Howieson DB, Kaye JA. Age-related brain changes associated with motor function in healthy older people. *J Am Geriatr Soc*. 1999;47(3):330-334.

2. Sheridan PL, Solomont J, Kowall N, Hausdorff JM. Influence of executive function on locomotor function: divided attention increases gait variability in Alzheimer's disease. *J Am Geriatr Soc*. 2003;51(11):1633-1637.

3. Park S, Jayaraman S. Enhancing the quality of life through wearable technology. *IEEE Eng Med Biol Mag*. 2003;22(3):41-48.

4. Pollack ME. Intelligent technology for an aging population: the use of AI to assist elders with cognitive impairment. *AI Mag*. 2005; 26(2):9-24.

5. Vurgun S, Philpose M, Pavel M. A statistical reasoning system for medication prompting. In: Krumm J, Abowd GD, Seneviratne A, Strang T, eds. *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp '07)*. Berlin, Germany: Springer-Verlag; 2007:1-18.

6. Logan B, Healey J, Philpose M, Munguia-Tapia E, Intille S. A Longitudinal evaluation of sensing modalities for activity recognition. In: Krumm J, Abowd GD, Seneviratne A, Strang T, eds. *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp '07)*. Berlin, Germany: Springer-Verlag; 2007:483-500.

7. Jovanov E, Milenkovic A. A decision support system for automatic screening of non-proliferative diabetic retinopathy. *J Med Syst*. 2011; 35(1):17-24.

8. Klasnja P, Consolvo S, McDonald DW, Landay JA, Pratt W. Using mobile & personal sensing technologies to support health behavior change in everyday life: lessons learned. *AMIA Annu Symp Proc*. 2009;Nov 14:338-342.

9. Fitbit. http://www.fitbit.com/. [Web site]. Accessed Jun 21, 2010.

10. Apple. Apple – Nike + iPod. 2010. http://www.apple.com/ipod/nike/run.html. Accessed Jun 21, 2010.

11. Hynes M, Wang H, Kilmartin L. Off-the-shelf mobile handset environments for deploying accelerometer based gait and activity analysis algorithms. *Conf Proc IEEE Eng Med Biol Soc*. 2009;2009:5187-5190.

12. Baek HJ, Lee HB, Kim JS, Choi JM, Kim KK, Park KS. Nonintrusive biological signal monitoring in a car to evaluate a driver's stress and health state. *Telemed J E Health*. 2009;15(2):182-189.

13. Intille SS. A new research challenge: persuasive technology to motivate healthy aging. *IEEE Trans Inf Technol Biomed*. 2004;8(3):235-237.

14. Intille SS. Ubiquitous computing technology for just-in-time motivation of behavior change. *Stud Health Technol Inform*. 2004;107(Pt 2):1434-1437.

15. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198.

16. Dipietro L, Caspersen CJ, Ostfeld AM, Nadel ER. A survey for assessing physical activity among older adults. *Med Sci Sports Exerc*. 1993;25(5):628-642.

17. SF-36: A community for measuring health outcomes using SF tools. 2008. http://www.sf-36.org. Accessed Jun 2, 2008.

18. Hawthorne G. Measuring social isolation in older adults: development and initial validation of the Friendship Scale. *Soc Indic Res*. 2006;77(3):521-548.

19. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1(3): 385-401.

20. Wyatt D, Choudhury T, Bilmes J. Conversation detection and speaker segmentation in privacy sensitive situated speech data. Paper presented at: Interspeech; August 2007; Antwerp, Belgium.

21. Choudhury T, Consolvo S, Harrison B, et al. The mobile sensing platform: an embedded activity recognition system. *Pervasive Comput*. 2008;7(2):32-41.

22. Consolvo S, McDonald DW, Toscos T, et al. Activity sensing in the wild: a field trial of UbiFit Garden. In: *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human factors in Computing Systems (CHI '08)*. New York, NY: ACM; 2008:1797-1806.

23. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181-188.

24. Tudor-Locke C, Sisson SB, Collova T, Lee SM, Swan PD. Pedometer-determined step count guidelines for classifying walking intensity in a young ostensibly healthy population. *Can J Appl Physiol*. 2005; 30(6):666-676.

25. Wenze SJ, Miller IW. Use of ecological momentary assessment in mood disorders research. *Clin Psychol Rev*. 2010;30(6):794-804.

26. Reid SC, Kauer SD, Dudgeon P, Sanci LA, Shrier LA, Patton GC. A mobile phone program to track young people's experiences of mood, stress and coping. Development and testing of the mobiletype program. *Soc Psychiatry Psychiatr Epidemiol*. 2009;44(6): 501-507.

27. Mattila E, Pärkkä J, Hermersdorf M, et al. Mobile diary for wellness management—results on usage and usability in two user studies. *IEEE Trans Inf Technol Biomed*. 2008;12(4):501-512.

28. France DJ, Shiavi RG, Silverman S, Silverman M, Wilkes DM. Acoustical properties of speech as indicators of depression and suicidal risk. *IEEE Trans Biomed Eng*. 2000;47(7):829-837.

29. Scherer KR. Vocal affect expression: a review and a model for future research. *Psychol Bull*. 1986;99(2):143-165.

30. Cowie R, Douglass-Cowie E. Automatic statistical analysis of the signal and prosodic signs of emotion in speech. In: *Proceedings of the Fourth International Conference on Spoken Language (ICSLP 96)*. Piscataway, NJ: IEEE; 1996:1989-1992.

31. Frick R. Communication emotion: the role of prosodic features. *Psychol Bull*. 1985;97(3):412-429.

32. Moore E II, Clements MA, Peifer JW, Weisser L. Critical analysis of the impact of glottal features in the classification of clinical depression in speech. *IEEE Trans Biomed Eng*. 2008;55(1):96-107.