

## Recruitment and baseline characteristics of the Community of Voices choir study to promote the health and well-being of diverse older adults

Julene K. Johnson<sup>a,b,\*</sup>, Steven E. Gregorich<sup>b,c</sup>, Michael Acree<sup>d</sup>, Anna M. Nápoles<sup>b,c</sup>, Jason D. Flatt<sup>a,b</sup>, Dana Pounds<sup>a</sup>, Alexandria Pabst<sup>e</sup>, Anita L. Stewart<sup>a,b</sup>

<sup>a</sup> Institute for Health & Aging, University of California, San Francisco, CA, USA

<sup>b</sup> Center for Aging in Diverse Communities, University of California, San Francisco, CA, USA

<sup>c</sup> Division of General Internal Medicine, Department of Medicine, University of California, San Francisco, CA, USA

<sup>d</sup> Osher Center for Integrative Medicine, University of California, San Francisco, CA, USA

<sup>e</sup> Department of Cognitive and Information Sciences, University of California, Merced, CA, USA

### ARTICLE INFO

#### Keywords:

Health disparities  
Community-based participatory research  
Music  
Psychometrics  
Recruitment  
Older adults

### ABSTRACT

**Objective:** To describe the recruitment and baseline results of the Community of Voices study that aims to examine the effect of a community choir intervention on the health and well-being of older adults from diverse racial/ethnic and socioeconomic backgrounds.

**Method:** Using community-based participatory research methods, we recruited adults age 60 and over from 12 Administration on Aging-supported senior centers in San Francisco into a 2-arm cluster-randomized controlled trial of the community choir intervention. Multiple outreach methods were used. We tracked outreach, screening, and recruitment metrics and collected demographics and baseline outcomes via community-based, interviewer-administered surveys and performance measures of cognition, physical function, and psychosocial variables.

**Results:** The study contacted 819 individuals, screened 636, and enrolled 390 diverse older adults over a 42-month, phased recruitment period. The mean age was 71.2 (SD = 7.3), and the majority were women. Two-thirds of the sample are non-white, and 20% of participants reported having financial hardship.

**Discussion:** Outreach and recruitment methods used in the Community of Voices trial facilitated enrollment of a large proportion of minority and lower-SES older adults in the final sample. Similar recruitment approaches could serve as a model for recruiting diverse racial/ethnic and socioeconomic older adults into research.

### 1. Introduction

Adults age 65 and over are the fastest growing segment of the American population. By 2033, older adults are projected to represent 20% of the U.S. population, and this will be the first time in history when older adults outnumber persons younger than age 18 [1]. By 2040, nearly half of older adults are expected to come from non-white racial/ethnic backgrounds [2]. Minority older adults are at high risk for poor health outcomes [3–5]. Socioeconomic disparities also disproportionately affect non-white older adults. In 2014, 19% of African Americans, 18% of Latinos, and 15% of Asians age 65 and over lived in poverty, more than twice the rate for older non-Latino whites (8%) [6]. Our current health and social systems are not well prepared for these dramatic changes in demographics. Novel and cost-effective approaches are needed to keep this increasingly diverse older adult population active and engaged. This can include development of interventions

delivered in local community settings that promote their health and well-being.

The challenge is to design and evaluate health promotion interventions that are appealing and appropriate for this increasingly diverse older adult population. Engagement in the arts (e.g., singing, acting, dancing) is a promising approach to help keep older adults active and engaged in meaningful ways and that can be culturally tailored. Two reviews found that that participating in the arts may have multiple, positive health benefits for older adults [7,8]. The studies reviewed document how engaging in the arts may address critical challenges in aging, such as improving quality of life and cognition and reducing loneliness and falls. In addition, arts interventions are relatively low cost to deliver in community settings and can be culturally tailored to take into account the background of participants. Most arts interventions, such as choir singing, involve multiple engagement components (e.g., social, physical, cognitive). There is evidence that such multi-

\* Corresponding author. Institute for Health & Aging, University of California, San Francisco, CA, USA.  
E-mail address: [julene.johnson@ucsf.edu](mailto:julene.johnson@ucsf.edu) (J.K. Johnson).

<http://dx.doi.org/10.1016/j.conctc.2017.09.006>

Received 9 May 2017; Received in revised form 31 August 2017; Accepted 14 September 2017

Available online 18 September 2017

2451-8654/ © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

modal interventions are particularly effective for promoting health [9–12].

To evaluate the effectiveness of such novel interventions for older adults requires the ability to recruit diverse minority and lower-socio-economic status (SES) older adults into research studies. Recruitment of minority and low-SES older adults has been a persistent challenge, and they continue to be underrepresented in health research [13]. Community-based participatory research (CBPR) methods are increasingly being used to help design culturally relevant studies and recruit and enroll persons from underserved populations [14]. CBPR approaches have been found to help recruit and retain diverse participants in research [15,16]. Recruitment of minority individuals into clinical trials using CBPR are typically more successful than studies that do not [17].

The Community of Voices study used CBPR methods to recruit older adults from a broad spectrum of racial/ethnic and SES backgrounds into a cluster-randomized controlled trial designed to examine the effects of a community choir program on cognitive, physical, and psychosocial outcomes [18]. Study recruitment began in February 2012 and was completed by August 2015. The purpose of this manuscript is to summarize the outreach, recruitment, and baseline results of this trial.

## 2. Methods

### 2.1. Overview of study design and intervention

A detailed description of the study protocol is summarized elsewhere [18]. Briefly, 12 senior centers were randomly assigned to either begin the Community of Voices choir intervention immediately or wait six months. Outcome measures were collected at baseline (prior to starting the intervention), 6 months (end of randomization phase), and 12 months (one year after enrollment). The choir intervention was culturally tailored for each senior center based on the ethnic/racial backgrounds and singing experiences of their clientele. It was designed to promote health and well-being by focusing on three hypothesized engagement components (i.e., physical, cognitive, and psychosocial). Participants in the study attended weekly, 90-min choir sessions (44 weeks) led by professional music directors at senior centers. The trial is registered at U.S. National Institutes of Health ([ClinicalTrials.gov](http://ClinicalTrials.gov)): #NCT01869179 (registered 9 January 2013).

### 2.2. Study screening and recruitment

The outreach and recruitment approach was developed and implemented in collaboration with the local Administration on Aging (AoA) and a network of AoA-supported community centers (Department of Aging and Adult Services or DAAS of San Francisco County). DAAS supports a large network of senior centers throughout San Francisco and serves older adults from diverse racial/ethnic and SES backgrounds. We recruited adults aged 60 and older from each of the 12 senior centers. The sampling frame included persons already attending each senior center and those within each of the senior center's geographic service area. Similar to our previous work recruiting ethnically diverse, lower-SES adults [16,19], we used multiple methods (e.g., flyers, presentations, word-of-mouth, and radio) shared with the senior centers and local community organizations, utilized a research team that reflected the diversity of the community (including bilingual and bicultural research associates), and completed all recruitment and screening assessment procedures in the ecommunity at each of the senior centers. The research team spent three months (2–3 days per week) on-site at each senior center to answer questions about the study and complete the screening and enrollment process. The research team passed out flyers and answered questions during events at the senior centers (e.g., food tabling, ongoing classes) and shared information with organizations (e.g., senior housing, libraries) in the geographic service area of each senior center. The study protocol, consent forms, and outcome measures were approved by the Institutional Review

Board at the University of California, San Francisco.

Through these outreach methods, interested individuals were invited to attend an informational meeting about the study at the senior centers or to provide their name and phone number to be called by a staff person to discuss the study over the phone. The choir intervention content, duration, study procedures, risks, and randomization procedures were explained in English or Spanish, and questions were answered. For those who attended an informational meeting and expressed interest in the study, a screening assessment was done in person. For those reached by phone, screening began on the phone using the identical screening script used for the in-person screening assessments except that the cognitive screen was done in-person during the start of the baseline assessment because it required responses on paper.

### 2.3. Eligibility criteria

Inclusion criteria were aged 60 and over, self-report (confirmed by research associates) of adequate visual and hearing acuity (with assistive devices), and English or Spanish fluency sufficient to complete study assessments (self-reported fluency in either language) including bilingual and monolingual Spanish speakers. Exclusion criteria included having a self-reported diagnosis of dementia or Alzheimer's disease or significant cognitive impairment (score of 0 or 1 on the Mini-Cog) [20]; having a serious medical or mental health condition that would limit participation in the study; and planning to move out of the area within 12 months. Persons who were already regularly singing in a choir (e.g., weekly) during the past six months were also excluded.

### 2.4. Demographic and health status variables

For individuals who began the screening interview, we collected five demographic variables: age, sex (male/female), language of screening interview (Spanish/English), ethnicity (Hispanic/Latino or not Hispanic/Latino) and race (American Indian/Alaskan Native, Asian, black/African American, Native Hawaiian/other Pacific Islander, white, and other).

All participants signed and were given a copy of the consent form along with a copy of the Bill of Rights for Research Participants before beginning the study. During the baseline assessment and after written consent was obtained, we collected additional demographic variables, including educational level, marital status (four categories), and nativity (yes/no born in a foreign country). Financial hardship was assessed by asking about problems paying for food, monthly bills, medical visits, or prescribed medications in the past 12 months. Hardship was indicated by endorsing difficulty paying for one or more of the items.

Regarding music background, we asked whether they previously sang in a choir as an adult (yes/no). We also asked participants to rate their overall music ability (poor, fair, good, very good, or excellent) and whether or not they could read music notation (yes/no). Music background was not considered for inclusion in the study, and our goal was to enroll participants with a range of musical abilities.

Self-rated health was assessed using a standard scale (five categories ranging from excellent to poor) and asking whether a doctor or other health professional had told them they had any of 11 different chronic conditions, adapted from the CDC's National Center for Health Statistics. Chronic conditions were categorized into seven physical health conditions (diabetes, heart disease, hypertension, stroke, cancer, arthritis, and emphysema/bronchitis/asthma) and one mental health category representing depression or anxiety. A total count of chronic physical health conditions ranged from 0 to 7. We also asked about health insurance (none, public only, or any private). Sleep quality was assessed using a single question about overall sleep quality in the past month (four categories ranging from very good to very bad) [21].

Three yes/no questions assessed shortness of breath [22]. Based on the published scoring algorithm, scores ranged from 3 to 6 with higher

scores indicating more shortness of breath. To assess voice quality, seven questions were selected from the Voice Handicap Index-10 [23]; we excluded three questions pertaining to voice problems. Participants rated the frequency of specific experiences with their voice on a five-point scale (1 = never to 5 = always). All item stems were framed negatively (e.g., my voice makes it difficult for people to hear me); therefore, we reversed the score so that a higher score indicates better voice quality. The seven-item scale had an internal consistency reliability of .87. These two surveys, in addition to sleep quality, were administered at baseline and follow-up assessments.

Our primary and secondary outcome measures, including the rationale and item descriptions, are described in detail elsewhere [18]. Here, we describe briefly the primary and secondary outcomes and scoring.

## 2.5. Primary outcomes

Our three primary outcomes include one from each of the hypothesized mechanisms of action of the choir intervention: cognitive, physical, and psychosocial engagement.

### 2.5.1. Cognition

The Trail Making Test (TMT), a test of executive function [24,25], was used as the primary cognitive outcome. The TMT has two conditions: Part A (connecting 25 numbers in order as quickly as possible) and Part B (connecting 25 numbers and letters, alternating in order as quickly as possible). We limited the time to complete each part to 180 s to reduce testing burden (instead of the traditional 300 s time limit) [26]. Our final measure is the time (seconds) to complete TMT-B minus the time to complete TMT-A, which is an index of executive function that isolates the executive control component of the TMT [27]; higher scores indicate better executive function. For descriptive purposes, we also report the time in seconds to complete parts A and B and the number of correct lines per minute on the TMT-B to take into account participants who were unable to complete TMT-B in the maximum time [28].

### 2.5.2. Physical function

The primary physical outcome was a timed test of repeated chair stands, one of several tests on the, a performance-based measure of lower body strength from the Short Physical Performance Battery (SPPB) [29]. We recorded the time in seconds to complete five chair stands with a maximum time allowed of 60 s. The possible range is 0–60 s, and higher scores indicate worse functioning (slower performance). In accordance with published guidelines, we also calculated five categories based on seconds to complete (0 = unable, 1 = > 16.7 s, 2 = 16.6–13.7 s, 3 = 13.6–11.2 s, and 4 = < 11.1 s) [29].

### 2.5.3. Psychosocial function

The eight-item Patient Health Questionnaire (PHQ-8) [30] assessed depressive symptoms as the primary psychosocial outcome. Raw scores were used as our primary outcome. The possible range was 0–24, with higher scores indicating more depressive symptoms. The internal-consistency reliability for the entire sample was 0.77. There were some differences by language of interview: Cronbach's alpha was 0.74 for the English speakers and 0.84 for the Spanish speakers. We also report frequencies for the published categories of 1–5 (ranging from no symptoms to severe symptoms) and the percent with a PHQ-8 score > 10, considered to indicate clinically significant depression [30].

## 2.6. Secondary outcomes

We included secondary outcomes for cognition, physical function, and psychosocial function. A majority of the secondary outcomes were from the National Institutes of Health Toolbox for Neurological and Behavioral Function (NIH Toolbox) [31].

### 2.6.1. Cognition

We used the NIH Toolbox Flanker Inhibitory Control and Attention Test, which measures both attention and inhibitory control [32]. Participants completed 20 trials, and the final score reflects a combination of accuracy and reaction time. We selected the Toolbox Computed Score which has a possible range of 0–10 with higher scores indicating better function.

We used a modified version of the NIH Toolbox Rey Auditory Verbal Learning Test as a measure of episodic memory. Fifteen words were presented orally (via computer) over three consecutive trials. After each trial, participants were asked to recall as many words as possible. The learning variable is the sum of correct words for all three trials (possible range 0–45, higher scores indicate better recall). We adapted this test to include two additional components. After completing the three trials, participants listened to 15 new words (distractor list) and asked to recall as many as possible of that list (possible range 0–15, higher scores indicate better recall). Immediately after recall of the distractor list, participants were asked to recall as many of the original 15 words as possible (short-term recall, possible score 0–15). The distractor and delayed recall conditions modifications were suggested by Dan Mungas (personal communication) to include a short delay condition. For this test, the words are different for the Spanish and English tests; thus scores are reported separately by language in addition to the total sample. The secondary outcome variable of interest is the delayed recall trial (after the distractor list).

### 2.6.2. Physical function

We used the NIH Toolbox Balance Accelerometry Measure that assesses static standing balance [33]. Participants assume and maintain five poses. Four involve standing with their feet side-by-side (together) with four variations: 1) eyes open on solid surface, 2) eyes closed on solid surface, 3) eyes open on foam surface, and 4) eyes closed on foam surface. Pose 5 is to stand tandem (one foot aligned in front of the other) on a solid surface with eyes open. Participants were asked to hold each position for 50 s; postural sway is recorded for each pose using an accelerometer worn by the participant. Scoring of each position, based on accelerometer data transmitted to the Assessment Center via computer, involves their ability to hold the position for the time specified and the postural sway. A belt with the accelerometer was placed on the hips [33]. We planned to use the two scores recommended by the Assessment Center (personal communication): 1) ratio of the score for pose 2 divided by the score for pose 1 (ratio 2/1), which reflects the ability to use input from somatosensory and vestibular systems to maintain balance, and 2) ratio of the score for pose 4 divided by pose 1 (ratio 4/1), which reflects the relative reduction in postural stability when visual and somatosensory inputs are simultaneously disrupted (typically representative of the effectiveness of vestibular function for postural control).

To assess gait speed, we used the NIH Toolbox performance measure of gait speed which is a measure of “walking pace” (time to walk 4 m at their usual pace) [42]. Participants completed two timed trials with the score (meters per second) on the fastest trial as the outcome. The possible range was 0.01–2.86 m per second (based on NIH Toolbox norms for all ages) [34] with a higher score indicating better (faster) performance.

### 2.6.3. Psychosocial function

As secondary outcomes, we included six NIH Toolbox measures of emotional well-being [43]. Two of these are computer adaptive testing (CAT) measures based on item response theory (IRT) methods. The other four are short forms in which all participants receive the same items. For all measures we used the T-scores which are normed such that the mean is 50 and the standard deviation is 10.

NIH Toolbox Sadness, a CAT measure, assesses the frequency in the past seven days that participants experienced feelings associated with sadness. There was an item pool of 22 items that pertained to feeling

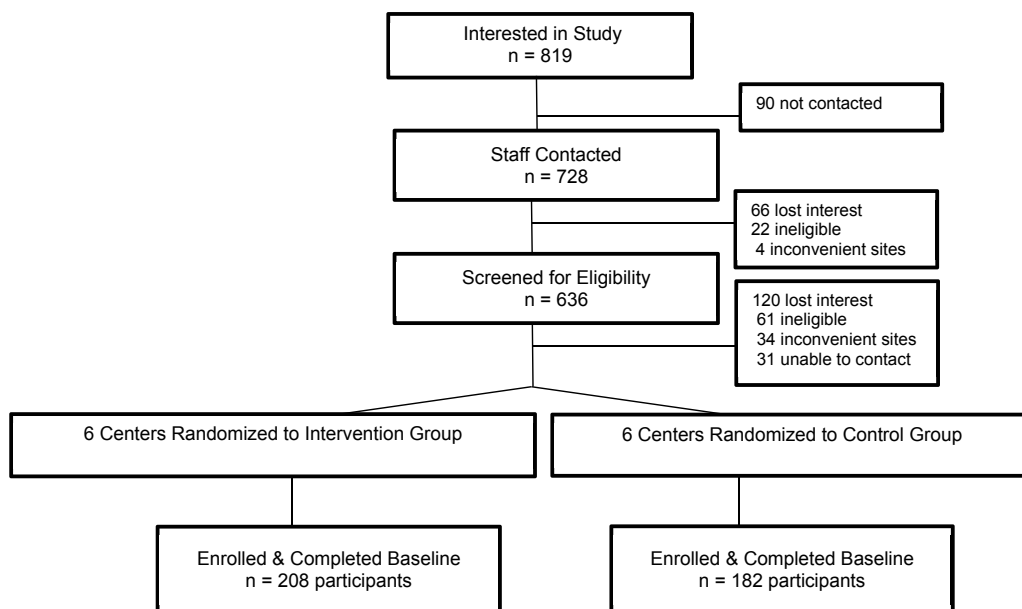


Fig. 1. CONSORT Diagram for the Recruitment and Baseline Assessments in the Community of Voices study.

worthless, helpless, sad, depressed, unhappy, hopeless, had nothing to look forward to, and felt like a failure. Higher scores indicate more sadness.

NIH Toolbox Positive Affect, a CAT measure, assesses the frequency in the past seven days the person experienced positive affect. An item pool of 44 items pertained to feeling cheerful, attentive, delighted, happy, joyful, enthusiastic, interested, peaceful, good natured, useful, understood, content, and liked oneself. Higher scores indicate more positive affect.

NIH Toolbox Fear/Affect is a seven-item short-form measure assessing the frequency in the past seven days the individual felt fearful, anxious, worried, nervous, uneasy, tense, and found it hard to focus. Higher scores indicate more fear/affect. In our own analyses of the seven items, the internal-consistency reliability was 0.91 and was the same for the English and Spanish speakers.

NIH Toolbox Apathy is a six-item short-form measure. Although it is labeled “apathy,” all items pertain to interest in life. Items ask the frequency in the past month the person felt interested in things, got things done, saw a job through, got things started on one's own, did interesting things, and was motivated. In our own analyses of the six items, the internal-consistency reliability was 0.76; it was .78 for the English speakers and .69 for the Spanish speakers. According to the Assessment Center, this score was reversed such that higher scores indicate less apathy (or more interest in things).

NIH Toolbox Loneliness, a five-item short-form measure, assesses the frequency in the past month the person felt alone, apart from others, left out, lonely, and no longer felt close to anyone. Higher scores indicate more loneliness. In our own analyses of the five items, the internal-consistency reliability was 0.88; it was the same for the English and Spanish speakers.

NIH Toolbox Self-Efficacy is a 10-item short-form measure that assesses one's confidence in handling problems. Items pertain to the extent to which the person (e.g., can solve problems, can accomplish goals, feels confident he/she can handle unexpected events, can think of solutions to troubles, and can handle whatever comes his/her way). Higher scores indicate more self-efficacy. In our own analyses the scale had an internal-consistency reliability of 0.88 which was the same in the English and Spanish samples.

Social support. We selected six items from the Medical Outcomes Study (MOS) Social Support Survey [35]. Three items each were chosen from the Positive Social Interaction Scale and the Emotional/Informational Support scale. Items assessed the extent to which participants had

someone to: have a good time with, get together with for relaxation, do something fun with, listen when you need to talk, give you advice if needed, and who understands your problems. Items asked how much of the time each of the kinds of support was available if needed (1 = none of the time, 5 = all of the time). We averaged all items to create a scale ranging from 1 to 5 with higher scores indicating more social support. The scale had an internal-consistency reliability of 0.90 which was similar in English and Spanish samples (0.90 and 0.88 respectively).

### 3. Results

#### 3.1. Screening and recruitment

The study recruitment period was from February 2012 to August 2015. A total of 819 individuals expressed interest in the study by providing their name and phone number, either after an informational meeting or by phone. See consort diagram in Fig. 1. Of these, 728 were contacted in person or by phone. Upon contact, 92 persons had lost interest in the study ( $n = 66$ ) or were deemed ineligible ( $n = 22$ ); four were waiting for a more convenient senior center site. The remaining 636 began the screening process and answered the five demographic questions (i.e., age, sex, language of assessment, race, and ethnicity).

Of the 636 who began the screening interview and answered the demographic questions, 120 (19%) decided not to join the study or complete the screening interview, 61 (10%) were ineligible, and 34 (5%) preferred to enroll at another site, and 31 (5%) were unable to be reached to schedule the baseline assessment or complete the screening. Thus, 390 (61%) enrolled in the study. Randomization were not revealed until after baseline assessments were completed for each pair of centers.

#### 3.2. Sample characteristics

The characteristics of the final enrolled sample ( $N = 390$ ) are described in Table 1. The mean age was 71.3 ( $SD = 7.2$ ), and 76% were female; 65% reported being of non-white (minority) racial/ethnic backgrounds. Only 32% were partnered. A wide range of education was found, with 14% having less than a high school education and 19% having a master's degree or more. Forty-one percent of the sample was foreign born, and 12% of the baseline assessments were completed in Spanish. Financial hardship was reported by 20% of participants. A large number ( $n = 160$ , 41%) reported having only public or no health



**Table 1**  
Sample characteristics.

Characteristic	N (%) <sup>a</sup> or M (SD)
<b>Demographics</b>	
Age (range 59–93) <sup>b</sup>	71.3 (7.2)
Sex (Female)	298 (76)
<b>Race/ethnicity</b>	
Non-Latino white	136 (35)
Black	78 (20)
Asian	104 (27)
Latino	72 (18)
<b>Education</b>	
6th grade or less	24 (6)
7th–12th grade (no diploma)	32 (8)
HS or GED	47 (12)
Some college or AA	85 (22)
Bachelor's degree	129 (33)
Master's degree or more	72 (19)
<b>Marital status</b>	
Married or living with a partner	125 (32)
Single/never married/widowed/divorced	265 (68)
Any financial hardship <sup>c</sup>	77 (20)
<b>Health insurance</b>	
None or public only	160 (41)
Any private (including public + private)	228 (59)
Mini-Cog total score, (range 0–5) <sup>d</sup>	4.3 (1.0)
<b>Music Background</b>	
Previously sung in a choir as an adult	175 (45)
<b>Self-report of musical ability</b>	
Poor	66 (17)
Fair	153 (39)
Good	113 (29)
Very good	46 (12)
Excellent	12 (3)

<sup>a</sup> Percent of non-missing.

<sup>b</sup> 1 person age 59 was included in the study due to incorrect report of age at screening.

<sup>c</sup> Financial hardship = any “yes” to problems paying for food, bills, medical visits or prescriptions.

<sup>d</sup> 2 participants with a Mini-Cog score of 0 were included in the study per PI decision.

insurance. The mean Mini-Cog score was also above the mid-point (4.3 out of 5 points). In terms of prior experience with music, only 45% had previously sung in a choir as an adult, and over half rated their overall musical ability as poor or fair.

**Table 2** summarizes baseline health status. Approximately one-quarter of the participants rated their overall health as poor or fair, and over half had two or more chronic medical conditions; hypertension and arthritis were the most commonly reported chronic condition. About one third of the participants reported at least one shortness of breath symptom, and the mean modified Voice Quality Index score was well above the midpoint (3.3 on a 0–4 scale). About one quarter of the participants rated their overall sleep quality as very bad or fairly bad.

### 3.3. Primary outcomes

**Table 3** summarizes the baseline descriptive statistics on the primary and secondary outcomes measures. For some outcomes, in addition to the analytic outcome measure, we summarize some supplemental measures for descriptive purposes (e.g., categories of severity of depression).

#### 3.3.1. Cognitive function

Based on the Trailmaking Test, the mean of our primary outcome measure of executive function (TMT-B minus TMT-A) was 73.3 s (SD = 38.5). In **Table 2** we also report on the separate conditions. For TMT-A, the mean time to complete was 48.4 s (observed range = 14.3–180, SD = 27.0). For TMT-B, 95 (26%) received the maximum score of 180 and were unable to complete the test in the time allotted. Including these 95 individuals, the mean TMT-B was 121.1 s

**Table 2**  
Baseline health status.

Characteristic	N (%) <sup>a</sup> or M (SD)
<b>Self-rated health</b>	
Excellent	36 (9)
Very good	87 (22)
Good	164 (42)
Fair or poor	103 (26)
<b>Chronic conditions</b>	
Diabetes	76 (21)
Hypertension	204 (55)
Any heart disease	76 (21)
Stroke	27 (7)
Cancer	60 (16)
Arthritis	186 (51)
Emphysema, bronchitis, or asthma	65 (18)
Depression or anxiety	89 (24)
<b>Number of chronic medical conditions</b>	
0	55 (15)
1 or more	313 (85)
2 or more	219 (60)
3 or more	113 (31)
Mean (SD) (range 0–6) <sup>b</sup>	1.9 (1.3)
<b>Frequency of shortness of breath</b>	
No symptoms	268 (69)
1 symptom	93 (24)
2–3 symptoms	29 (7)
<b>Voice quality, (range 0–4)</b>	
3.3 (0.7)	
<b>Sleep quality</b>	
Very bad	14 (4)
Fairly bad	69 (18)
Fairly good	189 (48)
Very good	118 (30)

Notes: M = Mean; SD = standard deviation; HS = High school; GED = General Education Diploma; AA = Associate of Arts degree.

<sup>a</sup> Percent of non-missing.

<sup>b</sup> Includes all chronic conditions except mental health.

(SD = 46.8). For the lines/minute to complete TMT-B, the mean was 13.3 lines per minute (SD = 7.4).

#### 3.3.2. Physical function

In our sample, the observed range for the time to complete five chair stands was 3.9–39.6 s with a mean of 13.1 s (SD = 5.3). Using the categorical measure, 83 (21%) were either unable to complete the chair stands or completed them in more than 16.7 s (categories 0 and 1).

Psychosocial function. On the PHQ-8, we observed nearly the full range (0–23), and the mean was 4.3 (SD = 3.9). Based on the PHQ-8 cut point of > 10 as indicative of possible major depression, 11% of our sample (n = 42) scored as at risk for depression.

### 3.4. Secondary outcomes

**Table 4** summarizes the secondary outcomes.

#### 3.4.1. Cognitive function

In our sample, the observed range of the NIH Toolbox Flanker Inhibitory Control and Attention Test was 2.6–9.4 with a mean of 7.3 (SD = 1.3).

In **Table 2**, we summarize results for the total sample and separately for those completing the NIH Toolbox Rey Auditory Verbal Learning Test in English or Spanish. For the total sample, we observed almost the full range of scores (1–14) with a mean of 5.7 (SD = 2.6). For the English version, we had a mean of 5.5 (SD = 2.6) and for the Spanish version, the mean was 6.6 (SD = 2.6). For descriptive purposes, we also report the results for the distractor list. For the total sample, the range was 1–10 words with a mean of 3.9 (SD = 1.5). Also, for the initial learning trial (immediate recall of the 15 words 3 times), we observed a range of 4–34 and a mean of 18.8 (SD = 5.6).

**Table 3**  
Baseline scores on primary outcome measures.

Variable/measure	N	Direction of high score	Possible range	Observed range	M (SD)
<b>Primary Outcomes</b>					
<b>Cognition</b>					
Trailmaking A: sec	383	Worse	0–180	14.3–180	48.4 (27.0)
Trailmaking B: sec	377	Worse	0–180	39.0–180	121.1 (46.8)
Executive function <sup>a</sup>	373	Worse	na	0–148.3	73.3 (38.5)
Trailmaking B: lines/minute	377	Better	0–60	0.33–36.92	13.3 (7.4)
<b>Physical function</b>					
Chair stands: sec to complete 5	369	Worse	1–60	3.9–39.6	13.1 (5.3)
Chair stands: categories, N (%)	390	Better	0–4	0–4	
Unable					21 (5)
> 16.7 s					62 (16)
16.6–13.7 s					64 (16)
13.6–11.2 s					96 (25)
≤ 11.1 s					147(38)
<b>Psychosocial function</b>					
Depressive symptoms (PHQ-8)	390	Worse	0–24	0–23	4.3 (3.9)
Major depression (PHQ-8 > 10), N (%)	390		0–1		42 (11)

Note: sec = seconds; m/sec = meters per second; na = not applicable; MOS = Medical Outcomes Study; PHQ-8 = Patient Health Questionnaire-8.

<sup>a</sup> Trails B time minus Trails A time.

**3.4.2. Physical function**

We initially planned to use two scores to reflect standing balance (ratio of poses 2/1 and ratio of poses 4/1). However, we had numerous technical problems, including problems with transmission of the accelerometer data to the Assessment Center, which resulted in a large number of missing scores for ratio 4/1 (133 persons or 34%). Therefore, we report only scores for the ratio 2/1. The observed range was 0.6–3.8; the mean was 1.5 (SD = 0.5). Eight percent of the scores (n = 32) were missing for ratio 2/1.

The observed range for walking speed (4-m walk) was 0.20–1.97 m per second, with a mean of 0.86 (SD = 0.25).

**3.4.2.1. Psychosocial function.** The T-scores for the NIH Toolbox Sadness CAT measure in our sample ranged from 34.2 to 69.1 with a mean of 47.8 (SD = 8.4).

For the NIH Toolbox Positive Affect CAT scale, T-scores in our sample ranged from 24.1 to 71.6 with a mean of 49.9 (SD = 9.4).

For the NIH Toolbox Fear/Affect short-form scale, T-scores ranged from 36.3 to 73.6 with a mean of 49.9 (SD = 8.7).

For the NIH Toolbox Apathy measure, the T-scores ranged from 52.8 to 80.6, and the mean was 72.8 (SD = 6.3). Higher scores indicate “less” apathy (contrary to the Toolbox technical manual). This measure was skewed toward higher levels of well-being, given that the mean of a T-score is supposed to be 50.

For the NIH Toolbox Loneliness short-form T score, we observed a mean of 50.9 (SD = 10.3).

The NIH Toolbox Self-Efficacy short form T-score has a mean of 49.5 (SD = 8.8).

For the modified MOS Social Support Scale, the mean was 3.54 (SD = 0.90).

**Table 4**  
Baseline scores on secondary outcome measures.

Variable/measure	N	Direction of high score	Possible range	Observed range	M (SD)
<b>Cognition</b>					
Flanker	381	Better	0–10	2.6–9.4	7.3 (1.3)
RAVLT, total sample	386	Better	0–45	4–43	188.8 (5.6)
English	338			4–34	18.7 (5.7)
Spanish	48			11–33	19.2 (5.1)
Distractor, total sample	378	Better	0–15	1–10	3.9 (1.5)
English	331			1–9	3.9 (1.5)
Spanish	47			1–10	3.6 (1.7)
Delayed recall, total sample	377	Better	0–15	1–14	5.7 (2.6)
English	329			1–14	5.5 (2.6)
Spanish	48			3–13	6.6 (2.6)
<b>Physical Function</b>					
Balance: ratio 2/1	358	Worse	na	0.6–3.8	1.5 (0.5)
Gait speed: m/sec	390	Better	0.01–2.86 <sup>a</sup>	0.20–1.97	0.86 (0.25)
<b>Psychosocial function</b>					
Sadness <sup>b</sup>	390	Worse	na	34.2–69.1	47.8 (8.4)
Positive affect <sup>b</sup>	390	Better	na	24.1–71.6	49.9 (9.4)
Fear/affect <sup>b</sup>	390	Worse	na	36.3–73.6	49.8 (8.7)
Loneliness <sup>b</sup>	390	Worse	na	37.6–82.9	50.9 (10.3)
Apathy <sup>b, c</sup>	390	Better <sup>§</sup>	na	52.8–80.6	72.8 (6.3)
Self-efficacy <sup>c</sup>	390	Better	na	25.1–68.4	49.5 (8.8)
MOS social support	390	Better	1–5	1–5	3.5 (0.9)

Note: sec = seconds; m/sec = meters per second; na = not applicable; RAVLT = Rey Auditory Verbal Learning Test.

<sup>a</sup> Based on minimum and maximum of norms for all ages [36].

<sup>b</sup> T-scores reported for NIH Toolbox measures.

<sup>c</sup> Scoring is reversed from technical manual due to a coding error by the Assessment Center; thus higher Apathy scores indicated less apathy (or more interest in life).

#### 4. Discussion

The Community of Voices study is the largest randomized study to date that aims to examine the effect of a community choir on health and well-being of diverse older adults. Recruitment goals in terms of inclusion of minority individuals and those with low SES were met. The study reached 819 individuals, screened 636, and enrolled 390 older adults. Thus, over half of those screened were enrolled over the 42-month, phased recruitment period. Only 10% of those screened were deemed ineligible, reflecting the study goal of being as inclusive as possible in terms of health status. Importantly, the study met the goal of recruiting minority older adults with 65% of the sample being represented by non-white individuals, which is similar to the proportion of non-whites in San Francisco County and much higher than the national average [37]. Diversity was also reflected in the range of socioeconomic and educational backgrounds, and almost half of the participants were foreign born. These outreach and recruitment goals were achieved through a successful collaboration with a local area agency on aging and by using CBPR methods tailored for diverse communities [16,38].

Because older adults, and in particular minority older adults, are underrepresented in clinical trials, research studies should consider being as inclusive as possible in terms of eligibility criteria. Increasingly, behavioral interventions, such as physical activity and volunteering, offer promise to promote health and well-being on a community level. However, the potential benefit from these types of interventions is left largely unrealized due to low participation rates of older adults from diverse backgrounds. Because interventions involving the arts may be particularly appealing to diverse older adults, recruitment strategies should facilitate their enrollment. Barriers to participation of minority older adults in research include narrow study eligibility criteria, language, transportation, multiple comorbidities, and lower health literacy rates [39]. Partnering with community-based organizations may help mitigate these barriers and help academic researchers design recruitment strategies that work best with minority and underserved populations [19].

The current study is one of the first to enroll a large sample of minority older adults into a trial using participation in the arts (i.e., community choir) to enhance cognitive, physical, and psychosocial engagement, which is hypothesized to promote health and well-being. Cohen and colleagues were the first to complete a large, non-randomized study of a community choir intervention for older adults, compared to a usual activity comparison group [40], but 93% of the participants were white. Coulter and colleagues recently completed a randomized trial of a 14-week choir intervention for older adults [41], but 98% of their participants were also white. These studies are important given the lack of prior studies focused on the arts and aging. However, arts-based studies may be particularly relevant to some underrepresented groups and may be a promising avenue for engaging them in research and, ultimately if efficacious, maintaining their health.

There are several limitations to the current study. The study was one of the early adopters of the NIH Toolbox in both Spanish and English. As noted, we encountered some technical challenges in collecting data at community center sites. Missing data occurred because of internet connectivity issues at the senior center community sites and for the NIH Toolbox Balance measure, transmission errors between the accelerometer and the computer accessing the Assessment Center. Internet connectivity issues affected a few of the tests. The use of legacy measures as primary outcomes was a good decision, as there were no missing data on the primary outcomes. Advantages of using the NIH Toolbox included more efficient administration of the tasks because of CAT and IRT methods and reduced time for data entry.

In conclusion, the outreach and recruitment methods used in the Community of Voices trial facilitated the recruitment and enrollment of a large sample of diverse racial/ethnic and lower-SES older adults, a population that remains underrepresented in health research. The

strategies used included multiple outreach methods, successful collaboration with a local area agency on aging, by using CBPR methods tailored for diverse communities (e.g., broad eligibility criteria, community-based recruitment and assessments in both English and Spanish), and the development of a culturally relevant intervention delivered in the community. Similar recruitment approaches could be used as a model for future studies to help increase enrollment of diverse older adults into research.

#### Acknowledgements

This project is supported by Grant Numbers 5R01AG042526 (Johnson) and P30AG15272 (Napoles) from the National Institute on Aging, National Institutes of Health (NIH) and the National Center for Advancing Translational Sciences, NIH through UCSF-CTSI Grant Number UL1 TR000004. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

We would like to acknowledge all of the Community of Voices research staff who helped with recruitment and collecting baseline data: Jasmine Santoyo-Olsson, Rachel Freyre, Dana Pounds, Merima Rubic, Maria Cora, and Ofelia Villero.

We would also like to acknowledge the San Francisco Department of Aging and Adult Services (Anne Hinton and Shireen McSpadden) and the participating senior centers: Mission Neighborhood Centers (Maria Bermudez), Centro Latino de San Francisco (Gloria Bonilla), Cathy Davis (Dr. George W. Davis Senior Center), Barbara Ockel (Bayview Opera House), Western Addition Senior Center (Robin Bill), San Francisco Senior Center – Aquatic Park (Sue Horst), 30th Street Senior Center (Valorie Villela), OMI Senior Center (Patty Clement-Cihak), Golden Gate Senior Services - Richmond Senior Center (Linda Murley), Golden Gate Senior Services - Castro Senior Center (Patrick Larkin), IT Bookman Community Center (Jackie Wright and Kristin Rosboro), and Veterans Equity Center (Luisa Antonio). We would also like to thank the San Francisco Community Music Center (Sylvia Sherman).

#### References

- [1] U.S. Census Bureau, Profile America Facts for Features, U.S. Census Bureau, Washington, D.C., 2016.
- [2] G.K. Vincent, V.A. Velkoff, The Next Four Decades. The Older Population in the United States: 2010 to 2050, U.S. Census Bureau, Washington, D.C., 2010.
- [3] Federal Interagency Forum on Aging-Related Statistics, Older Americans 2016: Key Indicators of Well-being, Federal Interagency Forum on Aging-Related Statistics, Washington, DC, 2016.
- [4] D.D. Dunlop, J. Song, L.M. Manheim, M.L. Daviglius, R.W. Chang, Racial/ethnic differences in the development of disability among older adults, *Am. J. Public Health* 97 (12) (2007 Dec) 2209–2215 PubMed PMID: 17971548.
- [5] G.H. Louie, M.M. Ward, Socioeconomic and ethnic differences in disease burden and disparities in physical function in older adults, *Am. J. Public Health* 101 (7) (2011 Jul) 1322–1329 PubMed PMID: 21164082.
- [6] Administration on Aging. A Profile of Older Americans: 2015, U.S. Department of Health and Human Services, Washington, D.C., 2015.
- [7] T. Noice, H. Noice, A.F. Kramer, Participatory arts for older adults: a review of benefits and challenges, *Gerontologist* 54 (5) (2014 Oct) 741–753 PubMed PMID: 24336875. Pubmed Central PMCID: 4229893.
- [8] K.D. Fraser, H.M. O'Rourke, H. Wiens, J. Lai, C. Howell, P. Brett-MacLean, A scoping review of research on the arts, aging, and quality of life, *Gerontologist* 55 (4) (2015 Aug) 719–729 PubMed PMID: 26179707.
- [9] V.R. Varma, E.J. Tan, A.L. Gross, G. Harris, W. Romani, L.P. Fried, et al., Effect of community volunteering on physical activity: a randomized controlled trial, *Am. J. Prev. Med.* 50 (1) (2016 Jan) 106–110 PubMed PMID: 26340864. Pubmed Central PMCID: 4691553.
- [10] D.E. Barnes, W. Santos-Modesitt, G. Poelke, A.F. Kramer, C. Castro, L.E. Middleton, et al., The Mental Activity and eXercise (MAX) trial: a randomized controlled trial to enhance cognitive function in older adults, *JAMA Intern. Med.* 173 (9) (2013 May 13) 797–804 PubMed PMID: 23545598.
- [11] A. Karp, S. Paillard-Borg, H.X. Wang, M. Silverstein, B. Winblad, L. Fratiglioni, Mental, physical and social components in leisure activities equally contribute to decrease dementia risk, *Dement. Geriatr. Cogn. Disord.* 21 (2) (2006) 65–73 PubMed PMID: 16319455.
- [12] L.P. Fried, M.C. Carlson, S. McGill, T. Seeman, Q.L. Xue, K. Frick, et al., Experience Corps: a dual trial to promote the health of older adults and children's academic success, *Contemp. Clin. Trials* 36 (1) (2013 Sep) 1–13 PubMed PMID: 23680986. Pubmed Central PMCID: 4112377.

- [13] L. Mody, D.K. Miller, J.M. McGloin, M. Freeman, E.R. Marcantonio, J. Magaziner, et al., Recruitment and retention of older adults in aging research, *J. Am. Geriatr. Soc.* 56 (12) (2008 Dec) 2340–2348 PubMed PMID: 19093934. Pubmed Central PMCID: 2695395.
- [14] L. Michener, J. Cook, S.M. Ahmed, M.A. Yonas, T. Coyne-Beasley, S. Aguilar-Gaxiola, Aligning the goals of community-engaged research: why and how academic health centers can successfully engage with communities to improve health, *Acad. Med.* 87 (3) (2012 Mar) 285–291 PubMed PMID: 22373619. Pubmed Central PMCID: 3292771.
- [15] S. Levkoff, H. Sanchez, Lessons learned about minority recruitment and retention from the centers on minority aging and health promotion, *Gerontologist* 43 (1) (2003 Feb) 18–26 PubMed PMID: 12604742.
- [16] A.M. Napoles, L.A. Chadiha, Advancing the science of recruitment and retention of ethnically diverse populations, *Gerontologist* 51 (Suppl 1) (2011 Jun) S142–S146 PubMed PMID: 21565815. Pubmed Central PMCID: 3092974. Epub 2011/05/20. eng.
- [17] D. De las Nueces, K. Hacker, A. DiGirolamo, L.S. Hicks, A systematic review of community-based participatory research to enhance clinical trials in racial and ethnic minority groups, *Health Serv. Res.* 47 (3 Pt 2) (2012 Jun) 1363–1386 PubMed PMID: 22353031. Pubmed Central PMCID: 3418827.
- [18] J.K. Johnson, A.M. Napoles, A.L. Stewart, W.B. Max, J. Santoyo-Olsson, R. Freyre, et al., Study protocol for a cluster randomized trial of the Community of Voices choir intervention to promote the health and well-being of diverse older adults, *BMC Public Health* 15 (2015) 1049 PubMed PMID: 26463176. Pubmed Central PMCID: PMC3361585.
- [19] J. Santoyo-Olsson, L. Phan, A.L. Stewart, C. Kaplan, G. Moreno-John, A.M. Napoles, A randomized trial to assess the effect of a research informational pamphlet on telephone survey completion rates among older Latinos, *Contemp. Clin. Trials* 33 (4) (2012 Jul) 624–627 PubMed PMID: 22449837. Pubmed Central PMCID: PMC3361585.
- [20] S. Borson, J. Scanlan, M. Brush, P. Vitaliano, A. Dokmak, The mini-cog: a cognitive 'vital signs' measure for dementia screening in multi-lingual elderly, *Int. J. Geriatr. Psychiatry* 15 (11) (2000 Nov) 1021–1027 PubMed PMID: 11113982.
- [21] D.J. Buysse, C.F. Reynolds 3rd, T.H. Monk, S.R. Berman, D.J. Kupfer, The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research, *Psychiatry Res.* 28 (2) (1989 May) 193–213 PubMed PMID: 2748771.
- [22] D.G. Blazer, C.F. Hybels, Shortness of breath as a predictor of depressive symptoms in a community sample of older adults, *Int. J. Geriatr. Psychiatry* 25 (10) (2010 Oct) 1080–1084 PubMed PMID: 20872930.
- [23] C.A. Rosen, A.S. Lee, J. Osborne, T. Zullo, T. Murry, Development and validation of the voice handicap index-10, *Laryngoscope* 114 (9) (2004 Sep) 1549–1556 PubMed PMID: 15475780.
- [24] E. Strauss, E.M.S. Sherman, O. Spreen, *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*, third ed., New York Oxford University Press, 2006.
- [25] R.M. Reitan, Validity of the Trailmaking Test as an indication of organic brain damage, *Percept. Mot. Ski.* 8 (1958) 271–276.
- [26] T.N. Tombaugh, Trail Making Test A and B: normative data stratified by age and education, *Arch. Clin. Neuropsychol.* 19 (2) (2004 Mar) 203–214 PubMed PMID: 15010086.
- [27] I. Sanchez-Cubillo, J.A. Perianez, D. Adrover-Roig, J.M. Rodriguez-Sanchez, M. Rios-Lago, J. Tirapu, et al., Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities, *J. Int. Neuropsychol. Soc.* 15 (3) (2009 May) 438–450 PubMed PMID: 19402930.
- [28] J.H. Kramer, J. Jurik, S.J. Sha, K.P. Rankin, H.J. Rosen, J.K. Johnson, et al., Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease, *Cogn. Behav. Neurol.* 16 (4) (2003 Dec) 211–218 PubMed PMID: 14665820.
- [29] J.M. Guralnik, E.M. Simonsick, L. Ferrucci, R.J. Glynn, L.F. Berkman, D.G. Blazer, et al., A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission, *J. Gerontol.* 49 (2) (1994 Mar) M85–M94 PubMed PMID: 8126356.
- [30] K. Kroenke, T.W. Strine, R.L. Spitzer, J.B.W. Williams, J.T. Berry, A.H. Mokdad, The PHQ-8 as a measure of current depression in the general population, *J. Affect. Disord.* 114 (1–3) (2009 Apr) 163–173 PubMed PMID: WOS:000264223900018. English.
- [31] R.C. Gershon, M.V. Wagster, H.C. Hendrie, N.A. Fox, K.F. Cook, C.J. Nowinski, NIH toolbox for assessment of neurological and behavioral function, *Neurology* 80 (11 Suppl 3) (2013 Mar 12) S2–S6 PubMed PMID: 23479538. Pubmed Central PMCID: 3662335.
- [32] P.D. Zelazo, J.E. Anderson, J. Richler, K. Wallner-Allen, J.L. Beaumont, K.P. Conway, et al., NIH Toolbox Cognition Battery (CB): validation of executive function measures in adults, *J. Int. Neuropsychol. Soc.* 20 (6) (2014 Jul) 620–629 PubMed PMID: 24960301. Pubmed Central PMCID: 4601803.
- [33] R.M. Rine, M.C. Schubert, S.L. Whitney, D. Roberts, M.S. Redfern, M.C. Musolino, et al., Vestibular function assessment using the NIH Toolbox, *Neurology* 80 (11 Suppl 3) (2013 Mar 12) S25–S31 PubMed PMID: 23479540. Pubmed Central PMCID: 3662339.
- [34] M. Kallen, J. Slotkin, J. Griffith, S. Magasi, J. Salsman, C. Nowinski, et al., NIH Toolbox Technical Manual, Motor Domain, Locomotion Subdomain, NIH Toolbox 4-Meter Walk Gait Speed Test, (2012).
- [35] C.D. Sherbourne, A.L. Stewart, The MOS social support survey, *Soc. Sci. Med.* 32 (6) (1991) 705–714 PubMed PMID: 2035047. Epub 1991/01/01. eng.
- [36] J. Slotkin, M. Kallen, J. Griffith, S. Magasi, J. Salsman, C. Nowinski, et al., NIH Toolbox Technical Manual, National Institutes of Health and Northwestern University, 2012.
- [37] United States Census Bureau, U.S. Census Bureau QuickFacts Washington, D.C.: U.S. Census Bureau, [cited 2017 July 1, 2017] (2016) Available from: <https://www.census.gov/quickfacts/fact/table/US/RHI225216>.
- [38] C. Heller, J.E. Balls-Berry, J.D. Nery, P.J. Erwin, D. Littleton, M. Kim, et al., Strategies addressing barriers to clinical trial enrollment of underrepresented populations: a systematic review, *Contemp. Clin. Trials* 39 (2) (2014 Nov) 169–182 PubMed PMID: 25131812.
- [39] A.P. Herrera, S.A. Snipes, D.W. King, I. Torres-Vigil, D.S. Goldberg, A.D. Weinberg, Disparate inclusion of older adults in clinical trials: priorities and opportunities for policy and practice change, *Am. J. Public Health* 100 (Suppl 1) (2010 Apr 1) S105–S112 PubMed PMID: 20147682. Pubmed Central PMCID: 2837461.
- [40] G.D. Cohen, S. Perlstein, J. Chapline, J. Kelly, K.M. Firth, S. Simmens, The impact of professionally conducted cultural programs on the physical health, mental health, and social functioning of older adults, *Gerontologist* 46 (6) (2006 Dec) 726–734 PubMed PMID: 17169928.
- [41] S. Coulton, S. Clift, A. Skingley, J. Rodriguez, Effectiveness and cost-effectiveness of community singing on mental health-related quality of life of older people: randomised controlled trial, *Br. J. Psychiatry* (2015 Jun 18) PubMed PMID: 26089304.
- [42] D.B. Reuben, S. Magasi, H.E. McCreath, R.W. Bohannon, Y.C. Wang, D.J. Bubela, W.Z. Rymer, J. Beaumont, R.M. Rine, J.S. Lai, R.C. Gershon, Motor assessment using the NIH Toolbox, *Neurology* 80 (11 Suppl 3) (2013 Mar 12) S65–S75, <http://dx.doi.org/10.1212/WNL.0b013e3182872e01>.
- [43] J.M. Salsman, Z. Butt, P.A. Pilkonis, J.M. Cyranowski, N. Zill, H.C. Hendrie, M.J. Kupst, M.A. Kelly, R.K. Bode, S.W. Choi, J.S. Lai, J.W. Griffith, C.M. Stoney, P. Brouwers, S.S. Knox, D. Cella, Emotion assessment using the NIH Toolbox, *Neurology* 80 (11 Suppl 3) (2013 Mar 12) S76–S86, <http://dx.doi.org/10.1212/WNL.0b013e3182872e11>.