ORIGINAL RESEARCH

A pilot randomized controlled trial using EEG-based brain–computer interface training for a Chinese-speaking group of healthy elderly

Tih-Shih Lee[1]
Shin Yi Quek[1]
Siau Juinn Alexa Goh[1]
Rachel Phillips[2]
Cuntai Guan[3]
Yin Bun Cheung[4]
Lei Feng[5]
Chuan Chu Wang[3]
Zheng Yang Chin[3]
Haihong Zhang[3]
Jimmy Lee[6]
Tze Pin Ng[5]
K Ranga Rama Krishnan[1]

[1]Department of Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, Singapore; [2]Singapore Clinical Research Institute, Singapore; [3]Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore; [4]Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore; [5]Department of Psychological Medicine, National University of Singapore, Singapore; [6]Department of General Psychiatry/Department of Research, Institute of Mental Health, Singapore

Correspondence: Tih-Shih Lee
Department of Neuroscience and Behavioral Disorders, Duke-NUS Graduate Medical School, 8 College Road, Singapore
Tel +65 6516 7461
Email tihshih.lee@duke-nus.edu.sg

Background: There is growing evidence that cognitive training (CT) can improve the cognitive functioning of the elderly. CT may be influenced by cultural and linguistic factors, but research examining CT programs has mostly been conducted on Western populations. We have developed an innovative electroencephalography (EEG)-based brain–computer interface (BCI) CT program that has shown preliminary efficacy in improving cognition in 32 healthy English-speaking elderly adults in Singapore. In this second pilot trial, we examine the acceptability, safety, and preliminary efficacy of our BCI CT program in healthy Chinese-speaking Singaporean elderly.

Methods: Thirty-nine elderly participants were randomized into intervention (n=21) and wait-list control (n=18) arms. Intervention consisted of 24 half-hour sessions with our BCI-based CT training system to be completed in 8 weeks; the control arm received the same intervention after an initial 8-week waiting period. At the end of the training, a usability and acceptability questionnaire was administered. Efficacy was measured using the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), which was translated and culturally adapted for the Chinese-speaking local population. Users were asked about any adverse events experienced after each session as a safety measure.

Results: The training was deemed easily usable and acceptable by senior users. The median difference in the change scores pre- and post-training of the modified RBANS total score was 8.0 (95% confidence interval [CI]: 0.0–16.0, P=0.042) higher in the intervention arm than waitlist control, while the mean difference was 9.0 (95% CI: 1.7–16.2, P=0.017). Ten (30.3%) participants reported a total of 16 adverse events – all of which were graded "mild" except for one graded "moderate".

Conclusion: Our BCI training system shows potential in improving cognition in both English- and Chinese-speaking elderly, and deserves further evaluation in a Phase III trial. Overall, participants responded positively on the usability and acceptability questionnaire.

Keywords: cognitive training, neuro-feedback, memory, attention

Introduction

There is growing evidence that late-life cognitive activity may have a protective effect on cognition in the elderly. Cohort studies have found a 40%–50% reduction in the risk of dementia for high-level late-life mental activity after controlling for other covariates, including education.[1] Systematic reviews have also demonstrated the beneficial impact of cognitive training (CT) for healthy elderly[2,3] and patients with Alzheimer's disease.[4] Cognitive gains from CT have been reported to last up to 5 years;[5] in fact, a meta-analysis has demonstrated that the protective effects of CT on the cognition of healthy elderly can persist years after training.[6]

However, the bulk of studies examining the utility of CT programs in elderly have been conducted in Western cultures. It is believed that cultural and linguistic factors may influence the impact of CT. For instance, CT programs that are based on mnemonic strategies for the English language may not be applicable to Chinese-speaking populations.[7] Despite the fact that Chinese is the most commonly spoken language in the world,[8] few researchers have developed and examined the use of CT programs for Chinese-speaking elderly. A literature review yielded only five independent studies thus far examining the efficacy of CT programs in elderly populations in Shanghai,[9] Beijing,[10] and Hong Kong.[7,11,12]

Singapore is a multiethnic and multicultural society of over 5 million people. The most common ethnic group in Singapore is Chinese (74.2%),[13] and the most common languages spoken and taught in schools are English and Mandarin. Many citizens are bilingual to a large extent; but invariably, most people prefer to use one language or are more proficient in one language. In Singapore, the most frequently spoken language at home among citizens aged 65 years or above is Chinese; in general, the Chinese-speaking group is 1.5 times larger than the English-speaking group here.[14] Furthermore, proficiency in the English language is associated with higher education level among citizens of Chinese ethnicity. The ratio of those who chose English versus those who chose Chinese as their most frequently spoken home language was 1.2 among university graduates, while this ratio was 0.31 among those who attained the equivalent of 12th grade or lower.

In 2010, we developed an innovative brain–computer interface (BCI) computerized CT program. BCI refers to the interfacing of a computer with signals from the neuronal activity of the brain.[15] Electroencephalography (EEG) waves are captured via two dry electrodes on a headband placed approximately on the frontal (Fp1 and Fp2) positions. Our system then utilizes an algorithm to quantify the user's attention level according to these recorded EEG waves. Our BCI system has showed potential in improving inattention symptoms in attention deficit hyperactivity disorder (ADHD) children.[16,17]

In 2013, we initiated a pilot trial to study the potential of adapting our BCI intervention for improving attention and memory in a group of healthy English-speaking elderly.[18] Our CT program showed promise, particularly in improving immediate memory, delayed memory, attention, visuospatial skills, and global cognitive functioning.

In this study, we replicated our previous study on a sample of healthy, predominantly Chinese-speaking elderly, who outnumbered English-speaking elderly in Singapore and tended to have a lower educational level. This study was done to determine the generalizability of our system and training task to a different linguistic population. In particular, our aim was to:

a) Determine the usability and acceptability of our BCI system to a group of Chinese-speaking elderly;

b) Assess if any concerns are reported by the Chinese-speaking elderly; and

c) Obtain a preliminary efficacy signal in a Chinese-speaking elderly cohort to determine the plausibility of a Phase III trial.

Methods

The study design, methodology, training protocol, and outcome measures used in the present study are identical to that used in our previous study,[18] with the following exceptions: a) participants recruited were more proficient in Chinese and preferred the usage of Chinese instead of English; b) training task instructions and outcome measures, including the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), were translated into Chinese; c) some items used in the translated RBANS were modified so that they would be more culturally relevant to the local population.

Ethics statement

This study was approved by the National University of Singapore Institutional Review Board. Written informed consent from each participant was obtained prior to study admission. The study were carried out in accordance with approved guidelines (Clinicaltrials.gov registration no NCT01661894).

Study design

This study was a double-arm, randomized, waitlist-controlled trial. Intervention consisted of 24 half-hour sessions of our BCI-based CT for 8 weeks (three times a week); the control arm received the same intervention after an initial 8-week waiting period (Figure 1). Potential participants were recruited primarily from the Singapore Longitudinal Aging Study (SLAS), an extensive cohort study of elderly participants in Singapore.[19,20] Additionally, participants were also included by referrals from recruited participants. Participants were qualified for the study if all of the following criteria were met at screening: Chinese ethnicity, predominantly proficient in Chinese, 60–70 years old, Geriatric Depression Scale (GDS) total score of ≤9, mini–mental state examination (MMSE) total score of ≥26, global Clinical Dementia
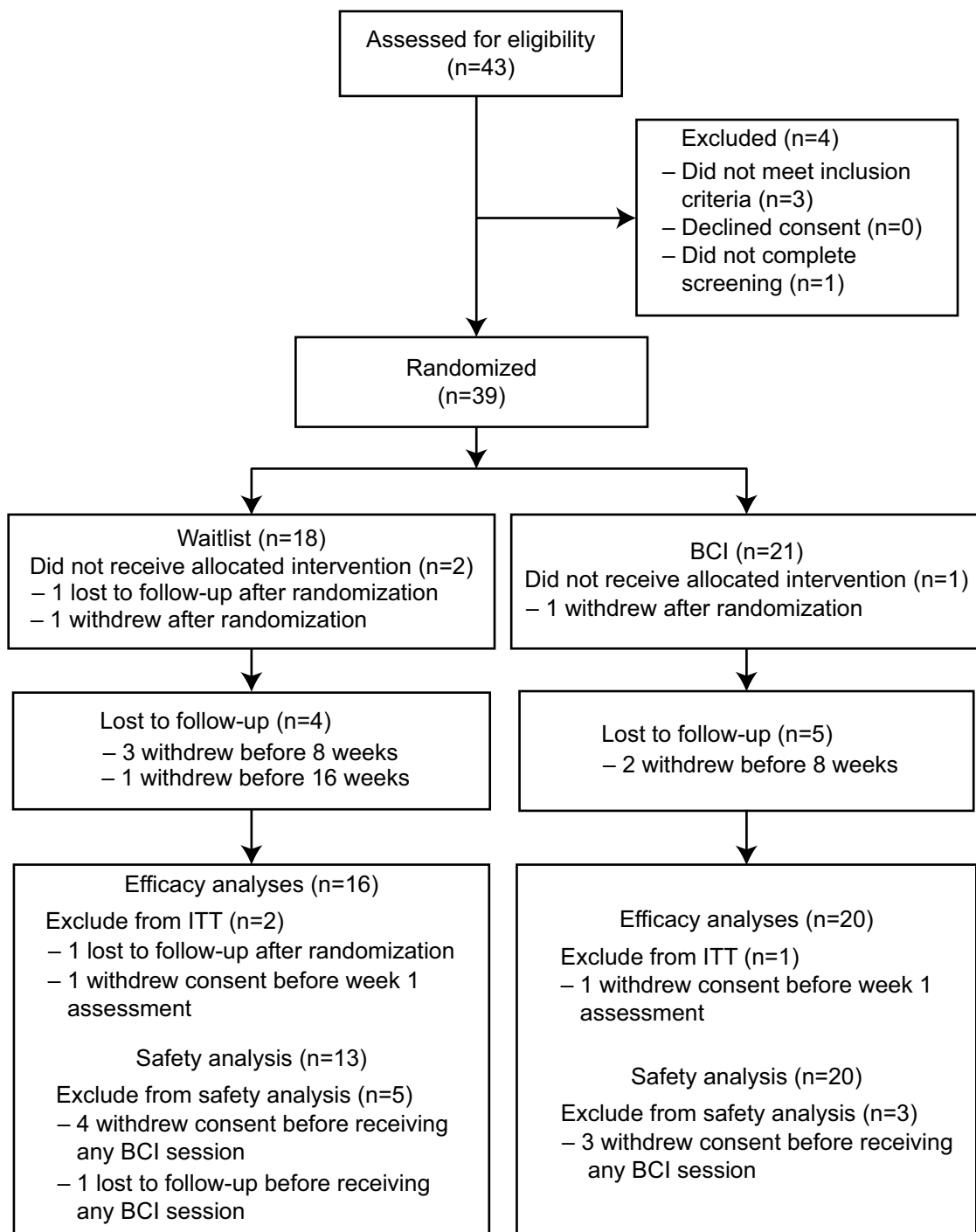
Figure 1 CONSORT flow diagram.
Abbreviations: BCI, brain–computer interface; ITT, intention-to-treat.

Rating (CDR) rating of 0–0.5, could travel to study site without assistance, not diagnosed with neuropsychiatric disorders (such as epilepsy or mental retardation), and were not involved in other clinical research trials at the time of participation (apart from the SLAS). We recruited participants with any level of education – as long as they were literate (able to read and write) in Chinese.

Randomization and masking

Participants were randomized into one of the two arms (the intervention arm or the waitlist control arm) via a password-protected internet-based randomization program. The allocation ratio was 1:1 stratified by education level. Blocking was utilized in the randomization permuted block scheme.[21] A biostatistician determined the block length. However, as per International

Procedure

Prior to the training session, each participant was required to go through a calibration session by answering a series of color Stroop tasks, so that our BCI system could generate an individualized EEG profile that discriminates each participants' attentive and inattentive state.[16–18] During the Stroop tasks, a word that spells the name of a color was displayed on the computer screen. The color used to display the word however, differs from the word (name of a color) (eg the word "blue" displayed in green font). Two choices are then displayed to the user: one choice is the word itself (eg blue) and the other choice is the name of the color used to display the word (eg green). The participant is required to identify the latter rather than the former. Through these tasks, participants are required to be attentive, and thus, their attentive state would be captured through the EEG device. Participants are also given rest periods that last for 7 seconds each. During the rest periods, they are instructed to relax their eyes and be in an inattentive state.

The EEG data collected in the calibration session was then analyzed using our BCI system to identify the EEG parameters. Our BCI system builds a computational model that employs filter banks to cover the full frequency range of EEG rhythms, together with common spatial pattern filtering to determine subject-specific spatial-spectral patterns in the EEG for discriminating attentive and inattentive states.[17] The system then employed a classifier to transform the patterns into a variable that represents the level of attention. This computation model was then employed in the training sessions, where the incoming EEG data was processed and quantified into an attention score at every 200-millisecond interval. This attention score ranged from 0 (low attention) to 100 (high attention), which allowed the participants to control the computer game based on their attention level.[17]

During the training sessions, participants performed a card-matching memory task in which they had to identify matching pairs of cards that show objects of different categories (eg, animals, vegetables, flags).[18] The categories were to provide a variety of stimulus and did not interfere with the difficulty of the game. They were required to focus their attention (sustain an attention score of above 50 for at least 2 seconds) in order to flip the cards; the higher the attention score, the faster they could flip the cards over. For the memory training, participants were required to remember the matching pairs of cards at each level of the game. All cards were revealed at the beginning of the game for a few seconds. Participants were required to remember the matching pairs after all cards were flipped over. They started at Level 1, where there are three matching pairs (six cards). As they proceeded to a more difficult level, the number of matching pairs increased. To proceed to a new level, participants could not exceed a predetermined maximum number of clicks for the existing level. This predetermined numbers of clicks requirement corresponded to the number of pairing mistakes made. For example, Level 1 of the game consists of three matching pairs and the maximum number of clicks is six in order to proceed to Level 2. If a participant made eight clicks in this level, it would mean one pairing mistake was made and he would be required to replay this level until the passing grade was achieved. The difficulty of this game is determined by the number of pairs to be remembered, starting with three pairs. The number of pairs to be remembered increased as participants progressed to higher levels. The highest level achieved in this study was Level 154, with 12 pairs of cards to be remembered.

The intervention arm underwent training from Week 1 to Week 8. Training sessions occurred three times a week, with each session lasting approximately 30 minutes. Training procedures were identical for the waitlist control arm from Week 9 to Week 16.

As our primary efficacy measure, translated and adapted versions of the RBANS were conducted on participants at Weeks 1, 8, and 16 for the intervention arm; and Weeks 1, 9, and 16 for the waitlist control arm.

The first recruitment of the study started on March 2013 and the last follow-up was completed in December 2013.

Outcome measures

Each participant completed a usability and acceptability questionnaire after their last training session (on the 24th session of training). Participants indicated their agreeableness on seven statements on a scale of 1–7 (1= strongly disagree, 7= strongly agree; Table 1). This questionnaire is adapted from IBM's computer usability satisfaction questionnaires.[22] In its original form, the questionnaire is highly reliable to detect usability of the system (r=0.97). Some irrelevant questions (eg, "The information [such as online help, on-screen messages, and other documentation] provided with this system was clear", "The organization of information on the system screens was clear", "This system has all the functions and capabilities I expect it to have") were replaced with more relevant questions (eg, "I will recommend this device to my friends and family", "I think the device is useful in training my memory and attention"). Participants were verbally assessed

Table 1 Descriptive summary of responses to all items in the usability and acceptability questionnaire

Questionnaire item	Mean (SD)	Median (range)
1. Overall, I am satisfied with how easy it is to use this device	6.7 (0.7)	7 (4–7)
2. I feel comfortable using this device	6.4 (0.7)	7 (5–7)
3. I enjoyed playing the game	6.7 (0.7)	7 (4–7)
4. I think the device is useful in training my memory and attention	6.6 (0.8)	7 (4–7)
5. I will recommend this device to my friends and family	6.5 (0.8)	7 (4–7)
6. Overall, I am satisfied with the interface of the game	6.6 (0.8)	7 (3–7)
7. Overall, I am satisfied with the whole system	6.6 (0.7)	7 (4–7)

Abbreviation: SD, standard deviation.

for any discomfort experienced after each training session as a measure for adverse events (AE) or serious AEs (SAE). If participants expressed that they experienced discomfort, the research assistant would proceed to complete a standard AE form by asking the participant further details (Table 1).

Our primary efficacy outcome measure was the modified RBANS. In its original form, RBANS is a comprehensive battery of neuropsychological tests that was developed to assess cognitive status. It is especially sensitive for detecting and characterizing dementia, and comes in four versions or forms.[23] The battery assesses five domains of cognitive function, namely, immediate and delayed memory, language, attention, and visuospatial/construction.[23]

In the current study, we used a version of RBANS A that had previously been translated into Chinese and culturally adapted for the Singaporean geriatric population.[24] We then identified and modified items in the remaining three forms of RBANS that may not be culturally relevant to the local population. For example, in the Story Memory components of the various RBANS forms, the American cities and states featured were changed into Asian cities and countries, such as "Miami, Florida" to "Sumatra, Indonesia" in Form B, and "Chicago, Illinois" to "Hokkaido, Japan" in Form C. This step was undertaken as test administrators reported that a sizeable proportion of participants did not recognize or comprehend some of the more culturally specific terms during administration of RBANS in our previous study on English-speaking participants;[18] they believed that this may have affected the participants' performance. After cultural adaptation of items was completed, RBANS forms B, C, and D were then translated into Chinese by a professional translation company that is unrelated to investigators.

To minimize practice effects, different forms of modified RBANS were counterbalanced for administration at different time-points. Modified RBANS assessments were conducted by study assistants trained in the field of psychology. As this was a small pilot study, we found it impractical to blind these assistants. Therefore, this procedure was not implemented. Nevertheless, RBANS is a manual test with standard administration instructions and objective scoring guidelines, and the lack of blinding was not deemed to pose a potential confound to the results of the study.

The primary endpoints were: a) responses on the usability and acceptability questionnaire; b) safety; and c) changes on the total scale index score of the modified RBANS between Weeks 1 and 8 in the intervention and waitlist control arms.

The secondary endpoints were: a) differences between intervention and waitlist control arms in the changes of the five modified RBANS domain scores from pre- to post-training; b) pooled changes for both arms pre- and post-training for both the modified RBANS total scale index score and the five domains; and c) adherence to training protocol, which was defined as the percentage of participants who finished more than 19 of the 24 (80%) training sessions.

Statistical methods

A precision (width of 95% confidence interval [CI]) of approximately ±13% in the estimation of the proportion of participants who gave positive feedback on acceptability, assuming the true proportion was approximately 80%, required a sample size of 32 participants. We also aimed to estimate the preliminary efficacy of the BCI training system at the end of training to determine if a larger scale study is warranted. For this purpose, Simon's randomized selection design was appropriate. A sample size of 32 participants was also determined to assure an 80% probability of accurately selecting the intervention arm as the superior arm (if it is truly superior) compared to the waitlist control arm by a moderate effect size of 0.3.[25,26] This design was not to confirm the superiority of the intervention arm, but to determine if this arm is worthy of further evaluation in a larger trial.

All data were analyzed using SAS software (v9.3; SAS Institute Inc, Cary, NC, USA). The tests of significance and CIs were based on two-tailed hypotheses. The *P*-value for statistical significance was 0.05 and all CIs were calculated at the 95% level.

The acceptability and safety analysis was conducted on the treated population. The treated population was defined as participants that received at least one session of BCI. We reported the mean and median ratings on the usability and acceptability questionnaire. The evaluation of the treatment safety was based on the AEs reported during the study and numbers and percentages were presented for the pooled data.

All efficacy analyses were performed on the intention-to-treat (ITT) population. The ITT population was defined as follows – as randomized and contactable after randomization. Missing scores were handled using the method of last observation carried forward (LOCF). A per-protocol analysis was also conducted to assess the sensitivity of the ITT results to the method of LOCF. The per-protocol population was defined as follows – as randomized and for whom data concerning efficacy endpoint measures were available.

The differences in the RBANS total scale index scores from Week 1 to Week 8 were compared between the intervention and waitlist control arms using the Mann–Whitney U-test. We estimated the median differences between the two arms and the Hodges–Lehmann CI associated with it. The Wilcoxon signed-rank test was employed to estimate the differences between pre- and post-BCI trainings. For this analysis, data of the five modified RBANS domain scores and the total scale index score from pre- and post-training assessments were combined across arms. Due to the explorative nature of the study, adjustments for multiplicity were not made for the multiple tests in comparisons of RBANS scores.

Participants

Forty-three participants were screened for eligibility, among which three were ineligible for meeting the exclusion criteria and one failed to complete the screening assessment. A total of 39 participants were randomized, with 21 allocated to the intervention group and 18 to the waitlist control group. Two participants withdrew after randomization (one from each arm); neither received any intervention. One participant from the waitlist control arm was lost to follow-up after randomization. This left a total of 36 participants in the ITT population, with 20 in the intervention arm and 16 in the waitlist control arm. All 20 in the intervention arm received treatment while only 13 in the waitlist control arm completed the control period and received treatment.

The mean age of the participants (12 male, 27 female) was 65.2 (standard deviation [SD]: 2.8) years. The majority of participants (84.6%) had an educational attainment of 12th grade or below, with the rest attaining above 12th grade. The average GDS score was 1.7 (SD: 1.6), with participants in the waitlist control arm scoring slightly higher (mean: 1.9, SD: 1) than participants in the intervention arm (mean: 1.5, SD: 1.4). The average total scores for MMSE were similar between arms, with a mean of 27.6 (SD: 1.6). Fifty-one point three percent of the participants reported themselves as being familiar with computers.

Results

Primary outcome measures

Usability and acceptability

The number of participants who completed the usability and acceptability questionnaire was 31. The mean satisfaction rating for all items was ≥6.4 (median =7 for all items) on a scale of 1= "strongly disagree" to 7= "strongly agree". On the last item "Overall, I am satisfied with the whole system", 90.9% (95% CI: 74% to 98%) of participants gave a positive rating of 4 or more (Table 2).

Safety

Ten out of the 33 treated participants (30.3%) reported a total of 16 AEs over the entire duration of the study. There were five participants in the intervention arm reporting six AEs

Table 2 Adverse event form

Item on form	Response options
Adverse event	1= fatigue, 2= seizures, 3= syncope/dizziness, 4= sweating, 5= nausea, 6= headache, 7= others
Status	0= absent, 1= present, 2= NA
CTCAE grade	1= grade 1 (mild), 2= grade 2 (moderate), 3= grade 3 (severe), 4= grade 4 (life-threatening), 5= grade 5 (death)
Date of onset	DD/MM/YYYY
Date stopped (if applicable)	DD/MM/YYYY
Outcome	1= recovered/resolved, 2= recovered/resolved with sequelae, 3= not recovered/not resolved, 4= fatal
Relationship	1= none, 2= unlikely, 3= possible, 4= probable, 5= definite, 6= unknown, 7= NA
Action taken with study treatment	1= none, 2= discontinued, 3= interrupted, 4= modified, 5= interrupted and modified, 6= not applicable
Medication used to treat this adverse event?	Y/N
Serious adverse event?	Y/N

Abbreviations: CTCAE, Common Terminology Criteria for Adverse Events; NA, not applicable; Y, yes; N, no; DD, day; MM, month; YYYY, year.

Table 3 AEs reported and their characteristics

	Number of subjects (N=33)	Number of events
Participants who ever experienced any AE	10 (30.3)	16
Type of AE reported		
Fatigue	3 (9.1)	3
Headache	4 (12.1)	5
Others	2 (6.1)	3[a]
Syncope/dizziness	4 (12.1)	5
Severity, n (%)		
Mild	10 (30.3)	15
Moderate	1 (3.0)	1
Relationship, n (%)		
None	1 (3.0)	1
Unlikely	3 (9.1)	6
Possible	5 (15.2)	5
Probable	2 (6.1)	3
Unknown	1 (3.0)	1
Medication used to treat this AE?, n (%)		
No	10 (30.3)	16
Outcome, n (%)		
Recovered/resolved	9 (27.3)	15
Not recovered/not resolved	1 (3.0)	1
Action taken to study treatment, n (%)		
None	9 (27.3)	15
Interrupted	1 (3.0)	1

Note: [a]The three events reported were: "discomfort in the head but not headache" from one subject, and two reports of "eye pain, tearing" from another subject.
Abbreviation: AE, adverse event.

and five participants in the waitlist control arm reporting ten AEs (Table 3).

The most frequently reported AEs were "headache" and "syncope/dizziness" (four participants reported five events) followed by "fatigue" (three participants reported three events). All AEs were given the lowest severity grading of 1 (mild), except for one report of "others – eye strain, tearing", which was given a severity grading of 2 (moderate) by trained research assistants. All AEs except one had resolved either by the end of the same session that they presented in or by the following day without medication or intervention from the research team. The resolution of one AE of mild fatigue was not verified by the research team due to oversight; the participant had reported coming into the session tired due to poor rest the previous night and subsequently did not report any AE in her following session. One participant was advised by her physician to discontinue participation in the study as she had a chronic health problem of low blood pressure, and her physician deemed it more prudent for her to abstain from participating in any research study.

No serious AEs were reported throughout the course of the study.

Change in modified RBANS total scale index score

The median (range) of the changes between pre- and post-training (Weeks 1 and 8) in the modified RBANS total scale index score was 0.5 (−10 to 29) in the intervention arm

and −1.0 (−20 to 13) in the waitlist control arm. The associated Hodges–Lehmann estimate of the median difference was 8.0 (95% CI: 0.0 to 16.0, P=0.042). The effect size was 0.9 SDs larger than the hypothesized 0.3 SDs. The results did not notably differ in the per-protocol analysis. More detailed results are presented in Table 4.

An exploratory analysis was done to examine the primary outcome according to participants' familiarity with computers. The Hodges–Lehmann estimate of the median difference in the change scores (change in total scale index score on RBANS from Week 1 to Week 8) between arms among subjects who are familiar with computers indicated an increase of 12.0 (95% CI: −2.0 to 26.0, P=0.085) in the intervention arm compared to the waitlist control arm. Among participants who are unfamiliar with computers, the median difference in the change score was 4.0 (95% CI: −5.0 to 14.0, P=0.422) higher in the intervention arm than the control arm.

Secondary outcome measures
Changes in modified RBANS domain scores

The intervention arm showed improvements in the immediate memory, visuospatial/constructional, attention, and delayed memory domains of the modified RBANS as reported by the Hodges–Lehmann estimate of median differences of pre- and post-training score changes between arms (Table 4). However, the difference in the delayed memory domain was

Table 4 A comparison of change in RBANS domain and total scale index scores between Weeks 1 and 8 for the intervention and waitlist control arms

Change in RBANS scores between Weeks 1 and 8	Intervention (n=20)	Waitlist (n=16)	P-value[a]	Median difference (95% CI)[b]
RBANS domain index scores				
Immediate memory				
Mean (SD)	5.7 (14.3)	−2.3 (15.9)		
Median (range)	3.5 (−22 to 29)	0.0 (−32 to 26)	0.129	7.0 (−3.0, 20.0)
Visuospatial/constructional				
Mean (SD)	4.8 (12.9)	−1.3 (10.4)		
Median (range)	5.0 (−17 to 44)	0.0 (−27 to 17)	0.082	6.0 (0.0, 11.0)
Language				
Mean (SD)	0.3 (16.2)	−1.4 (16.1)		
Median (range)	−1.0 (−24 to 36)	0.0 (−39 to 25)	1.000	0.0 (−10.0, 11.0)
Attention				
Mean (SD)	4.1 (10.2)	−0.5 (7.2)		
Median (range)	3.0 (−9 to 34)	0.0 (−10 to 21)	0.127	−4.0 (−10.0, 1.0)
Delayed memory				
Mean (SD)	3.0 (9.9)	−6.4 (11.5)		
Median (range)	0.0 (−19 to 24)	−2.5 (−24 to 9)	0.042	8.0 (0.0, 17.0)
RBANS total scale index score				
Mean (SD)	5.5 (11.8)	−3.5 (8.8)		
Median (range)	0.5 (−10 to 29)	−1.0 (−20 to 13)	0.042	8.0 (0.0, 16.0)

Notes: [a]P-value from the Mann–Whitney U-test; [b]Hodges–Lehmann estimation and its associated 95% CI.
Abbreviations: CI, confidence interval; RBANS, Repeatable Battery for the Assessment of Neuropsychological Status; SD, standard deviation.

the only one to reach statistical significance. The associated Hodges–Lehmann estimate of the median difference was 8.0 (95% CI: 0.0 to 17.0, P=0.042). The results did not notably differ in the per-protocol analysis.

Pooled analysis
The median (range) of the differences in the modified RBANS total scale index score was 1.0 (−15 to 29; Wilcoxon signed-rank test, P=0.039), indicating a statistically significant difference in median score post-training compared to pre-training, with total modified RBANS scores increasing after an 8-week period of BCI (Table 5).

The median of the pre- and post-training differences in delayed memory and attention domain scores were also statistically significant, 1.5 (−12 to 34, P=0.008) and 0.0 (−19 to 24, P=0.039), respectively.

The results did not notably differ in the per-protocol analysis.

Adherence
The proportion of randomized participants that received at least one BCI session was 33/39 (84.6%). The adherence rate among participants who received at least one BCI session was 31/33 (93.9%). Of the 20 in the intervention arm that received treatment, 18 satisfied the ≥80% BCI adherence rate and completed the study; of the 13 in the waitlist control

arm, all 13 satisfied the ≥80% BCI adherence rate and 12 completed the study.

Discussion
Feedback from participants was positive regarding the usability of the system and the perceived efficacy of training, as well as how enjoyable the task was. Participants were also highly motivated to return for training as suggested by the high adherence rate. While AEs were reported, there appears to be no clear indication from the present study that the device poses a risk to users. A majority of participants (70.7%) did not report any AE throughout the entire course of training, and most events reported were of the lowest severity grading of "mild". It is interesting to note that in contrast to the presence of AEs in the present study, there were no AEs reported at all by the English-speaking sample of our previous study using the same system and device. We postulate that this difference could be due to the higher educational attainment (42.9% attained above secondary educational level versus 15.4% in the present cohort) and greater familiarity with computers (80.0% self-reported as being familiar versus 48.7% in the present cohort) of the English-speaking sample. It may be because the educated participants could perhaps better understand the working mechanisms of our system and were more at ease with using it. It is also notable that the most frequent AEs reported were "headache", "syncope/dizziness",

Table 5 Changes in RBANS domain and total scale index scores between pre- and post-intervention, pooling data from the intervention and waitlist control groups

Change in RBANS scores between Weeks 1 and 8	Summary statistics (n=36)	P-value[a]
RBANS domain index scores		
Immediate memory		
Mean (SD)	4.1 (16.0)	
Median (range)	1.5 (−23 to 39)	0.126
Visuospatial/constructional		
Mean (SD)	1.7 (11.3)	
Median (range)	0.0 (−19 to 44)	0.462
Language		
Mean (SD)	0.22 (16.6)	
Median (range)	0.0 (−45 to 36)	0.984
Attention		
Mean (SD)	4.1 (9.3)	
Median (range)	1.5 (−12 to 34)	0.008
Delayed memory		
Mean (SD)	3.1 (9.1)	
Median (range)	0.0 (−19 to 24)	0.039
RBANS total scale index score		
Mean (SD)	4.1 (10.4)	
Median (range)	1.0 (−15 to 29)	0.039

Note: [a]P-value from Wilcoxon signed-rank test.
Abbreviations: RBANS, Repeatable Battery for the Assessment of Neuropsychological Status; SD, standard deviation.

and "fatigue", which could reasonably be due to an extended period of sustained concentration rather than to usage of the device per se. Nevertheless, we will continue to rigorously monitor safety outcomes in our future trials.

In terms of preliminary indications of efficacy, the differences in the change score of the modified RBANS total scale index score from pre- to post-training between the intervention and waitlist control arms was statistically significant. It is remarkable that statistical significance was achieved despite the small sample size of our pilot trial, and can be taken as highly promising evidence for the efficacy of our intervention. As further support for this, differences in the modified RBANS total scale index score pre- and post-training were likewise found to be statistically significant when data from both arms were pooled, with scores improving after training.

It is also notable that differentiated improvements in the modified RBANS domain scores were achieved, with statistically significant improvements found in domains that our intervention was targeting, namely memory and attention. This suggested that our training task was specific and valid. In particular, the intervention arm demonstrated statistically significant improvements in delayed memory pre- and post-training as compared to the waitlist control arm; pooled data showed that improvements in attention and delayed memory pre- and post-training were statistically significant. Improvements in immediate memory and visuospatial/construction

were also found to be positive pre- and post-training, though these did not reach statistical significance. In our previous study of English-speaking elderly, we postulated that though our training did not directly target visuospatial/construction skills, participants could have honed attentiveness to graphic stimuli due to the nature of our training tasks, which rely heavily on visuospatial memories of pictures, resulting in incidental gains in this domain. As expected, there were no changes pre- and post-training in the language domain, as the training task neither targeted nor depended on language. At this point, it may be interesting to note our previous study on English-speaking participants showed a similar trend of results, with pooled data manifesting statistically significant improvements pre- and post-training in all domains except for language. It appears that language does not play a large role in the effect of our CT as long as clear instructions were given in the participants' dominant language. Statistical analysis to compare and contrast the results from English- and Chinese-speaking subjects is not feasible due to the incomparable nature of the modified RBANS outcome measure. Unless validation studies have been done to demonstrate equivalence of these two versions, it is not statistically sound to combine the results into one statistical analysis for comparison.

The main limitation in our study is that the adapted and translated RBANS forms B, C, and D were not validated or normed. Due to the lack of norms, raw test scores of the Chinese-speaking participants in the present study are not

directly comparable to the scores of the English-speaking participants in our previous study. However, RBANS is the only comprehensive neuropsychological assessment tool that consists of parallel versions to minimize practice effect in a repeated measurement setting. We were not able to find validated tools in Mandarin that suited the purpose of our study. While there exists a validated version of form A translated in Shanghai, the People's Republic of China,[27] it is not linguistically appropriate for our local population, as the grammatical structure and vocabulary used are catered toward the mainland Chinese population. There is also a locally translated and adapted version of RBANS, but only form A was translated. This version is normed and validated, and has shown similar validity with the original English RBANS form A in the local population.[24] Thus, we decided on this version and got a local professional translation company to translate forms B, C, and D of RBANS in a culturally relevant manner. The translation process of forms B, C, and D ensured their "face validity".[28] Furthermore, as the four translated forms have been randomized and counterbalanced across participants and time-points, there is little chance of systematic bias due to the assessment tools.

An exploratory analysis revealed familiarity with computers may play a role in the improvements following CT. This factor has not been examined in the literature but it warrants further examination in our upcoming larger trial.

As the purpose of this pilot study was to determine the usability, safety, and preliminary efficacy as an assessment of value for a future larger trial, the relatively small sample size is not considered to be a limitation.

Conclusion

In conclusion, the combined findings of our previous and current pilot studies suggest that: i) our BCI system, device, and training program are deemed acceptable and easily usable by both English-speaking and Chinese-speaking elderly; ii) safety concerns, while reported only by the Chinese-speaking elderly, remain low in frequency and mild in nature; and iii) preliminary efficacy signals are promising. This intervention is deemed suitable to proceed to an efficacy trial with larger statistical power according to Simon's randomized design,[24] which aims to discern if an intervention is worthy of further evaluation. In this larger trial, which is scheduled to begin in 2015, we plan to broaden our sample to include participants with mild cognitive impairment and early dementia. We also plan to include functional magnetic resonance imaging scanning and gene analyses so as to examine

possible changes in the brain that may occur due to CT, as well as identify any potential factors (eg, familiarity with computers) that may aid or hinder individuals in progressing and benefiting through late-life CT. Knowledge of these aspects may aid researchers in designing more personalized and targeted CT interventions for the elderly.

Disclosure

The authors report no conflicts of interest in this work. The BCI technology used in this study is owned by Agency for Science, Technology and Research and Exploit Technologies Pte Ltd, Singapore.

References
1. Valenzuela MJ, Breakspear M, Sachdev P. Complex mental activity and the aging brain: molecular, cellular and cortical network mechanisms. *Brain Res Rev*. 2007;56(1):198–213.
2. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. *PLoS One*. 2012;7(7): e40588.
3. Tardif S, Simard M. Cognitive stimulation programs in healthy elderly: a review. *Int J Alzheimers Dis*. 2011;2011:378934.
4. Sitzer DI, Twamley EW, Jeste DV. Cognitive training in Alzheimer's disease: a meta-analysis of the literature. *Acta Psychiatr Scand*. 2006; 114(2):75–90.
5. Willis SL, Tennstedt SL, Marsiske M, et al; ACTIVE Study Group. Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA*. 2006;296(23):2805–2814.
6. Valenzuela M, Sachdev P. Can cognitive exercise prevent the onset of dementia? Systematic review of randomized clinical trials with longitudinal follow-up. *Am J Geriatr Psychiatry*. 2009;17(3):179–187.
7. Kwok T, Wong A, Chan G, et al. Effectiveness of cognitive training for Chinese elderly in Hong Kong. *Clin Interv Aging*. 2013;8:213–219.
8. Encyclopædia Britannica [webpage on the Internet]. Language. Encyclopædia Britannica Inc.; 2014 [updated January 10, 2014; cited August 7, 2014]. Available from: http://www.britannica.com/EBchecked/topic/329791/language. Accessed August 7, 2014.
9. Cheng Y, Wu W, Feng W, et al. The effects of multi-domain versus single-domain cognitive training in non-demented older people: a randomized controlled trial. *BMC Med*. 2012;10:30.
10. Niu YX, Tan JP, Guan JQ, Zhang ZQ, Wang LN. Cognitive stimulation therapy in the treatment of neuropsychatric symptoms in Alzheimer's disease: a randomized controlled trial. *Clin Rehabil*. 2010;24(12): 1102–1111.
11. Man DW, Chung JC, Lee GY. Evaluation of a virtual reality-based memory training programme for Hong Kong Chinese older adults with questionable dementia: a pilot study. *Int J Geriatr Psychiatry*. 2012; 27(5):513–520.

12. Lee GY, Yip CC, Yu EC, Man DW. Evaluation of a computer-assisted errorless learning-based memory training program for patients with early Alzheimer's disease in Hong Kong: a pilot study. *Clin Interv Aging*. 2013;8:623–633.

13. Department of Statistics, Ministry of Trade and Industry, Republic of Singapore. *Yearbook of Statistics Singapore 2013*. Singapore: The Department; 2013 [cited August 7, 2014]. Available from: http://www.scribd.com/doc/206444277/Yearbook-of-Statistics-Singapore-2013. Accessed August 7, 2014.

14. Department of Statistics, Ministry of Trade and Industry, Republic of Singapore. *Census of Population 2010 Statistical Release 1: Demographic Characteristics, Education, Language and Religion*. Singapore: The Department; 2010 [cited August 7, 2014]. Available from: http://www.singstat.gov.sg/Publications/population.html. Accessed August 7, 2014.

15. Langmoen IA, Berg-Johnsen J. The brain-computer interface. *World Neurosurg*. 2012;78(6):573–575.

16. Lim CG, Lee TS, Guan CT, et al. Effectiveness of a brain-computer interface based programme for the treatment of ADHD: a pilot study. *Psychopharmacol Bull*. 2010;43(1):73–82.

17. Lim CG, Lee TS, Guan CT, et al. A brain-computer interface based attention training program for treating attention deficit hyperactivity disorder. *PLoS One*. 2012;7(10):e46692.

18. Lee TS, Goh AS, Quek SY, et al. A brain-computer interface based cognitive training system for healthy elderly: a randomized control pilot study for usability and preliminary efficacy. *PLoS One*. 2013;8(11):e79419.

19. Ng TP, Feng L, Niti M, Kua EH, Yap KB. Tea consumption and cognitive impairment and decline in older Chinese adults. *Am J Clin Nutr*. 2008;88(1):224–231.

20. Feng L, Ng TP, Chuah L, Niti M, Kua EH. Homocysteine, folate, and vitamin B-12 and cognitive performance in older Chinese adults: findings from the Singapore Longitudinal Ageing Study. *Am J Clin Nutr*. 2006;84(6):1506–1512.

21. Pocock SJ. *Clinical Trials: A Practical Approach*. Chichester, UK: Wiley; 1983:23.

22. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *Int J Hum Comput Interact*. 1995;7(1):57–78.

23. Randolph C. *Repeatable Battery for the Assessment of Neuropsychological Status, manual*. San Antonio, TX: Psychological Corporation; 1998.

24. Lim ML, Collinson SL, Feng L, Ng TP. Cross-cultural application of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): performance of elderly Chinese Singaporeans. *Clin Neuropsychol*. 2010;24(5):811–826.

25. Simon R, Wittes RE, Ellenberg SS. Randomized phase II clinical trials. *Cancer Treat Rep*. 1985;69(12):1375–1381.

26. Liu PY. Phase II selection designs. In: Crowley J, editor. *Handbook of Statistics in Clinical Oncology*. New York, NY: Marcel Dekker; 2001:119–127.

27. Cheng Y, Wu W, Wang J, et al. Reliability and validity of the Repeatable Battery for the Assessment of Neuropsychological Status in community-dwelling elderly. *Arch Med Sci*. 2011;7(5):850–857.

28. Cheung YB. Chapter 12: validity and reliability. *Statistical Analysis of Human Growth and Development*. Boca Raton, FL: Taylor and Francis; 2014.