The Reliability and Validity of the Chinese version of Confusion Assessment Method based Scoring System for Delirium Severity (CAM-S)

Xinchun Mei[a,#], **Yupeng Chen**[a,#], **Hailin Zheng**[a], **Zhongyong Shi**[a], **Edward R. Marcantonio, MD, SM**[b], **Zhongcong Xie, MD, PhD**[c], **Yuan Shen, MD, PhD**[a,*]

[a]Department of Psychiatry, Shanghai 10th People's Hospital, Anesthesia and Brain Research Institute, Tongji University, Shanghai, P.R. China

[b]Divisions of General Medicine and Primary Care and Gerontology, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA

[c]Geriatric Anesthesia Research Unit, Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA, USA

Abstract

Previous studies showed that the Confusion Assessment Method based delirium severity evaluation tool (CAM-S) had good reliability and validity. However, there is no Chinese version of CAM-S. Therefore, we set out to perform a prospective investigation in older Chinese patients who had total joint replacement surgery under general anesthesia in Tenth People's Hospital in Shanghai, P.R. China. A total of 576 participants of 60 year-old and older were screened, 179 participants were enrolled and 125 of them were included for the final analysis. Pre-operative evaluations were conducted one day before the surgery. Postoperative evaluations were conducted twice daily from postoperative day 1 to day 3. The incidence of postoperative delirium was 24.8%. The Chinese version of CAM-S [including a Short Form (CAM-S Short Form) and a Long Form (CAM-S Long Form)] had an optimal reliability reflected by internal consistency (*Cronbach's α* = 0.748 and 0.839 for CAM-S Short Form and CAM-S Long Form respectively), split-halves reliability (*Pearson correlation coefficient* = 0.372 and 0.384 for CAM-S Short Form and CAM-S Long Form respectively), and inter-rater reliability (*intra-class correlation coefficients* = 0.629 and 0.945 for CAM-S Short Form and CAM-S Long Form respectively). Besides, the Chinese version of CAM-S also showed a good discriminate validity. The domain scores of CAM-S were inversely correlated with corresponding domain scores of MMSE. Finally, a receiver operating characteristic (ROC) analysis obtained an optimal cutoff point of 2.5 for CAM-S Short Form and 3.5 for CAM-S Long Form in recognizing delirium diagnosed by CAM. The areas under the ROC were 0.989 (95% CI 0.972 – 1.000, p < 0.001) and 0.964 (95% CI 0.946 – 0.982, p < 0.001), respectively. These data suggest that the Chinese version of CAM-S has good reliability and validity in

*Correspondence to: Yuan Shen, M.D., Ph.D., Department of Psychiatry, Shanghai 10th People's Hospital, Anesthesia and Brain Research Institute, Tongji University, 301 Yanchang Rd., Outpatient Clinic Building, Floor 4th, Room 4405, 200072 Shanghai, P.R. China. Tel.:/Fax: +86 21 66303649; kmshy@tongji.edu.cn.
#:These authors have contributed equally to this work.

CONFLICT OF INTEREST/DISCLOSURE STATEMENT
The authors have no conflict of interest to report.

evaluating postoperative delirium in geriatric Chinese patients and may be a useful tool to assess the severity of delirium.

Keywords

confusion assessment method; postoperative delirium severity; reliability; validity

INTRODUCTION

Delirium is a dysfunction of attention, thinking and consciousness, accompanied by changes in cognition or perceptual disturbances, with an acute onset and a fluctuating nature, which is not attributable to pre-existing psychiatric disorders or substance-induced states [1]. Postoperative delirium is one of the most common postoperative complications in hospitalized older adults [2]. It is estimated that postoperative delirium occurs in 12% to 56% of elderly patients following surgery under anesthesia [3–5]. Importantly, postoperative delirium is associated with increased morbidity, mortality, length of stay and placement in nursing homes [6–8].

Measurement of delirium severity is essential to assess the impact of postoperative delirium on postoperative outcomes. The quantification of delirium will facilitate the clinical investigation across different settings and studies. Based on the Confusion Assessment Method (CAM), the most often utilized diagnostic tool for delirium, a new scoring system to determine delirium severity called CAM-S has been developed [9]. The CAM-S has been demonstrated a strong association with clinical outcomes and significant gradients across severity categories [9]. The CAM-S includes two forms: a Short Form (CAM-S Short Form, based on the 4-item algorithm) and a Long Form (CAM-S Long Form, based on the 10-item CAM instrument). The reliability and validity of CAM-S in the measurement of delirium has been established and has provided a new measure for evaluating delirium severity [9]. However, there is no Chinese version of CAM-S and the feasibility of CAM-S in the Chinese population remains uninvestigated.

Therefore, we set out to perform a prospective investigation in older Chinese patients who had total joint replacement surgery under general anesthesia in Tenth People's Hospital in Shanghai, P.R. China. The objective of this study was to evaluate the validity and reliability of the CAM-S (including CAM-S Short Form and CAM-S Long Form) in Chinese. We focused on the CAM-S in the current studies to test a hypothesis that Chinese version of CAM-S has good reliability and validity to assess the severity of delirium.

MATERIALS AND METHODS

PARTICIPANTS IN THE STUDY

The protocol was approved by the Human Research Ethics Committee of Shanghai 10th People's Hospital, P. R. China [SHSY-IEC-3.0/15–78/01]. A tsiotal of 576 participants who had scheduled orthopedic surgery at the Shanghai 10th People's Hospital were screened from June 22, 2016 to April 11, 2017. Among them, 179 participants were enrolled in the study and signed the written informed consent. The inclusion criteria were: (1) 60 year-old

or older; (2) had general anesthesia; (3) American society of anesthesiologist (ASA) class I to III; and (4) normal cognition at enrollment [Mini-mental State Examination (MMSE) scores above than 24]; Participants were excluded from the studies if they had: (1) pre-existing delirium determined by CAM; (2) neurological diseases (e.g., stroke, Parkinson's disease, etc.) (3) mental disorders (e.g., acute episode of major depressive disorder, schizophrenia, mental retardation, etc.) diagnosed by using the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [1] and (4) impaired vision or auditory function which may affect the assessments. All participants signed written informed consent before being enrolled in the study.

Sample size was estimated according to our previous studies of postoperative delirium, in which the incidence of postoperative delirium was 25.6% [10]. Considering a dropout rate of 20% at follow-up assessments [11], 125 participants was needed with $p < 0.05$ as the significance level.

PRE-OPERATIVE INTERVIEW

Participants were admitted two to three days before the scheduled surgery and the pre-operative assessments were performed one day before surgery. The assessments included demographic characteristics (e.g., age, gender, height, weight, and education), clinical features (e.g., diagnosis, past medical history and type of surgery) and co-morbid diseases determined by Charlson Comorbidity Index (CCI) scores [12]. General cognitive function was assessed using Mini-mental State Examination (MMSE) [13]. Attention was measured with Digit Span Test (DST) forward and backward [14, 15]. The CAM algorithm and CAM-S were performed one day before the surgery to exclude pre-existing delirium.

ANESTHESIA AND SURGERY

All of the participants had Total Hip Replacement (THR) or Total Knee Replacement (TKR) under general anesthesia. Among all the participants, 58 participants received intravenous anesthetic, propofol, while 67 participants received inhalational anesthetic, sevoflurane. The participants had standardized perioperative care, including pre-operative medication, general anesthetics, and postoperative pain control, which included standardized postoperative pain management. All of the enrolled participants used patient-controlled analgesia. There were no major complications among the participants during the immediate postoperative period.

POSTOPERATIVE INTERVIEW

The CAM and CAM-S were performed twice daily at two time-intervals: 9:00–11:00 am and 4:00–6:00 pm on the first, second and third day post-surgery. The assessments on the 1st and 2nd postoperative day included DST and certain cognitive domains of MMSE (see Evaluation of validity). On the 3rd postoperative day, a complete MMSE was re-performed to assess postoperative cognitive function apart from CAM and CAM-S.

The incidence of postoperative delirium was assessed using CAM. The CAM algorithm consists of four clinical criteria: (1) acute onset and fluctuating course, (2) inattention, (3) disorganized thinking, and (4) altered level of consciousness. To diagnose delirium, both the first and the second criteria have to be present, as well as either the third and/or the fourth

criteria. The CAM in the Chinese language has been proven to have good reliability and validity among the Chinese elderly population [14]. Two psychiatrists who performed the delirium assessments in the current study had good training and went through quality control procedures.

The medical records were reviewed daily to check for delirium occurrence beyond the two assessment times.

CONFUSION ASSESSMENT METHOD-SEVERITY (CAM-S)

CAM-S includes a short-form (CAM-S Short Form) and long-form (CAM-S Long Form) scoring system. The Short Form includes the same four features as CAM: acute change or fluctuation, inattention, disorganized thinking and altered level of consciousness. Each feature, except fluctuation, is scored as absent (0), mild (1) or marked (2). Acute onset of fluctuation was scored as absent (0) or present (1). The sum score of CAM-S Short Form ranges from 0 (no) to 7 (most severe). The Long Form (CAM-S Long Form) contains 10 items: acute onset or symptom fluctuation, inattention, disorganized thinking, altered level of consciousness, disorientation, memory impairment, perceptual disturbances, psychomotor agitation, psychomotor retardation and sleep-wake cycle disturbance. Similarly, each item is rated from 0 to 2, except acute onset or fluctuation. The sum score of CAM-S Long Form ranges from 0 (no) to 19 (most severe) [9].

The translation and back-translation methods were used to create the Chinese version of CAM-S. The CAM-S was first translated into Chinese by X.M. and then back-translated into English by Z.S. The original English version and the back-translated English version of CAM-S were compared, and variations were identified for inspection of consistency. All items with variances were then translated into Chinese and back-translated into English again according to suggestions from consistency discussions. The final Chinese version of CAM-S was generated only after the back-translated English version was consistent with the original English version of CAM-S.

STATISTICAL ANALYSIS

The *One-Sample Kolmogorov-Smirnov* method was used to test the normality of all variables. They were generally divided into continuous parameters and categorical parameters. Continuous parameters, including age, Body Mass Index (BMI), education, CCI Scores and baseline MMSE scores were presented as mean ± standard deviation (SD). Categorical factors, such as gender, were presented as percentage (%). For all analysis, we used the CAM-S scores (including CAM-S Short Form scores and CAM-S Long Form scores) and the cognitive domain scores at each time-point preoperatively and postoperatively. Since each participant was evaluated once preoperatively and twice daily post operation, 875 person-times assessments (including 125 person-times preoperative assessments and 750 person-times postoperative assessments) were yielded.

EVALUATION OF RELIABILITY

Inter-item reliability was determined by using *Cronbach's alpha coefficient,* to assess internal consistency between items [17, 18]. Split-halves reliability was determined by

Pearson correlation coefficient of two parts of CAM-S, namely odd items and even items [19]. Then inter-rater reliability was evaluated using 81 postoperative assessments of 42 participants that were randomly selected for mutual assessment by two independent evaluators (X. M. and Y. C.) who were blinded to each other. *Cohen's kappa coefficient* was calculated to determine the consistence of diagnosing delirium by using CAM between two raters [20]. In comparison, *intra-class correlation coefficients* (*ICC*) was applied to determine consistence of assessing delirium severity between two raters by using CAM-S [21].

EVALUATION OF VALIDITY

Discriminate validity of CAM-S was evaluated by *Student t-test*, comparing the CAM-S scores between participants who developed postoperative delirium and those who did not develop delirium diagnosed by CAM [22]. Given that CAM-S Long Form evaluated multiple psychiatric domains including inattention, disorientation and memory impairment, DST and corresponding domains of MMSE were selected as the criterion of criteria validity of CAM-S Long Form [23]. For the analysis of criteria validity, *spearman correlation coefficients* between domain scores of CAM-S and MMSE as well DST were calculated. They were: (1) inattention: item 2 of CAM-S and sum of item 12 of MMSE (serial 7s) as well DST scores [24]; (2) disorientation: item 5 of CAM-S and sum of item 1 to item 10 of MMSE [24] and (3) memory: item 6 of CAM-S and sum of item 11 and 13 of MMSE [24].

THE OPTIMAL CHINESE VERSION OF CAM-S CUTOFF POINT IN RECOGNIZING POSTOPERATIVE DELIRIUM

ROC analysis was performed to determine the optimal CAM-S cutoff point for recognizing delirium defined by CAM algorithm. The optimal cutoff was obtained according to *Youden* index (maximum of [sensitivity + specificity - 1]). The predictive power was evaluated by the area under the curve (AUC) in the receiver operating characteristic (ROC) analysis.

All analyses were conducted with SPSS version 20.0 (SPSS Inc., Chicago, IL) with $p < 0.05$ as the significance level.

RESULTS

CHARACTERISTICS OF PARTICIPANTS

A total of 576 participants over 60 year-old who were admitted for orthopedic surgery were screened and 397 of them were excluded. Thus, 179 participants were enrolled and provided written informed consent for the study. Among them, 54 participants were further excluded due to: baseline MMSE less than 24 (N=22), change of anesthesia methods (N=14), cancelation of operation (N=5), withdraw of inform consent (N=12) and early discharge before scheduled date (N=1). In the end, 125 participants were included for the final data analysis (Figure 1). The characteristics of the participants were presented in Table 1. Among all the participants, 31 participants developed postoperative delirium. The incidence of postoperative delirium was 24.8%. The prevalence of delirium in this participant population on postoperative day 1, day 2 and day 3 was 19.2%, 10.4% and 4.0%, respectively.

RELIABILITY

The overall *Cronbach's alpha* coefficients were 0.748 ($p < 0.001$) for CAM-S Short Form and 0.839 for CAM-S Long Form ($p < 0.001$). *Pearson correlation* coefficients between total scores of odd items and even items were 0.372 ($p < 0.001$) and 0.384 ($p < 0.001$) for CAM-S Short Form and CAM-S Long Form respectively. The *Cohen Kappa* for inter-rater reliability of delirium diagnosis using CAM was 1.00, ($p < 0.001$). The inter-rater *ICC* of delirium severity assessment was 0.629 (95% CI: 0.424–0.762, $p < 0.001$) using the CAM-S Short Form and 0.945 (95% CI: 0.915–0.965, $p < 0.001$) using the CAM-S Long Form.

VALIDITY

The domain scores of CAM-S were inversely correlated with corresponding domain scores of MMSE including attention ($r = -0.519$, $p < 0.001$), orientation ($r = -0.815$, $p < 0.001$), and memory ($r = -0.751$, $p < 0.001$). The participants who developed postoperative delirium had a higher CAM-S score than the participants who did not develop postoperative delirium both in CAM-S Short Form (3.22 ± 0.55 *vs.* 1.37 ± 0.64, $p < 0.001$) and CAM-S Long Form (5.69 ± 1.58 *vs.* 3.01 ± 1.09, $p < 0.001$).

THE OPTIMAL CHINESE VERSION OF CAM-S CUTOFF IN DESCRIBING POSTOPERATIVE DELIRIUM

ROC analysis was performed to determine the optimal CAM-S cutoff point in recognizing delirium defined by CAM algorithm. The area under ROC curve was 0.989 (95% CI 0.972 – 1.000, $p < 0.001$) for CAM-S Short Form and 0.964 (95% CI 0.946 – 0.982, $p < 0.001$) for CAM-S Long Form. The optimal cutoff point for CAM-S Short Form was 2.5, at which a sensitivity of 0.980 and a specificity of 0.994 were achieved. For CAM-S Long Form, an optimal cutoff point of 3.5 could achieve a sensitivity of 0.922 and a specificity of 0.890 (Fig.2). Using cutoff 2.5 of CAM-S Short Form, all these 31 participants who diagnosed of postoperative delirium by CAM were identified as having delirium. Using cutoff of 3.5 of CAM-S Long Form, 29 out of 32 participants diagnosed of postoperative delirium by CAM were identified as having delirium. These data indicated that the participants who had a CAM-S Short Form score over 3 and CAM-S Long Form score over 4 would more likely to be diagnosed as postoperative delirium.

DISCUSSION

This study provided evidence for the reliability and validity of the Chinese version of CAM-S in measuring severity of postoperative delirium in older adults who had surgery. The study demonstrated that the Chinese version of CAM-S had good validity and high reliability. Thus, this Chinese version of CAM-S is a simple but reliable tool for the assessment of the severity of postoperative delirium happened under different types of general anesthesia in Chinese population. CAM-S is also feasible for delirium severity assessment in non-surgical patients, as used for preoperative (non-surgical) assessments in present study.

The CAM-S is designed to quantify the severity of delirium symptoms observed at bedside. The original CAM-S has demonstrated good psychometric properties, high inter-rater reliability and strong association with clinical outcomes related to delirium [9]. The results

from the current studies revealed the good internal consistency, good spilt-halves reliability and a high inter-rater reliability [16–22]. It suggested that the Chinese version of CAM-S is a reliable measurement to assess the severity of postoperative delirium. We plan to make the CAM-S Chinese Version as a research tool publicly available.

Chinese version of CAM-S could identify participants who develop postoperative delirium effectively. In our study, the participants who developed postoperative delirium had a higher CAM-S score than the participants who did not develop postoperative delirium both in CAM-S Short Form and CAM-S Long Form. Regarding the criteria validity, an inverse relationship was observed between CAM-S domain scores and corresponding domain scores in MMSE score. The Mini-Mental State Examination (MMSE) is the most widely used cognitive test for screening cognitive disorders. Several studies have been conducted in China during the past few decades to assess the efficiency of MMSE in Chinese for identifying participants with cognitive dysfunction [25]. In particular, the Chinese version of CAM-S showed strong correlation with disorientation and memory impairment defined by MMSE. These results further indicate that the Chinese version of CAM-S could be a reliable tool to reveal the cognitive deficits of delirious patients (summarized in Table 2).

In the current study, an optimal CAM-S Short Form cutoff point of 2.5 and CAM-S Long Form cutoff point of 3.5 was obtained by ROC analysis based on the combined prevalence of delirium through postoperative day 1 to day 3. In this way, participants who obtained 3 or more in CAM-S Short Form and 4 or more in CAM-S Long Form were more likely to develop postoperative delirium. Employing this cutoff score, we were able to achieve an optimal sensitivity and specificity using the Chinese version of CAM-S to identify delirium. All participants who developed postoperative delirium obtained 3 or more by CAM-S Short Form while 29 out of 32 participants obtained 4 or more by CAM-S Long Form. Three participants obtained 3 by CAM-S Short Form and CAM-S Long Form.

There were several limitations in this study. First, the participants in the current studies only included the patients who had the surgery of THR or TKR. The future studies may need to include patients who have different types of surgery (e.g., cardiac surgery). Second, only cognitive domains in MMSE were used to validate the criteria validity of CAM-S. Future studies may use other tools [e.g. Memorial Delirium Assessment Scale (MDAS)] to investigate the validity of the Chinese version of CAM-S. Third, the sample size was relatively small in our study, which led to the relative low power for cutoff values and particular type of aged surgical patient group in this study.

In conclusion, the results from this study showed that the Chinese version of CAM-S could be an effective and reliable method to measure the severity of postoperative delirium in Chinese patients. Examining severity of postoperative delirium is important to better define the clinical impact of the full spectrum of delirium in older patients. Thus, widely adopting the Chinese version of CAM-S will likely promote more clinical studies of postoperative delirium and its therapeutic/preventive strategies.

ACKNOWLEDGEMENTS

REFERENCES

[1]. 4th ed Washington, DC, American Psychiatric Association (1997) Diagnostic and Statistical Manual of Mental Disorders (DSM-IV).

[2]. Marcantonio ER (2012) Postoperative delirium: a 76-year-old woman with delirium following surgery. JAMA 308, 73–81. [PubMed: 22669559]

[3]. Breitbart W, Rosenfeld B, Roth A, Smith MJ, Cohen K, Passik S (1997) The Memorial Delirium Assessment Scale. J Pain Symptom Manage 13, 128–137. [PubMed: 9114631]

[4]. DeCrane SK, Sands L, Ashland M, Lim E, Tsai TL, Paul S, Leung JM (2011) Factors associated with recovery from early postoperative delirium. J Perianesth Nurs 26, 231–241. [PubMed: 21803271]

[5]. Rudolph JL, Marcantonio ER (2011) Review articles: postoperative delirium: acute change with long-term implications. Anesth Analg 112, 1202–1211. [PubMed: 21474660]

[6]. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40, 373–383. [PubMed: 3558716]

[7]. Lawton MP, Brody EM (1969) Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 9, 179–186. [PubMed: 5349366]

[8]. McCusker J, Cole M, Abrahamowicz M, Primeau F, Belzile E (2002) Delirium predicts 12-month mortality. Arch Intern Med 162, 457–463. [PubMed: 11863480]

[9]. Inouye SK, Kosar CM, Tommet D, Schmitt EM, Puelle MR, Saczynski JS, Marcantonio ER, Jones RN (2014) The CAM-S: development and validation of a new scoring system for delirium severity in 2 cohorts. Ann Intern Med 160, 526–533. [PubMed: 24733193]

[10]. Shi Z, Wu Y, Li C, Fu S, Li G, Zhu Y, Celeste AS, Marcantonio ER, Xie Z, Shen Y (2014) Using the Chinese version of Memorial Delirium Assessment Scale to describe postoperative delirium after hip surgery. Front Aging Neurosci 6, 297. [PubMed: 25414664]

[11]. Bernd R, Jean-Baptist du P, Daniel W, Robert K, Maria B (2014) Sample Size Calculation in Clinical Trials. Allergol Immunopathol (Madr) 42, 485–592. [PubMed: 24280317]

[12]. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40, 373–383. [PubMed: 3558716]

[13]. Robert K, Zhang M, Ya-Qu O, Wang Z, William TL, Elena Y, Wong SC, David PS, Igor G (1988) A Chinese version of the mini-mental state examination; Impact of illiteracy in a Shanghai dementia survey. Journal of Clinical Epidemiology 41, 971–978. [PubMed: 3193141]

[14]. Huang X, Du X, Song H, Zhang Q, Jia J, Xiao T, Wu J (2015) Cognitive impairments associated with corpus callosum infarction: a ten cases study. Int J Clin Exp Med 8, 21991–21998. [PubMed: 26885171]

[15]. Martins S, Lourenco C, Pinto-de-Sousa J, Conceicao F, Paiva JA, Simoes MR, Fernandes L (2015) Validation study of the European Portuguese version of the Confusion Assessment Method (CAM). Int Psychogeriatr 27, 777–784. [PubMed: 25275655]

[16]. Leung J, Leung V, Leung CM, Pan PC (2008) Clinical utility and validation of two instruments (the Confusion Assessment Method Algorithm and the Chinese version of Nursing Delirium Screening Scale) to detect delirium in geriatric inpatients. Gen Hosp Psychiatry 30, 171–176. [PubMed: 18291299]

[17]. Henrica CW, Mokkink LB, Mosmuller DG, Terwee CB (2017) Spearmane Brown prophecy formula and Cronbach's alpha: different faces of reliability and opportunities for new applications. Journal of Clinical Epidemiology 85, 45–49. [PubMed: 28342902]

[18]. Gilbert B (2000) Creating comparability among reliability coefficients: the case of Cronbach alpha and cohen kappa. Psychological Reports 87, 1171–1182. [PubMed: 11272758]

[19]. Groppe DM, Makeig S, Kutas M (2009) Identifying reliable independent components via split-half comparisons. Neuroimage 45, 1199–1211. [PubMed: 19162199]

[20]. Tang W, Hu J, Zhang H, Wu P, He H (2015) Kappa coefficient: a popular measure of rater agreement. Shanghai Arch Psychiatry 27, 62–67. [PubMed: 25852260]

[21]. Dichter MN, Schwab EM (2018) Item distribution and inter-rater reliability of the German version of the quality of life in Alzheimer's disease scale (QoL-AD) proxy for people with dementia living in nursing homes, BMC Geriatrics 18, 145. [PubMed: 29914389]

[22]. Hewston P, Deshpande N (2017) The Short Version of the Activities-Specific Balance Confidence Scale for Older Adults with Diabetes-Convergent, Discriminant and Concurrent Validity: A Pilot Study. Can J Diabetes 41, 266–272. [PubMed: 28268190]

[23]. Moreno R, Martínez RJ, Muñiz J (2015) Guidelines based on validity criteria for the development of multiple choice items. Psicothema 27, 388–394. [PubMed: 26493578]

[24]. Gabriel FJ, Santesteban O, Trzepacz P, Bernal C, Valencia C, Ocampo MV, Pablo JD, Gaviria AM, Vilella E (2014) MMSE items that predict incident delirium and hypoactive subtype in older medical inpatients. Psychiatry Res 220, 975–981. [PubMed: 25307690]

[25]. Yang Z, Holt HK, Fan JH, Ma L, Liu Y, Chen W, Como P, Zhang L, Qiao YL (2016) Optimal Cutoff Scores for Alzheimer's Disease Using the Chinese Version of Mini-Mental State Examination Among Chinese Population Living in Rural Areas. Am J Alzheimers Dis Other Demen 31, 650–657. [PubMed: 27659393]
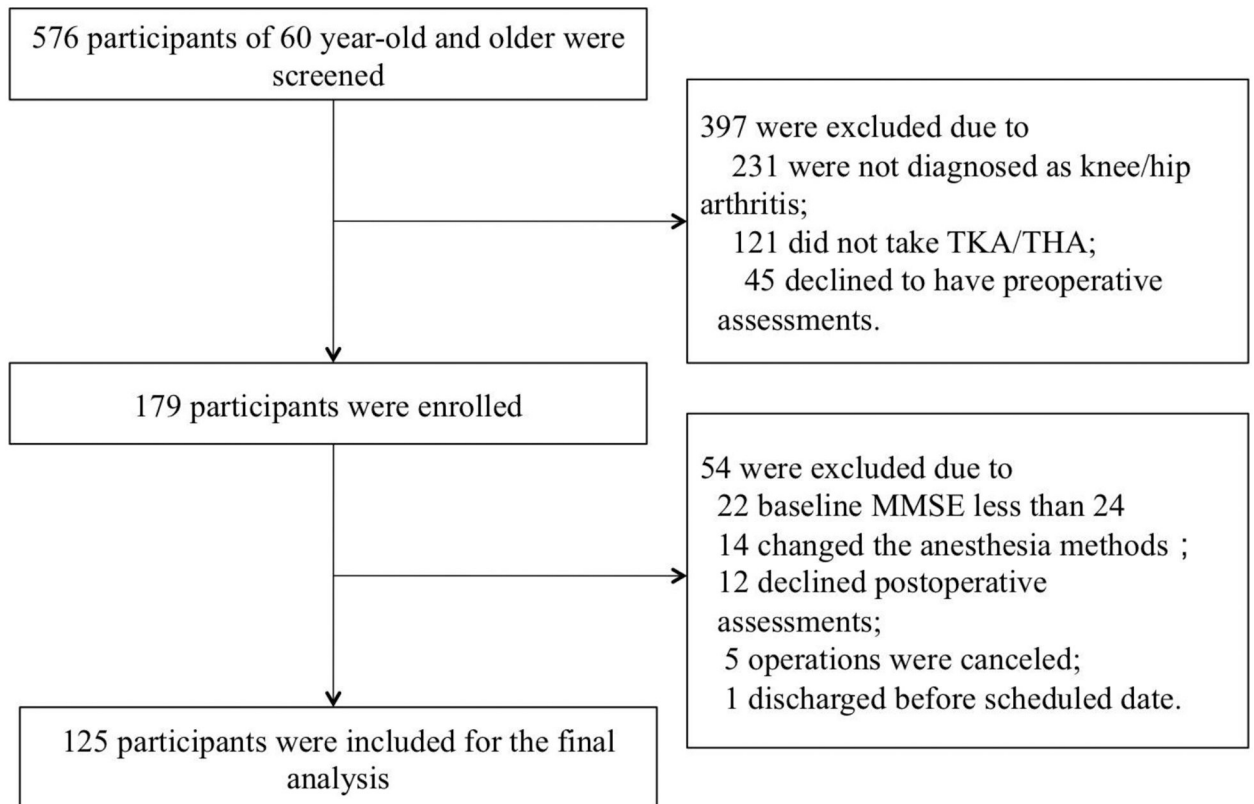
Figure 1: Flow diagram.

The flow diagram shows that 576 participants of 60 year-old and older were screened for the study and 179 participants were enrolled in the study. There were 54 participants who were further excluded, resulting in 125 participants for the final data analysis.
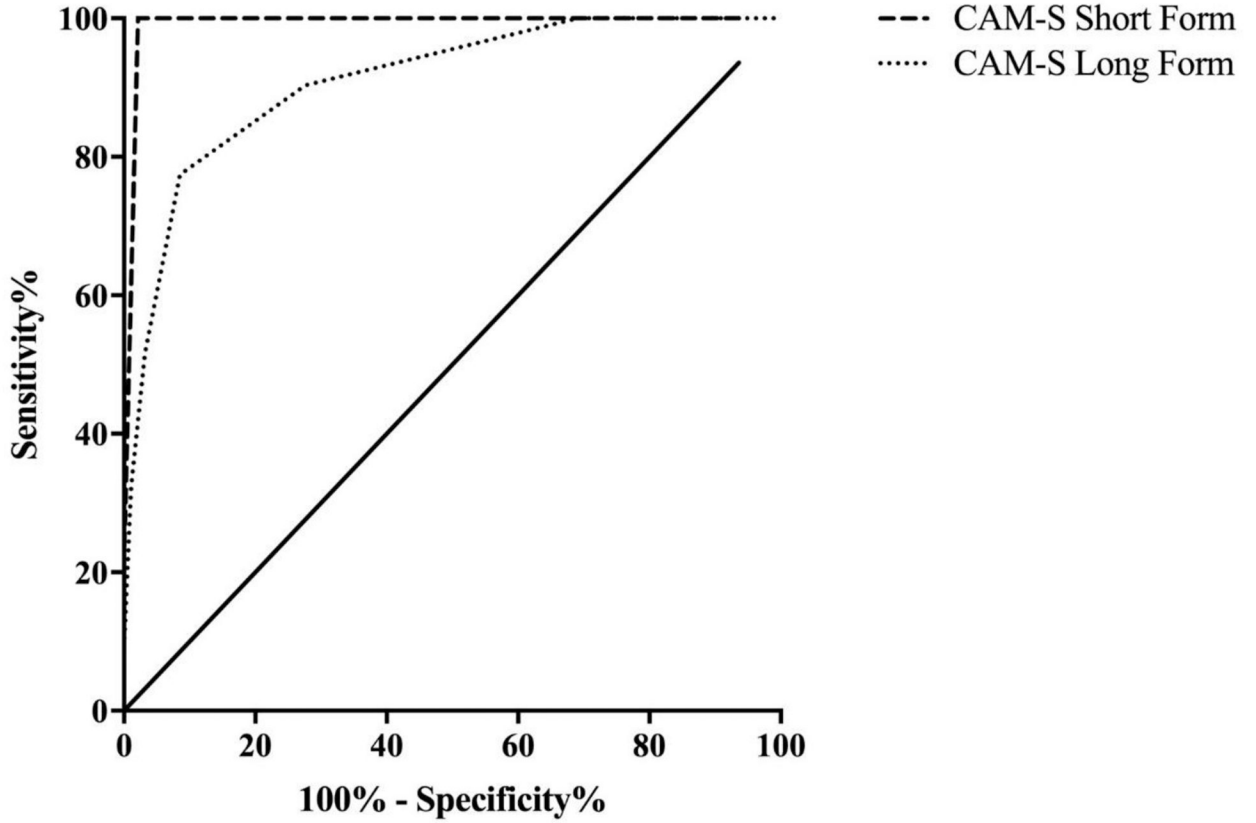
Figure 2: Receiver operating characteristics (ROC) curve.
The ROC analysis was used for determination of the diagnostic sensitivity and specificity of the optimum value of the CAM-S Short Form and CAM-S Long Form *vs.* the CAM. The area under ROC curve was 0.989 (95% CI 0.972 – 1.000, $p < 0.001$) for CAM-S Short Form and 0.964 (95% CI 0.946 – 0.982, $p < 0.001$) for CAM-S Long Form. The optimal cutoff point for CAM-S Short Form was 2.5, at which a sensitivity of 0.980 and a specificity of 0.994 were achieved. For CAM-S Long Form, an optimal cutoff point of 3.5 could achieve a sensitivity of 0.922 and a specificity of 0.890.

Table 1.

Demographic characteristics of the study participants

Variables (N=125)	Value
Male sex, n (%)	45 (36.0)
Age, mean ± SD, yr	70.3 ± 6.5
BMI, mean ± SD, kg/m2	26.0 ± 3.4
Education, mean ± SD, yr	9.1 ± 3.8
CCI, mean ± SD, score	3.2 ± 1.1
Baseline MMSE, mean ± SD, score	26.8 ± 1.6

Abbreviation: BMI, body mass index; CCI, Charlson Comorbidity Index; MMSE, Mini-mental State Examination; yr, year; kg, kilogram; SD, standard deviation.

Note: Continuous parameters were presented as Mean ± standard deviation (SD). Categorical factors were presented as percentage. BMI was calculated dividing weight by the square of height.

Table 2.

Summary of reliability and validity of Chinese version of CAM-S

	Reliability		
	Internal consistency[a]	Split-halves reliability[b]	Inter-rater reliability[c]
CAM-S Short Form	0.748	0.372	0.629
CAM-S Long Form	0.839	0.384	0.945
	Discriminate validity[d]		
	POD	Non POD	p value
CAM-S Short Form	3.22 ± 0.55	1.37 ± 0.64	< 0.001
CAM-S Long Form	5.69 ± 1.58	3.01 ± 1.09	< 0.001
	Criterion-related validity[e]		
	r value		p value
Attention domain	−0.519		< 0.001
Orientation domain	−0.815		< 0.001
Memory domain	−0.751		< 0.001

Abbreviation: POD, participants who developed postoperative delirium; Non POD, participants who did not develop postoperative delirium.

Notes:

[a.] Internal consistency was assessed by using *Cronbach's alpha coefficient.*

[b.] Split-halves reliability was determined by *Pearson* correlation coefficient of two parts of CAM-S, namely odd items and even items.

[c.] Inter-rater reliability was assessed by using *intra-class correlation coefficients* (*ICC*).

[d.] Discriminate validity was evaluated by *Student t-test*, comparing the CAM-S scores between participants who developed postoperative delirium and those who did not develop delirium diagnosed by CAM.

[e.] Criterion-related validity was assessed by *spearman correlation coefficients* between domain scores of CAM-S and concurrent cognitive domains.