Supplementary Material 3

Clinical data include Hamilton anxiety scale (HAMA), Hamilton depression scale (HAMD) [1] and functional activities questionnaire (FAQ)[2]. Neuropsychological evaluations include mini-mental state examination (MMSE) [3], AVLT[4], shape trails test (STT) [5], symbol digit modalities test (SDMT) [6], Boston naming test (BNT) [7], and complex figure test (CFT)[8]. The evaluation details are as follows[9]:

Hamilton Anxiety Scale (HAMA)

This test includes 14 items: 1 Anxious mood; 2 Tension; 3 Fears; 4 Insomnia; 5 Intellectual retardation; 6 Depressed mood; 7 General somatic symptoms (muscular); 8 General somatic symptoms (sensory); 9 Cardiovascular symptoms; 10 Respiratory symptoms; 11 Gastrointestinal symptoms; 12 Genito-urinary symptoms; 13 Autonomic symptoms; 14 Behavior at interview. Each item is divided into 5 levels: 0-asymptomatic; 1 mild; 2 moderate; 3 severe; 4 extremely severe. Total scale score (items 1-14) is 56. No anxiety:0~7; Possible anxiety:7~14; There must be anxiety: 14~21; There must be obvious anxiety: 21~29; Severe anxiety: >29.

Hamilton Depression Scale (HAMD)

This test includes 24 items: 1 Depressed mood; 2 Guilt feelings; 3 Suicidal thoughts; 4 Initial insomnia (difficulty falling asleep);5 Middle insomnia (sleeping lightly); 6 Early morning awakening; 7 Work and interests; 8 Psychomotor retardation (refers to slowed thinking and speech, difficulty concentrating, reduced initiative); 9 Agitation; 10 Psychological anxiety; 11 Somatic anxiety (refers to physiological symptoms of anxiety, including dry mouth, abdominal bloating, diarrhea, hiccups, abdominal cramps, palpitations, headache, hyperventilation and sighing, frequent urination, and sweating); 12 Gastrointestinal symptoms; 13 General somatic symptoms; 14 Sexual symptoms (such as decreased libido and menstrual irregularities); 15 Hypochondriasis; 16 Loss of weight; 17 Insight (self-awareness); 18 Diurnal variation (if symptoms worsen in the morning or evening, specify which, and rate the degree of variation); 19 Depersonalization or derealization (feeling unreal or having delusions of unreality);

Ma J, *et al. J Neurol Neurosurg Psychiatry* 2024; 95:902–911. doi: 10.1136/jnnp-2023-333152

20 Paranoid symptoms; 21 Obsessive symptoms (obsessive thoughts and compulsive behaviors); 22 Feeling of incapacity; 23 Hopelessness; 24 Self-esteem feelings. Each item is divided into 5 levels: 0- asymptomatic; 1 mild; 2 moderate; 3 severe; 4 extremely severe. Normal:0~8; Possible depression: 8~20; Diagnosed depression: 21~35; Severe depression:>35.

Functional Activities Questionnaire (FAQ)

FAQ reflecting daily living abilities. This FAQ includes a total of 10 questions, with each question encompassing two dimensions: accuracy and completion speed. Each dimension consists of five levels, represented as 0, 1, 2, 3, and 9. A score of 9 indicates that the task is not applicable or cannot be completed.

Mini-Mental State Examination (MMSE)

This test is similar to the English language version of the MMSE, and is scored out of 30.

Auditory Verbal Learning Test (AVLT)

The examiner read out 12 two-character words. Each group of 12 words contained three different categories with four words in each. The different word types were presented randomly with 1-s intervals between the words. Immediately after the entire list had been presented, the participant was asked to recall the words. This learning and recall phase were repeated three times. The participant was given a 5-min non-verbal test, and was then asked to recall the 12 words for the fourth time. Next, the participant completed another 20-min non-verbal task, and was then asked to recall the word list for the fifth time (AVLT-long delayed recall, AVLT-LR or AVLT-N5). Finally, the participant was shown the word list.

Shape Trails Test (STT)

The shape trails test parts A and B: in the STT-A, the participant is asked to connect a

Ma J, *et al. J Neurol Neurosurg Psychiatry* 2024; 95:902–911. doi: 10.1136/jnnp-2023-333152

series of Arabic numbers (1-25) in their numerical order. In the Chinese version of the STT-B, Arabic digits (1-25) are surrounded by either a square or a circle, and the participant is asked to connect the digits in a sequence such that the shapes surrounding the digits alternate. We recorded the time taken to complete the task.

Symbol Digit Modalities Test (SDMT)

In this test, the participant was asked to match a series of symbols to corresponding digits according to a symbol-digit pairing illustration. We recorded the number of items
correctly completed within 90 s. After 90 s, the participant was asked to recall the symbols that corresponded to each digit [known as SDMT-incidental learning (SDMT-IL); i.e., participants were not told in advance that they would be asked to recall the information].

Boston Naming Test (BNT) (30-item version)

In this test, the participants were asked to name the subjects of 30 pictures with no time limit; the total possible score was 30.

Complex Figure Test (CFT)

Participants were asked to copy a visual figure (CFT-copy), then to draw the figure from memory after approximately 25 min had passed (CFT-delayed recall, CFTLR). We used the scoring standard established by Taylor in 1981, with a total possible score of 36.

Quality Control of Data Collection

Ensuring the accuracy and reliability of our data is of utmost importance to us, and we took several steps to maintain high standards of quality control:

1) Dual independent data entry: To mitigate the risk of data entry errors, we adopted a dual-entry system. This involved two distinct operators entering the same data

Ma J, *et al. J Neurol Neurosurg Psychiatry* 2024; 95:902–911. doi: 10.1136/jnnp-2023-333152

independently, with any discrepancies being identified and rectified through a rigorous verification process.

2) Specialized training for data entry staff: Our team underwent a comprehensive training regimen that covered both the use of our data entry software and a deep dive into the cognitive assessments and scales utilized in our study, ensuring their understanding of the significance, scoring ranges, and coding rules for each scale.

3) Systematic quality assurance procedures: We periodically sampled entered data for quality checks, comparing them against original records to ensure accuracy. This process allowed us to maintain continuous oversight and implement immediate corrections where necessary.

4) Continuous process evaluation: The data entry process was subject to ongoing evaluation through regular review meetings. These meetings allowed our team to address any challenges promptly and refine our data entry methodologies as needed.

5) Comprehensive audit trails: The audit trail functionality of our data entry system documented every action, from initial data entry to subsequent modifications. This level of transparency and traceability was instrumental in ensuring the integrity of our data entry process.

Table S3-1. Normal range of score for each cognitive function assessment scale.

Indicators	Age Group		
	50-59	60-69	70-79
HAMD		8	
HAMA		7	
MMSE		24	
AVLT	24	22	19
AVLT_N5	5	4	3
BNT		22	
STT_B, Middle School	180	200	260
STT_B, University	180	190	220
SDMT_correct		34	
CFT_copy	33	31	30

AVLT: Auditory verbal learning test; AVLT_N5: Auditory verbal learning test, long-term delay recall; BNT: Boston naming test; HAMA: Hamilton Anxiety Scale; HAMD: Hamilton depression scale; MMSE: Mini-Mental State Examination; STT_B: Shape trails test, part B; SDMT_correct:

Symbol digit modalities test, correct number; CFT_copy: Complex figure test, copy part.

In our study, individuals participate in daily low-intensity walking exercises. Sleep duration was determined according to the following two questions[10]: (1) How many minutes of actual sleep on average did you get at night (this may be shorter than the number of hours you spent in bed) during the past 6 months? (2) How many minutes on average did you take a nap after lunch during the past 6 months?

References

1. Maier W, Buller R, Philipp M, Heuser I. The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. *Journal of affective disorders* 1988; **14**(1): 61-8.

2. Marshall GA, Zoller AS, Lorius N, et al. Functional Activities Questionnaire Items that Best Discriminate and Predict Progression from Clinically Normal to Mild Cognitive Impairment. *Curr Alzheimer Res* 2015; **12**(5): 493-502.

3. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; **12**(3): 189-98.

4. Zhao Q, Guo Q, Liang X, et al. Auditory Verbal Learning Test is Superior to Rey-Osterrieth Complex Figure Memory for Predicting Mild Cognitive Impairment to Alzheimer's Disease. *Current Alzheimer research* 2015; **12**(6): 520-6.

5. Zhao Q, Guo Q, Li F, Zhou Y, Wang B, Hong Z. The Shape Trail Test: application of a new variant of the Trail making test. *PloS one* 2013; **8**(2): e57333.

6. Wang B, Guo Q, Zhao Q, Hong Z. Memory deficits for non-amnestic mild cognitive impairment. *Journal of neuropsychology* 2012; **6**(2): 232-41.

7. Mack WJ, Freed DM, Williams BW, Henderson VW. Boston Naming Test: shortened versions for use in Alzheimer's disease. *Journal of gerontology* 1992; **47**(3): P154-8.

8. Youn YC, Pyun JM, Ryu N, et al. Use of the Clock Drawing Test and the Rey-Osterrieth Complex Figure Test-copy with convolutional neural networks to predict cognitive impairment. *Alzheimers Res Ther* 2021; **13**(1): 85.

9. Ma J, Wang Y, Zhang Y, Zhen X, Guo Q, Shi M. Neuropsychological features in post-stroke cognitive impairment with no dementia patients with different Traditional Chinese Medicine syndromes. *J Tradit Chin Med* 2019; **39**(1): 97-102.

10. Zhong W, Wang F, Chi L, Yang X, Yang Y, Wang Z. Association between Sleep Duration and Depression among the Elderly Population in China. *Exp Aging Res* 2022; **48**(4): 387-99.