

Five-year follow-up study of multi-domain cognitive training for healthy elderly community members

Wei FENG¹, Chunbo LI^{2*}, You CHEN³, Yan CHENG¹, Wenyuan WU^{1*}

Background: Cognitive training, a safe non-pharmacological intervention, may help mitigate cognitive decline and prevent the development of dementia in elderly individuals.

Objective: Evaluate the long-term effects of cognitive training among healthy elderly community members.

Methods: Healthy individuals 70 years of age or older from one urban community in Shanghai were screened and the 151 individuals who met inclusion criteria were assigned either to an intervention group ($n=90$) or a control group ($n=61$). The intervention involved twice-weekly training in reasoning, memory, and strategy that continued for 12 weeks (a total of 24 sessions). Participants were assessed at baseline and at 12 weeks, and 5 years after enrollment using the Chinese versions of the Neuropsychological Test Battery for Elderly persons (NTBE), the Stroop Color-Word Test, and a general health questionnaire.

Results: Forty-nine (54%) intervention group subjects and 33 (54%) control group subjects completed the 5-year follow-up. There were few differences in the baseline neurocognitive measures of those who did and did not complete the 5-year follow-up, and there were few differences between those who dropped out of the intervention group compared to those who dropped out of the control group. At the 5-year follow-up, individuals in the intervention group performed better than those in the control group on only 5 measures (in the Trails Making A Test and the Cancellation Test 3) of the 61 measures assessed by NTBE and the Stroop tests, but none of these differences met the pre-determined required level of statistical significance ($p=0.0008$).

Conclusion: We do not confirm the results of previous studies that report long-term benefits of brief cognitive training courses for elderly community residents. Our failure to identify differences in cognitive functioning five years after cognitive training is not likely due to differential dropout between the intervention and control groups but may be related to the relatively small sample and the large number of measures being assessed. Future intervention studies for cognitive training in the elderly should be hypothesis driven (i.e., focused on a single outcome measure of interest), use much larger samples, and include regular booster sessions as part of the cognitive training package.

Key words: aged, cognition, neuropsychological tests, follow up studies, neuropsychology, China

1. Introduction

Signs of cognitive decline, including memory loss, decreased processing speed and difficulty concentrating, are commonly seen among elderly people.^[1] Studies have found that standardized cognitive training can significantly delay cognitive decline and reduce the risk of dementia.^[2-4] Most cognitive training studies in China^[5] have focused on single cognitive domains such as memory, reasoning or processing speed. However, the content of such single-domain training is relatively dull and participants' interest and compliance may diminish after a few sessions. To address this issue,

our team developed an integrated multi-domain cognitive training package tailored for elderly urban community members and administered the three-month intervention to a sample of elderly people in Shanghai in 2006. Previous reports on the study indicated that compared to cognitive functioning in a control group, individuals in the intervention group had better reasoning, memory and executive functioning at the end of the training and that these differences persisted for 1 year after the training.^[6-8] The current paper reports on a 5-year follow-up assessment of the individuals enrolled in this project.

doi: <http://dx.doi.org/10.3969/j.issn.1002-0829.2014.01.005>

¹ Department of Psychiatry, Tongji Hospital, Tongji University, Shanghai, China

² Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

³ Shanghai Yangpu District Mental Health Center, Shanghai, China

*correspondence: Wenyuan WU: wuwu@tongji.edu.cn; Chunbo LI: chunbo_li@yahoo.com

A full-text Chinese translation will be available at www.saponline.org from April 15, 2014.

2. Methods

2.1 Sample

The enrollment and follow-up of subjects for the study is shown in Figure 1. The sample were elderly residents of two neighborhoods of one of the 9 sub-districts of the Putuo District of Shanghai (one of Shanghai's 19 districts). A total of 374 elderly community members from these neighborhoods were recruited by the neighborhood committees (i.e., local administrative offices) and screened from April to May 2006. Inclusion criteria were: (a) at least 70 years of age; (b) ability to self-care with no physical disability or severe physical disease; (c) no mental disorders; and (d) ability to read, write, see, and hear. A total of 151 elderly individuals met these inclusion criteria including 83 males and 68 females; their age ranged from 70 to 89 years with a mean (sd) age of 74.8 (3.7) years.

Recruitment took place at the offices of the neighborhood committee. In order to avoid possible contamination due to communication between participants in the intervention and control groups, assignment to the intervention or control groups was done sequentially. The first 50 screened individuals who met eligibility criteria were asked to participate in the intervention group, the next 50 screened individuals who met eligibility criteria were asked to participate in the control group, the third group of 50 screened who met eligibility criteria were invited to participate in the intervention group, and so forth. Using this process, 90 individuals were recruited in the intervention group and 61 in the control group. There were no significant differences in gender ($\chi^2=1.38$, $p=0.241$), age ($t=0.35$, $p=0.725$), or educational level ($\chi^2=0.39$, $df=3$, $p=0.942$) between the two groups at baseline. This study was approved by the Ethics Committee of Tongji Hospital of Tongji University and all participants provided written informed consent. Five years after the intervention (February to March in 2012), a total of 82 participants were followed-up including 49 in the intervention group and 33 in the control group. There were no significant differences between those who completed the 5-year follow from the intervention and control groups in gender ($\chi^2=1.56$, $p=0.212$), age ($t=-0.05$, $p=0.959$), or educational level ($\chi^2=2.98$, $df=3$, $p=0.395$).

2.2 Assessment tools

The Chinese version of the WHO Neuropsychological Test Battery for Elderly persons (NTBE) was used to evaluate eight domains of cognitive functioning: auditory verbal learning; sorting; cancellation; language; motor functioning; visual function; spatial construction; and trail making.^[9,10] The test-retest correlation coefficients for the auditory verbal learning, cancellation, visual function, spatial construction and trail making subtests ranged from 0.64-0.92; the split-half correlation coefficient was 0.85; and the correlation coefficients between domains of NTBE ranged from 0.10 to 0.42.^[11] The Stroop Color-Word Test was used to assess executive functioning by testing the accuracy and speed of reading words in different colors.^[12,13]

Baseline assessments were conducted from June to September 2006. This included physical examinations, lab tests, NTBE, the Stroop test, and a general health questionnaire.

2.3 Intervention

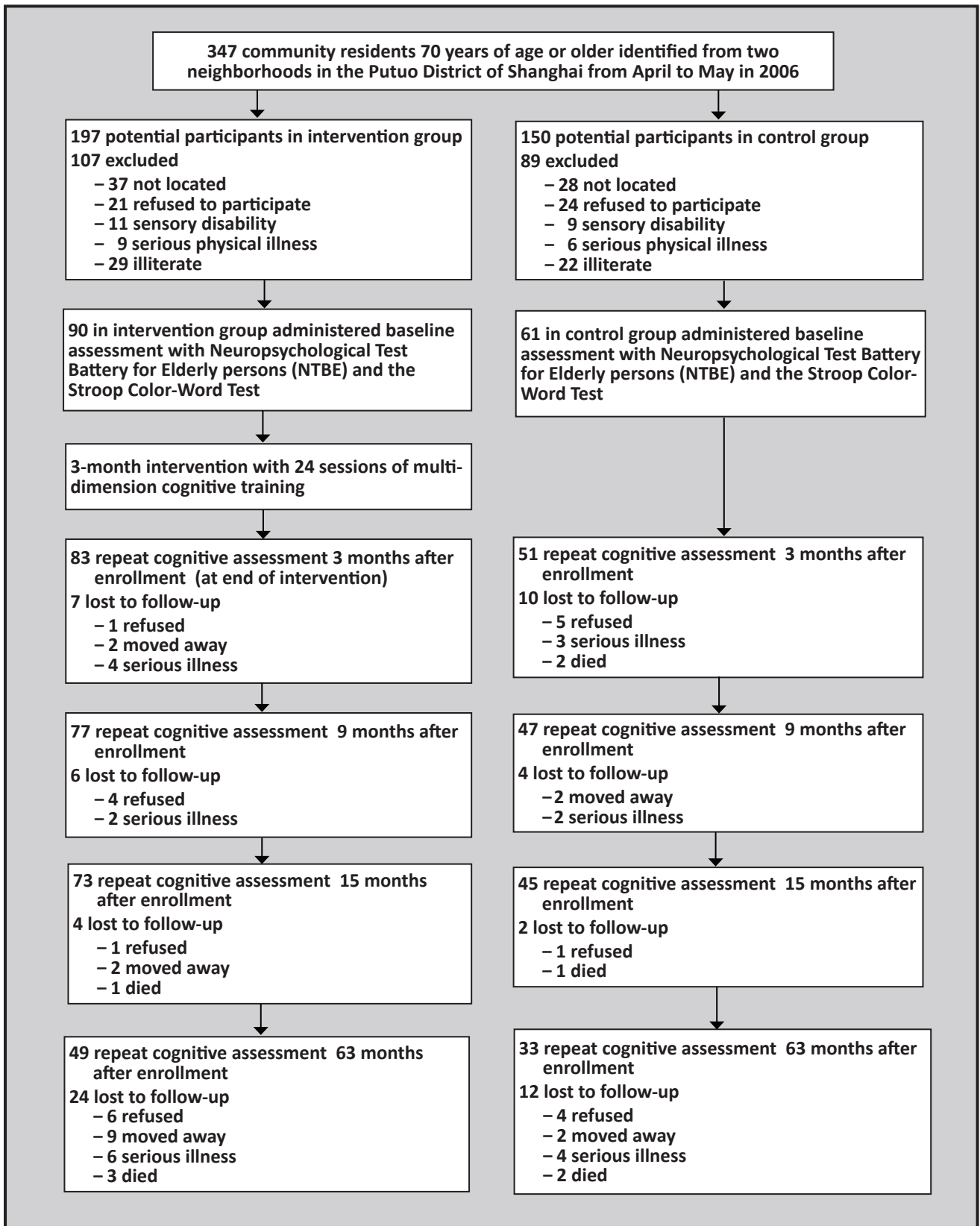
Intervention group members received multi-domain cognitive training from October 2006 to January 2007. Two graduate student psychiatrists provided 24 face-to-face training sessions over this 12-week period to the 90 individuals in the intervention group. The 90 individuals were divided into six groups of 15 individuals each for the training sessions. The length of each session was 60 minutes. Participation in the sessions varied from a high of 97% (87/90) at the first session to a low of 63% (57/90) at the twenty-second session; the mean level of participation over the 24 sessions was 76%. Domains of training included reasoning (i.e., the identification of patterns in a group of words, numbers, or pictures), memory (i.e., memorizing pictures and words), problem solving (i.e., forming strategies for different tasks), and behavioral exercises (i.e., handwriting and handcrafts). Each session covered one domain. After each session participants provided feedback about the difficulty level, perceived usefulness, and interestingness of the session (information that was subsequently used to restructure the sessions). Between training sessions, participants in the intervention group were encouraged to do physical exercise and to finish the homework assigned during the session (including reading, calligraphy, painting, etc.). More details about the training can be found in our previous reports on this project.^[14-16] Individuals in the control group did not receive any cognitive training.

Three months after enrollment (i.e., at the end of the cognitive training in the intervention group) and 9 months, 15 months and 63 months after enrolment all available individuals in the intervention and control groups were re-assessed using the same battery of instruments used at the baseline assessment. These evaluations were conducted by five graduate students in psychiatry who were trained in the use of the instruments and had good inter-rater reliability. The interclass correlation coefficients for the various NTBE sub-tests when these five raters simultaneously assessed four anxious elderly inpatients were all above 0.80. These evaluators were blind to the group membership of the individuals they evaluated.

2.4 Statistical analysis

Epidata 3.0 software was used for data entry and SPSS17.0 software was used for data analyses. Descriptive statistics, chi-squared test, one sample t-test, Mann-Whitney Z-test (continuous data that is not normally distributed), paired t-test and analysis of covariance (ANCOVA) were used depending on the type of data. Three separate analyses were conducted: (a) comparing baseline demographic and neuropsychological test results between the 82 individuals who completed the 5-year follow-up with the 69 who did not complete the 5-year follow-up; (b) comparing

Figure 1. Flowchart of the study



baseline characteristics and neuropsychological test results for the 41 intervention group individuals who did not complete the 5-year follow-up with the 28 control group individuals who did not complete the 5-year follow-up; and (c) comparing the characteristics and neuropsychological test results at baseline (adjusting for age and educational status), and at 3-months post-baseline and 63 months post-baseline (adjusting for age, educational status and baseline value) between the 49 individuals in the intervention group and the 33 individuals in the control group who completed the 5-year follow-up assessment. A total of 61 different measures (59 measures from the subtests on the NTBE and 2 measures from the Stroop Color-Word Test) were assessed at each time interval, so to limit possible bias due to multiple testing, the p -value for statistical significance was set at 0.0008 (i.e., 0.05 / 61).

3. Results

3.1 Comparison of baseline characteristics and neuropsychological test scores between those who did and did not complete the 5-year follow-up

At the time of the 5-year follow-up it was only possible to evaluate 54% (49/90) of the individuals originally enrolled in the intervention group and 54% (33/61) of the individuals originally enrolled in the control group. In the intervention group 12 withdrew consent, 13 moved away (typically to live with children), 12 developed serious physical illnesses that precluded participation, and 4 died. In the control group 10 withdrew consent, 4 moved away, 9 developed serious physical illnesses and 5 died. It is possible that some of these dropouts, particularly those that had serious illnesses or died, were directly or indirectly related to dementia or cognitive decline. This could potentially compromise the comparability of the remaining participants.

To determine whether or not individuals we evaluated five years after the intervention were representative of all enrolled individuals, we compared the baseline demographic characteristic and neuropsychological results of the 82 individuals who completed the 5-year evaluation and the 69 individuals who did not complete the 5-year evaluation. There were no significant differences between these two groups by gender ($\chi^2=1.63$, $p=0.197$), age ($t=1.32$, $p=0.192$), or educational level ($\chi^2=2.05$, $df=3$, $p=0.541$). Five of the 61 neuropsychological measures were different at baseline between those who did and did not complete the 5-year follow-up: compared to those who dropped out during the five years of follow-up, at baseline those who completed the 5-year evaluation had fewer correct responses on Cancellation Test 1 (24.23 [2.76] v. 25.04 [1.49], $t=2.17$, $p=0.032$) and Cancellation Test 2 (10.75 [0.47] v. 10.89 [0.35], $t=2.00$, $p=0.048$), more missing items on Cancellation Test 2 (0.26 [0.47] v. 0.11 [0.35], $Z=2.41$, $p=0.016$), and required more reminders during the Trails Making A test (0.56 [0.97] v. 0.29 [0.66], $Z=2.29$, $p=0.022$) and during the Trails Making B test (1.00 [1.56] v. 0.50 [1.26], $Z=2.11$, $p=0.009$). Given the large number of tests that were compared, none of these differences

were considered statistically significant.

We also compared the baseline characteristics of the 41 individuals who dropped out of the intervention group over the five years with those of the 28 individuals who dropped out of the control group over the five years. There were no significant differences between the groups by gender ($\chi^2=0.15$, $p=0.808$), age ($t=0.57$, $p=0.569$), or educational level ($\chi^2=1.18$, $df=3$, $p=0.241$). Only two of the 61 neuropsychological measures were different at baseline: those who dropped out of the intervention group had more inserted responses in Auditory Verbal Learning Test 1 than those who dropped out of the control group (0.49 [0.75] v. 0.11 [0.32], $Z=2.46$, $p=0.014$), but those who dropped out of the control group had more inserted responses to the Auditory Verbal Learning Test 6 (1.11 [1.23] v. 0.46 [0.81], $Z=2.66$, $p=0.008$). Here, again, given the large number of tests considered, neither of these differences were considered statistically significant.

3.2 Comparison of neuropsychological test results at baseline, at the end of the 3-month intervention and at 5-year follow-up for the 49 intervention-group subjects and 33 control-group subjects who completed the 5-year follow-up

The comparison of the baseline, 3-month and 5-year neuropsychological test results for the 49 individuals from the intervention group and the 33 individuals from the control group who completed the 5-year follow-up are shown in Table 1.

Results for 5 of the 61 assessed measures suggest that individuals in the intervention group who completed the 5-year follow up had somewhat better baseline functioning than individuals in the control group who completed the 5-year follow-up. After controlling for age and educational level, compared to individuals in the control group, at baseline those in the intervention group had more correct responses on Cancellation Test 1 ($p=0.006$), fewer missing items on Cancellation Test 1 ($p=0.005$) and Cancellation Test 3 ($p=0.040$), and better results on the Contact Function Test ($p=0.029$) and the Semantic Relations Test ($p=0.003$). But none of these differences reached our pre-determined level of statistical significance ($p=0.0008$).

At the end of the three-month intervention (or 3 months after enrollment in the control group), after controlling for age, educational level and baseline level of the measure, 6 of the 61 test scores suggested that the cognitive functioning of the intervention group was better than that of the control group: compared to the control group, individuals in the intervention group recalled more items on the Naming Recall Test ($p=0.020$), had a higher score on the Semantic Relations Test ($p=0.020$), performed better on the Visual Matching and Reasoning Test ($p=0.036$), had less duplicated responses in the Auditory Verbal Learning Test 7 ($p=0.024$), had fewer errors on Trails Making Test A ($p=0.033$), and had less color interference in the Stroop Color-Word Test ($p=0.017$).

At the time of the 5-year follow-up, after adjusting for age, educational level and the baseline value for the measure, intervention group individuals performed better than control group individuals on three measures of the Trails Making A Test – fewer errors ($p=0.041$), fewer minor errors ($p=0.026$), and fewer reminders

($p=0.045$). They also performed better on two measures of Cancellation Test 3—more correct responses ($p=0.015$) and fewer missed items ($p=0.018$). None of these differences reached our pre-determined level of statistical significance ($p=0.0008$).

Table 1. Comparison of scores of 59 neuropsychological measures from the Neuropsychological Test Battery for Elderly persons and 2 measures from the Stroop Color-Word Test between the 49 individuals in the cognitive training intervention group and 33 individuals in the control group who completed all three evaluations at baseline, 3 months post-enrollment and 63 months post-enrollment

Item (measure)	time of evaluation	Intervention group mean (sd)	Control group mean (sd)	ANCOVA adjusted for age, education and (for 3 rd and 63 rd month) baseline result F-value (p-value)
Neuropsychological Test Battery for Elderly persons				
Auditory Verbal Learning Test 1 (number of correct responses)	baseline	4.63 (1.62)	4.97 (2.07)	1.45 (0.232)
	3 rd month	5.88 (2.00)	6.15 (1.89)	1.14 (0.289)
	63 rd month	5.98 (2.25)	6.64 (2.40)	1.46 (0.231)
Auditory Verbal Learning Test 1 (number of duplicated responses)	baseline	0.31 (0.77)	0.03 (0.17)	3.89 (0.052)
	3 rd month	0.43 (0.68)	0.48 (0.67)	1.64 (0.204)
	63 rd month	0.47 (0.82)	0.45 (0.87)	0.12 (0.734)
Auditory Verbal Learning Test 1 (number of inserted responses)	baseline	0.51 (1.26)	0.27 (0.52)	1.35 (0.250)
	3 rd month	0.69 (1.00)	0.36 (0.60)	2.34 (0.130)
	63 rd month	0.31 (0.85)	0.30 (0.68)	0.22 (0.645)
Auditory Verbal Learning Test 2 (number of correct responses)	baseline	7.18 (2.10)	6.67 (1.87)	0.52 (0.474)
	3 rd month	8.53 (2.50)	8.18 (2.46)	0.05 (0.832)
	63 rd month	7.61 (2.17)	8.15 (2.61)	2.98 (0.089)
Auditory Verbal Learning Test 2 (number of duplicated responses)	baseline	0.57 (0.82)	0.52 (0.76)	0.00 (0.949)
	3 rd month	0.82 (1.17)	0.94 (1.44)	0.40 (0.528)
	63 rd month	0.61 (1.15)	0.73 (0.98)	0.20 (0.656)
Auditory Verbal Learning Test 2 (number of inserted responses)	baseline	0.29 (0.82)	0.12 (0.42)	1.65 (0.202)
	3 rd month	0.47 (0.84)	0.27 (0.52)	0.75 (0.389)
	63 rd month	0.35 (0.86)	0.24 (0.61)	0.38 (0.541)
Auditory Verbal Learning Test 3 (number of correct responses)	baseline	8.71 (2.18)	8.27 (2.39)	0.08 (0.783)
	3 rd month	9.71 (2.49)	9.61 (2.42)	0.15 (0.698)
	63 rd month	8.63 (2.57)	8.70 (2.47)	0.50 (0.483)
Auditory Verbal Learning Test 3 (number of duplicated responses)	baseline	1.12 (1.09)	0.82 (1.04)	1.41 (0.239)
	3 rd month	1.00 (1.32)	1.27 (1.70)	1.26 (0.264)
	63 rd month	1.06 (1.65)	0.61 (1.04)	1.42 (0.237)
Auditory Verbal Learning Test 3 (number of inserted responses)	baseline	0.14 (0.35)	0.36 (0.65)	3.32 (0.072)
	3 rd month	0.43 (0.91)	0.21 (0.49)	1.62 (0.207)
	63 rd month	0.24 (0.63)	0.27 (0.52)	0.08 (0.776)
Auditory Verbal Learning Test 4 (number of correct responses)	baseline	9.45 (2.32)	9.06 (2.37)	0.18 (0.674)
	3 rd month	10.47 (2.32)	10.27 (2.66)	0.08 (0.776)
	63 rd month	9.45 (2.81)	9.64 (2.50)	1.13 (0.291)
Auditory Verbal Learning Test 4 (number of duplicated responses)	baseline	1.02 (1.41)	1.00 (1.32)	0.00 (0.960)
	3 rd month	1.16 (1.25)	1.45 (1.73)	1.12 (0.293)
	63 rd month	1.02 (1.71)	0.82 (1.04)	0.64 (0.426)
Auditory Verbal Learning Test 4 (number of inserted responses)	baseline	0.29 (0.54)	0.18 (0.39)	1.02 (0.315)
	3 rd month	0.49 (0.82)	0.39 (0.70)	0.57 (0.453)
	63 rd month	0.18 (0.57)	0.33 (0.65)	1.61 (0.208)

Table 1. Comparison of scores of 59 neuropsychological measures from the Neuropsychological Test Battery for Elderly persons and 2 measures from the Stroop Color-Word Test between the 49 individuals in the cognitive training intervention group and 33 individuals in the control group who completed all three evaluations at baseline, 3 months post-enrollment and 63 months post-enrollment (cont'd)

Item (measure)	time of evaluation	Intervention group mean (sd)	Control group mean (sd)	ANCOVA
				adjusted for age, education and (for 3 rd and 63 rd month) baseline result F-value (p-value)
Neuropsychological Test Battery for Elderly persons				
Auditory Verbal Learning Test 5 (number of correct responses)	baseline	10.27 (2.53)	9.73 (2.35)	0.44 (0.510)
	3 rd month	11.12 (2.51)	10.24 (2.59)	0.84 (0.362)
	63 rd month	10.02 (2.78)	9.79 (2.51)	0.02 (0.879)
Auditory Verbal Learning Test 5 (number of duplicated responses)	baseline	1.57 (1.80)	1.33 (1.34)	0.30 (0.586)
	3 rd month	1.80 (2.20)	1.42 (1.68)	0.35 (0.555)
	63 rd month	0.96 (1.26)	0.91 (1.16)	0.08 (0.776)
Auditory Verbal Learning Test 5 (number of inserted responses)	baseline	0.27 (0.61)	0.12 (0.33)	2.42 (0.124)
	3 rd month	0.55 (0.84)	0.36 (0.55)	0.73 (0.396)
	63 rd month	0.29 (0.79)	0.21 (0.55)	0.00 (0.994)
Sorting Test (score)	baseline	9.45 (0.74)	9.58 (0.97)	0.77 (0.384)
	3 rd month	9.71 (0.61)	9.33 (1.02)	3.49 (0.066)
	63 rd month	9.67 (0.66)	9.48 (0.91)	1.12 (0.294)
Cancellation Test 1 (number of correct responses)	baseline	25.43 (1.29)	24.45 (1.60)	7.94 (0.006)
	3 rd month	25.33 (1.31)	24.79 (2.70)	0.19 (0.667)
	63 rd month	24.65 (2.92)	24.24 (3.67)	0.00 (0.985)
Cancellation Test 1 (number of missing items)	baseline	0.55 (1.29)	1.55 (1.60)	8.31 (0.005)
	3 rd month	0.67 (1.31)	1.21 (2.70)	0.18 (0.671)
	63 rd month	1.31 (2.92)	1.55 (3.64)	0.04 (0.840)
Cancellation Test 1 (number of incorrect responses)	baseline	0.06 (0.32)	0.45 (1.92)	2.88 (0.094)
	3 rd month	0.04 (0.29)	0.27 (0.91)	3.46 (0.067)
	63 rd month	0.08 (0.34)	0.30 (0.98)	1.83 (0.180)
Cancellation Test 1 (completion time in seconds)	baseline	56.22 (19.34)	50.61 (15.34)	3.03 (0.086)
	3 rd month	49.78 (15.78)	50.24 (14.15)	0.34 (0.564)
	63 rd month	56.51 (24.40)	54.61 (23.79)	0.31 (0.578)
Cancellation Test 2 (number of correct responses)	baseline	10.90 (0.37)	10.88 (0.33)	0.00 (0.927)
	3 rd month	10.86 (0.41)	10.88 (0.33)	0.22 (0.644)
	63 rd month	11.86 (7.18)	10.55 (1.25)	1.33 (0.252)
Cancellation Test 2 (number of missing items)	baseline	0.10 (0.37)	0.12 (0.33)	0.00 (0.927)
	3 rd month	0.14 (0.41)	0.09 (0.29)	0.80 (0.373)
	63 rd month	0.16 (0.37)	0.45 (1.25)	1.92 (0.170)
Cancellation Test 2 (number of incorrect responses)	baseline	0.20 (1.02)	0.12 (0.55)	0.35 (0.558)
	3 rd month	0.14 (0.61)	0.00 (0.00)	2.22 (0.140)
	63 rd month	0.04 (0.29)	0.18 (0.58)	2.06 (0.155)
Cancellation Test 2 (completion time in seconds)	baseline	27.35 (12.61)	27.39 (11.65)	0.02 (0.891)
	3 rd month	28.82 (11.29)	26.70 (10.57)	1.19 (0.278)
	63 rd month	32.92 (24.58)	29.15 (16.21)	0.71 (0.402)

Table 1. Comparison of scores of 59 neuropsychological measures from the Neuropsychological Test Battery for Elderly persons and 2 measures from the Stroop Color-Word Test between the 49 individuals in the cognitive training intervention group and 33 individuals in the control group who completed all three evaluations at baseline, 3 months post-enrollment and 63 months post-enrollment (cont'd)

Item (measure)	time of evaluation	Intervention group mean (sd)	Control group mean (sd)	ANCOVA
				adjusted for age, education and (for 3 rd and 63 rd month) baseline result F-value (p-value)
Neuropsychological Test Battery for Elderly persons				
Cancellation Test 3 (number of correct responses)	baseline	17.24 (2.18)	15.94 (3.76)	2.74 (0.102)
	3 rd month	17.45 (1.65)	16.70 (2.07)	1.12 (0.294)
	63 rd month	17.35 (2.11)	15.42 (3.15)	6.24 (0.015)
Cancellation Test 3 (number of missing items)	baseline	1.55 (2.04)	3.09 (3.74)	4.36 (0.040)
	3 rd month	1.37 (1.55)	2.30 (2.07)	2.06 (0.156)
	63 rd month	1.51 (1.89)	3.33 (3.11)	5.83 (0.018)
Cancellation Test 3 (number of incorrect responses)	baseline	0.16 (0.83)	0.48 (1.72)	0.73 (0.397)
	3 rd month	0.12 (0.63)	0.21 (1.05)	0.34 (0.562)
	63 rd month	0.12 (0.60)	0.27 (0.91)	0.60 (0.441)
Cancellation Test 3 (completion time in seconds)	baseline	44.51 (14.71)	40.88 (17.00)	1.27 (0.263)
	3 rd month	44.65 (16.29)	41.94 (10.19)	0.27 (0.607)
	63 rd month	51.55 (34.45)	43.21 (19.55)	0.82 (0.368)
Articulation Test (score)	baseline	11.71 (0.74)	11.76 (0.66)	0.17 (0.681)
	3 rd month	11.67 (0.94)	11.82 (0.64)	1.25 (0.266)
	63 rd month	11.80 (0.82)	11.73 (0.88)	0.17 (0.678)
Naming Test (number of items named)	baseline	7.00 (0.00)	6.97 (0.17)	1.81 (0.183)
	3 rd month	6.98 (0.14)	7.00 (0.00)	0.28 (0.597)
	63 rd month	7.00 (0.00)	7.00 (0.00)	---
Naming Recall Test (number of items recalled)	baseline	5.65 (1.25)	5.91 (1.13)	1.54 (0.219)
	3 rd month	6.02 (0.83)	5.52 (0.97)	5.65 (0.020)
	63 rd month	4.76 (1.18)	5.00 (1.23)	1.88 (0.174)
Verbal Fluency Test (number of animals named)	baseline	16.61 (4.84)	15.03 (3.95)	1.64 (0.205)
	3 rd month	16.41 (4.59)	16.18 (4.42)	1.93 (0.169)
	63 rd month	14.73 (4.72)	14.09 (4.34)	0.55 (0.463)
Verbal Fluency Test (number of surnames mentioned)	baseline	17.78 (7.14)	16.18 (5.87)	0.50 (0.483)
	3 rd month	18.18 (6.09)	16.48 (4.95)	0.33 (0.568)
	63 rd month	18.02 (7.01)	16.97 (6.13)	0.15 (0.697)
Verbal Fluency Test (number of vegetables named)	baseline	14.18 (4.18)	14.73 (2.79)	0.34 (0.546)
	3 rd month	14.49 (4.22)	15.58 (3.90)	1.03 (0.314)
	63 rd month	13.82 (4.28)	15.06 (4.46)	1.84 (0.178)
Mini-Token Test (score)	baseline	2.49 (0.94)	2.21 (1.08)	0.36 (0.550)
	3 rd month	2.65 (0.97)	2.48 (0.94)	0.02 (0.890)
	63 rd month	2.51 (0.96)	2.06 (1.00)	2.43 (0.123)
Motor Test (score)	baseline	23.90 (4.70)	24.30 (3.26)	0.43 (0.513)
	3 rd month	25.94 (2.15)	24.73 (5.27)	2.64 (0.108)
	63 rd month	26.22 (3.22)	26.15 (2.15)	0.15 (0.699)
Contact Function Test (score)	baseline	3.51 (0.68)	3.06 (0.83)	4.92 (0.029)
	3 rd month	3.51 (0.65)	3.24 (0.83)	0.24 (0.628)
	63 rd month	3.69 (0.51)	3.45 (0.62)	0.18 (0.675)

Table 1. Comparison of scores of 59 neuropsychological measures from the Neuropsychological Test Battery for Elderly persons and 2 measures from the Stroop Color-Word Test between the 49 individuals in the cognitive training intervention group and 33 individuals in the control group who completed all three evaluations at baseline, 3 months post-enrollment and 63 months post-enrollment (cont'd)

Item (measure)	time of evaluation	Intervention group mean (sd)	Control group mean (sd)	ANCOVA
				adjusted for age, education and (for 3 rd and 63 rd month) baseline result F-value (p-value)
Neuropsychological Test Battery for Elderly persons				
Semantic Relations Test (score)	baseline	3.33 (0.99)	2.55 (1.25)	9.14 (0.003)
	3 rd month	3.37 (0.70)	2.76 (0.97)	5.67 (0.020)
	63 rd month	3.55 (0.65)	3.27 (0.88)	0.90 (0.345)
Recognition (number correctly identified items)	baseline	6.24 (1.27)	5.88 (1.29)	1.49 (0.227)
	3 rd month	6.16 (1.56)	6.00 (1.32)	0.00 (0.991)
	63 rd month	5.57 (1.28)	5.76 (1.15)	0.27 (0.602)
Recognition (number incorrect items identified)	baseline	1.18 (0.91)	1.21 (0.96)	0.00 (0.997)
	3 rd month	1.18 (1.07)	1.09 (0.77)	0.15 (0.702)
	63 rd month	1.37 (1.24)	1.52 (1.23)	0.37 (0.547)
Visual Matching and Reasoning (score)	baseline	5.27 (2.10)	4.48 (1.50)	1.97 (0.165)
	3 rd month	6.18 (1.87)	4.82 (2.14)	4.56 (0.036)
	63 rd month	5.69 (2.07)	4.52 (1.79)	3.21 (0.077)
Auditory Verbal Learning Test 6 (number of correct responses)	baseline	8.94 (3.09)	7.88 (3.25)	1.62 (0.208)
	3 rd month	10.22 (2.93)	9.21 (3.90)	0.16 (0.688)
	63 rd month	9.06 (3.78)	8.39 (3.98)	0.01 (0.913)
Auditory Verbal Learning Test 6 (number of duplicated responses)	baseline	0.92 (1.30)	0.58 (1.00)	1.41 (0.238)
	3 rd month	1.43 (1.88)	1.09 (1.59)	0.16 (0.692)
	63 rd month	1.02 (1.44)	0.91 (1.38)	0.08 (0.784)
Auditory Verb Learning Test 6 (number of inserted responses)	baseline	1.20 (1.58)	0.94 (1.06)	1.53 (0.220)
	3 rd month	1.04 (0.98)	0.88 (1.02)	0.20 (0.658)
	63 rd month	0.78 (1.20)	0.48 (0.67)	0.82 (0.369)
Auditory Verbal Learning Test 7 (number of correct responses)	baseline	3.80 (1.77)	3.61 (1.37)	0.12 (0.727)
	3 rd month	3.98 (1.71)	3.79 (1.93)	0.03 (0.857)
	63 rd month	3.33 (1.55)	2.94 (1.12)	1.32 (0.255)
Auditory Verbal Learning Test 7 (number of duplicated responses)	baseline	0.39 (0.79)	0.30 (0.59)	0.10 (0.757)
	3 rd month	0.24 (0.55)	0.55 (0.97)	5.30 (0.024)
	63 rd month	0.24 (0.60)	0.18 (0.47)	0.46 (0.501)
Auditory Verbal Learning Test 7 (number of inserted responses)	baseline	0.84 (1.21)	1.00 (1.20)	0.24 (0.624)
	3 rd month	0.84 (1.14)	0.82 (0.92)	0.27 (0.607)
	63 rd month	0.88 (1.25)	0.76 (1.17)	0.66 (0.418)
Auditory Verbal Learning Test 8 (number of correct responses)	baseline	7.78 (3.31)	7.33 (3.31)	0.04 (0.839)
	3 rd month	9.39 (3.70)	8.76 (3.89)	0.03 (0.858)
	63 rd month	8.31 (3.56)	8.30 (3.75)	0.29 (0.590)
Auditory Verbal Learning Test 8 (number of duplicated responses)	baseline	0.69 (1.18)	0.48 (0.97)	0.46 (0.501)
	3 rd month	0.71 (1.32)	0.48 (0.83)	0.61 (0.437)
	63 rd month	0.51 (1.21)	0.76 (1.17)	0.69 (0.408)
Auditory Verbal Learning Test 8 (number of inserted responses)	baseline	1.24 (1.69)	0.79 (0.82)	2.90 (0.092)
	3 rd month	0.98 (1.01)	0.94 (1.03)	0.00 (0.949)
	63 rd month	0.84 (1.43)	0.48 (0.76)	0.75 (0.391)

Table 1. Comparison of scores of 59 neuropsychological measures from the Neuropsychological Test Battery for Elderly persons and 2 measures from the Stroop Color-Word Test between the 49 individuals in the cognitive training intervention group and 33 individuals in the control group who completed all three evaluations at baseline, 3 months post-enrollment and 63 months post-enrollment (cont'd)

Item (measure)	time of evaluation	Intervention group mean (sd)	Control group mean (sd)	ANCOVA adjusted for age, education and (for 3 rd and 63 rd month) baseline result F-value (p-value)
Neuropsychological Test Battery for Elderly persons				
Spatial Construction Test (score)	baseline	5.45 (1.12)	5.27 (1.10)	0.17 (0.678)
	3 rd month	5.47 (0.83)	5.03 (1.38)	1.71 (0.194)
	63 rd month	5.61 (0.76)	5.30 (1.13)	0.97 (0.328)
Trails Making Test A (completion time in seconds)	baseline	114.49 (58.32)	123.39 (47.90)	0.05 (0.828)
	3 rd month	101.65 (41.39)	114.73 (48.99)	0.40 (0.530)
	63 rd month	107.84 (54.40)	120.12 (56.80)	0.24 (0.627)
Trails Making Test A (number of errors)	baseline	0.47 (1.17)	0.52 (0.91)	0.08 (0.784)
	3 rd month	0.29 (0.68)	1.12 (2.12)	4.73 (0.033)
	63 rd month	0.08 (0.28)	0.36 (0.82)	4.32 (0.041)
Trails Making Test A (number of times needed reminding)	baseline	0.31 (0.68)	0.27 (0.63)	0.24 (0.627)
	3 rd month	0.31 (0.62)	0.39 (0.66)	0.02 (0.881)
	63 rd month	0.19 (0.67)	0.61 (1.22)	4.16 (0.045)
Trails Making Test A (number of minor errors)	baseline	0.16 (0.43)	0.14 (0.50)	0.38 (0.541)
	3 rd month	0.16 (0.37)	0.15 (0.36)	0.05 (0.818)
	63 rd month	0.04 (0.20)	0.30 (0.68)	5.14 (0.026)
Trails Making Test B (completion time in seconds)	baseline	207.02 (98.69)	221.79 (79.56)	0.02 (0.891)
	3 rd month	186.51 (83.60)	204.79 (82.90)	0.18 (0.670)
	63 rd month	213.49 (146.72)	233.88 (109.82)	0.02 (0.888)
Trails Making Test B (number of errors)	baseline	0.82 (1.56)	0.67 (1.53)	0.64 (0.425)
	3 rd month	0.37 (0.86)	0.82 (1.21)	3.10 (0.082)
	63 rd month	0.33 (1.00)	0.79 (1.80)	1.31 (0.257)
Trails Making Test B (number of times needed reminding)	baseline	0.57 (1.51)	0.39 (0.75)	0.92 (0.341)
	3 rd month	0.35 (0.75)	0.73 (1.07)	3.43 (0.068)
	63 rd month	0.77 (1.33)	1.18 (1.29)	1.74 (0.192)
Trails Making Test B (number of minor errors)	baseline	0.22 (0.55)	0.18 (0.47)	0.68 (0.412)
	3 rd month	0.20 (0.41)	0.21 (0.42)	0.03 (0.869)
	63 rd month	0.27 (0.61)	0.36 (0.82)	0.07 (0.788)
STROOP COLOR-WORD TEST				
Color Interference (score)	baseline	47.00(26.77)	47.39 (21.04)	1.41 (0.239)
	3 rd month	39.55 (14.71)	45.03 (24.68)	5.94 (0.017)
	63 rd month	49.52 (20.71)	45.79 (35.65)	1.96 (0.166)
Word Interference (score)	baseline	18.08 (13.55)	15.09 (15.69)	0.04 (0.840)
	3 rd month	20.22 (13.35)	23.91 (14.89)	0.89 (0.349)
	63 rd month	25.58 (17.75)	29.00 (21.12)	3.74 (0.057)

4. Discussion

4.1 Main findings

Among individuals who were followed up five years after enrollment, after adjusting for baseline cognitive functioning, age and education there were no differences on a wide range of neurocognitive tests between those who had received a three-month cognitive training program and those who had not received the training. These results are based on 54% follow-up of the original sample so the failure to find a difference could be due to differential dropout from the two groups. But the proportion of dropouts was identical in the two groups (46%) and we found few differences in the baseline characteristics and neuropsychological profile of those who completed the study versus those who dropped out. Moreover, there were also few differences between those who dropped out from the intervention group versus those who dropped out of the control group. These results strongly suggest that the lack of differences in neuropsychological functioning five years after the three-month cognitive training course is real; it is not likely due to differential dropout rates in the two groups.

Thus our results do not confirm results from other countries that report long-term effectiveness of cognitive training. For example, Ball and colleagues randomly assigned 2832 community members aged 65 to 94 years into a memory training group, a reasoning training group, a processing speed training group and a control group; after 6 weeks of cognitive training, scores on corresponding cognitive functioning tests were improved and participants' cognitive functioning and daily functioning were significantly better than those of the control group five years after the intervention.^[17-19]

Why were we unable to replicate Ball's findings? The sample size of Ball's study was 18-fold larger than the sample size in our study so his study had the statistical power to identify small differences that we could not identify. Ball's study had younger participants (starting at 65 years of age while our study started at 70 years of age) so it's possible that a smaller proportion of the subjects in his study were affected by age-related decline (i.e., mild cognitive impairment) that would swamp the positive effects of a short cognitive training program. Most importantly, 60% of the participants in Ball's study were given reinforcement training 11 months and 35 months after the initial cognitive training, limiting the attenuation of the training effect over time; we did not conduct any booster sessions over the five-year follow-up period. The younger age of participants and use of booster training sessions in Ball's study were probably also a factor in the higher 5-year follow-up rate in his study compared to ours (67% v. 54%).

4.2 Limitations

The main limitation in this study is the relatively small sample size, which was magnified by the high dropout rate (46%) at the time of the 5-year follow-up. Given

the comprehensive battery of neuropsychological tests conducted (with 61 independent measures) an initial sample size of 151 individuals and a follow-up sample size of 82 individuals is much too small. The requirement to adjust the results for age, educational status and baseline values further weakened the power of the tests to identify differences between groups or over time. Thus many of the negative results in the study could be due to Type II errors – that is, failure to identify important differences between groups because the study sample was too small. Moreover, several of the indices used in the neuropsychological battery employed are not normally distributed, so it would be preferable to use non-parametric tests (e.g., Mann-Whitney tests) to compare the results across groups.

4.3 Implications

The belief that a short cognitive training program can have a prolonged effect on the cognitive functioning of elderly individuals is attractive, but probably not realistic. Changing the long-term trajectory of cognitive functioning, particularly in the elderly cohort who are experiencing a natural decline in their mental functioning, will probably require sustained and repetitive effort to encourage elderly individuals to adopt a 'cognitively healthy lifestyle' just like earlier public health efforts focused on getting middle-age and elderly adults to adopt a 'heart-healthy' lifestyle.

Given the advanced age of respondents in these studies there is inevitably going to be a high dropout rate as the follow-up period is extended. Most of these dropouts are not preventable; many respondents develop serious illnesses that prevent participation, some died and some move away (often to live with their children). Only about one-third (22/68) of those who dropped out in our study withdrew consent, so even if it is possible to make the programs so engaging that all participants are willing to continue participation, there will still be relatively high dropout rates. Sample sizes for such studies need to take this into consideration and the analysis of long-term outcomes must assess the possibility that there is differential rates and types of dropouts in the intervention and control groups and, if so, adjust the results accordingly.

When using comprehensive neuropsychological batteries with dozens of measures to compare groups or to compare a single group over time the likelihood of identifying statistically significant differences is greatly increased due to the number of statistical tests being conducted, particularly if the sample is large. In this scenario the 'significant' tests or measures will change every time the intervention is repeated and researchers may exhaustively debate the reasons for the differences in their studies without realizing that many of these results are statistical artifacts. To prevent this from happening, intervention research for cognitive training must move towards hypothesis-based testing in which the effectiveness of the outcome is based on a small number of specific measures identified *before* starting the intervention. The fishing expeditions for 'significant

variables' that many researchers currently undertake using the huge neuropsychological batteries currently available will not advance knowledge in the field.

Conflict of interest

The authors declare that they have no conflict of interest related to this manuscript.

Funding

This study was supported by the National Science Foundation of China (81200831 and 30770769), the China National Science and Technology Pillar Projects (2009BAI77B03), and the Shanghai Health Bureau (2012038).

社区健康老年人多域认知训练的 5 年随访研究

冯威, 李春波, 陈优, 成燕, 吴文源

背景: 认知训练是一种安全的非药物干预, 可能帮助老年个体减轻认知功能减退和预防痴呆症。

目标: 评估社区健康老年人认知训练的长期效果。

方法: 上海某城市社区 70 岁及以上的健康老年人进行筛选后, 151 名符合纳入标准的个体被分配至干预组 ($n=90$) 和对照组 ($n=61$)。干预组每周接受两次推理、记忆和策略的训练, 持续 12 周 (共 24 次)。于基线、12 周、和入组后 5 年时, 使用《老年人神经心理测验中文版》(NTBE)、《斯特鲁色词测验》、和《一般健康问卷》分别对个体进行评估。

结果: 干预组 49 人 (54%) 和对照组 33 人 (54%) 完成了 5 年的随访。完成与未完成 5 年随访的个体之间在基线时的神经认知测量没有显著差异, 并且干预组脱落个体和对照组脱落个体之间也没有差异。5 年随访时, 干预组个体在 NTBE 和斯特鲁色词测验的 61 项

评估中仅有 5 项评分 (路径 A 测试和划消测验 3) 比对照组好, 但这些差异没有达到预先确定的统计学水平 ($p=0.0008$)。

结论: 我们没有证实以往研究的结果, 它们报导对社区中老年居民进行简要认知培训会有长期效果。我们无法证实认知训练五年后认知功能的差异不可能是由干预组和对照组脱落不同所致, 但可能与样本相对较小和评估项目数目太多有关。今后对于老年人认知训练的干预研究应该建立在假设的基础上 (即专注于感兴趣的单个测量结果), 使用更大的样本, 并将定期强化课程纳入认知训练。

关键词: 老年人; 认知; 神经心理学测试; 随访研究; 神经心理学; 中国

本文全文中文版从 2014 年 4 月 15 日起在 www.saponline.org 可供免费阅读下载

References

- McAuley E, Kramer AF, Colcombe SJ. Cardiovascular fitness and neurocognitive function in older adults: a brief review. *Brain Behav Immun*. 2004; **18**(3): 214-220
- Wang JQ, Wu WY. [Research progress in cognitive training in old people]. *Tongji Da Xue Xue Bao (Yi Xue Ban)*. 2010; **31**(3): 125-128. Chinese. doi: <http://dx.doi.org/10.3969/j.issn.1008-0392.2010.03.032>
- Gates N, Valenzuela M. Cognitive exercise and its role in cognitive function in older adults. *Curr Psychiatry Rep*. 2010; **12**(1): 20-27. doi: <http://dx.doi.org/10.1007/s11920-009-0085-y>
- Shen Y, Li CB, Wu WY. [Cognitive aging, brain plasticity in old people and cognition reservation]. *Zhonghua Xin Wei Yi Xue Yu Nao Ke Xue Za Zhi*. 2009; **18**(8): 765-766. Chinese
- Feng W, Li CB, Wu WY. [Cognitive intervention on aging]. *Shanghai Jing Shen Yi Xue*. 2006; **18**(3): 172-174. Chinese
- Feng W, Li CB, Wu WY, Chen Y, Cheng Y. [Efficacy of comprehensive cognitive training for community healthy elderly: a controlled trial]. *Zhonghua Jing Shen Ke Za Zhi*. 2008; **41**(3): 152-155. Chinese. doi: <http://dx.doi.org/10.3321/j.issn:1006-7884.2008.03.007>
- Feng W, Li CB, Cheng Y, Wu WY. [Long-term effects of composite cognitive training for community healthy elderly: one year follow-up]. *Lin Chuang Jing Shen Bing Xue*. 2009; **19**(3): 145-147. Chinese
- Feng W, Wu WY, Chen Y, Cheng Y, Li CB. [Effects of composite cognitive training for healthy community elders in reasoning ability: a controlled trial]. *Zhonghua Xing Wei Yi Xue Yu Nao Ke Xue Za Zhi*. 2011; **20**(12): 1125-1127. Chinese. doi: <http://dx.doi.org/10.3760/cma.j.issn.1674-6554.2011.12.023>
- Xue HB, Xiao SF, Li CB, He YL, Wu WY, Zhang MY. [The neuropsychological test battery for elderly people]. *Zhonghua Yi Xue Za Zhi*. 2005; **85**(42): 2961-2965. Chinese
- World Health Organization. World Health Organization. *Battery of Cognitive Assessment Instrument for Elderly*. Geneva: WHO, 1996. 5-8
- Xue HB, Xiao SF, Zhang MY. [Reliability and validity of neuropsychological test battery for the elderly]. *J Intern Med Concepts Pract*. 2007; **2**(2): 103-105. Chinese. doi: <http://dx.doi.org/10.3969/j.issn.1673-6087.2007.02.009>
- Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychology*. 1935, **18**(6): 643-662
- van Boxtel MP, ten Tusscher MP, Metsemakers JF, Willems B, Jolles J. Visual determinants of reduced performance on the Stroop color-word test in normal aging individuals. *J Clin Exp Neuropsychol*. 2001; **23**(5): 620-627
- Xue ZQ, Feng W, Li CB, Wu WY. [Short-term effect of cognitive training intervention in community elders]. *Lin Chuang Jing Shen Yi Xue Za Zhi*. 2007; **17**(5): 292-295. Chinese. doi: <http://dx.doi.org/10.3969/j.issn.1005-3220.2007.05.002>

15. Feng W, Li CB, Chen Y, Cheng Y, Wu WY. Integrative cognitive training for healthy elderly Chinese in community: A controlled study. *Biomedical Research*. 2013; **24**(2): 223-229
 16. Cheng Y, Wu W, Feng W, Wang J, Chen Y, Shen Y, et al. The effects of multi-domain versus single-domain cognitive training in non-demented older people: a randomized controlled trial. *BMC Medicine*. 2012; **10**: 30. doi: <http://dx.doi.org/10.1186/1741-7015-10-30>
 17. Ball K, Berch DB, Helmers KF, Jobe JB, Leveck MD, Marsiske M, et al. Effects of cognitive training interventions with older adults: A randomized controlled trial. *JAMA*. 2002; **288**(18): 2271-2281. doi: <http://dx.doi.org/10.1001/jama.288.18.2271>
 18. Willis SL, Tennstedt SL, Marsiske M, Ball K, Elias J, Koepke KM, et al. Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA*. 2006; **296**(23): 2805-2814. doi: <http://dx.doi.org/10.1001/jama.296.23.2805>
 19. Unverzaqt FW, Smith DM, Rebok GW, Marsiske M, Morris JN, Jones R, et al. The Indiana Alzheimer disease center's symposium on mild cognitive impairment. Cognitive training in older adults: lessons from the ACYIVE study. *Curr Alzheimer Res*. 2009; **6**(4): 375-383
- (received: 2013-01-16; accepted: 2013-03-29)



Dr. Wei Feng received her Bachelor's degree in clinical medicine in 2004 and a PhD in biomedical engineering in 2009 from Tongji University School of Medicine. She has worked at the Tongji Hospital of Tongji University since 2009 where she is currently an attending psychiatrist in the Department of Psychiatry. Her main research interest is geriatric psychiatry.

Erratum

In the December 2014 article 'Characteristics of the gastrointestinal microbiome in children with autism spectrum disorder: a systematic review' by Xinyi Cao, Ping Lin, Ping Jiang, and Chunbo Li (*Shanghai Archives of Psychiatry*. 2013; **25**(6): 342-353. doi: <http://dx.doi.org/10.3969/j.issn.1002-0829.2013.06.003>), there were five errors in figure 1: (a) the number of articles identified from English-language databases should have been 5962 instead of 5961; (b) the time range searched in the ISI web of knowledge should have been 1994-2013, not 1986-2013; (c) the time range searched in Ovid/Medline should have been 1970-2013, not 1946-2013; (d) the time range searched in PsycINFO should have been 1966-2013, not 1806-2013; (e) the time range searched in Cochrane Library should have been 1967-2013, not all years. These changes were made in the online version of the journal on January 27, 2014.