Confusion assessment method: a systematic review and meta-analysis of diagnostic accuracy

Qiyun Shi[1,2]
Laura Warren[3]
Gustavo Saposnik[2]
Joy C MacDermid[1]

[1]Health and Rehabilitation Sciences, Western University, London, Ontario, Canada; [2]Stroke Outcomes Research Center, Department of Medicine, St Michael's Hospital, University of Toronto, Toronto, Ontario, Canada; [3]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

Background: Delirium is common in the early stages of hospitalization for a variety of acute and chronic diseases.

Objectives: To evaluate the diagnostic accuracy of two delirium screening tools, the Confusion Assessment Method (CAM) and the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU).

Methods: We searched MEDLINE, EMBASE, and PsychInfo for relevant articles published in English up to March 2013. We compared two screening tools to Diagnostic and Statistical Manual of Mental Disorders IV criteria. Two reviewers independently assessed studies to determine their eligibility, validity, and quality. Sensitivity and specificity were calculated using a bivariate model.

Results: Twenty-two studies (n = 2,442 patients) met the inclusion criteria. All studies demonstrated that these two scales can be administered within ten minutes, by trained clinical or research staff. The pooled sensitivities and specificity for CAM were 82% (95% confidence interval [CI]: 69%–91%) and 99% (95% CI: 87%–100%), and 81% (95% CI: 57%–93%) and 98% (95% CI: 86%–100%) for CAM-ICU, respectively.

Conclusion: Both CAM and CAM-ICU are validated instruments for the diagnosis of delirium in a variety of medical settings. However, CAM and CAM-ICU both present higher specificity than sensitivity. Therefore, the use of these tools should not replace clinical judgment.

Keywords: confusion assessment method, diagnostic accuracy, delirium, systematic review, meta-analysis

Introduction

Delirium is characterized by acute onset of an altered level of consciousness with fluctuating levels of orientation, memory, thought, and/or behavior.[1] It is commonly observed among patients with an acute medical condition, especially patients in the internal medicine, neurology, psychiatry, and geriatric wards. Delirium has been associated with unfavorable outcomes including higher mortality, longer hospitalization, and a greater degree of dependence after discharge.[2] Therefore, early recognition and prevention of delirium may improve outcomes in hospitalized patients.

Currently, two standard diagnostic criteria on delirium are the Diagnostic and Statistical Manual of Mental Disorders (DSM) IV criteria[3] and International Classification of Diseases-10.[4] However, neither of these diagnostic tools can be easily applied to daily bedside practice. Therefore, a variety of screening tools such as the Confusion Assessment Method (CAM),[5] the Delirium Rating Scale,[6] the Intensive Care Delirium Screening Checklist,[7] and the Nursing Delirium Screening

Correspondence: Qiyun Shi
Health & Rehabilitation Sciences, Western University, Room 1014, Elborn College, 1201 Western Road, London, Ontario, Canada, N6G 1H1
Email qshi26@uwo.ca

Scale[8] have been developed. Among them, CAM is one of the most widely used diagnostic instruments for clinical and research purposes with proven psychometric properties.[9] It was developed by Inouye et al based on DSM III for the purpose of enabling nonpsychiatric trained clinicians to identify delirium.[5,10] Ely et al developed the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU),[11,12] an instrument specifically designed to assess nonverbal patients (ie, mechanically ventilated) based on the CAM algorithm. The objective of this study was to investigate whether a nonlanguage based method has impact on diagnostic accuracy of the scale and CAM-ICU performance in identifying delirium in a different spectrum of patients in comparison to CAM. Therefore, in this review, we compared CAM and CAM-ICU to DSM IV, a golden standard for delirium diagnosis applied to verbal versus nonverbal patients.

Materials and methods
Search sources and searches
Two reviewers (QS and LW) conducted a literature search of MEDLINE, EMBASE, and PsychInfo in March 2013. Computer searches based on keywords were conducted. References from previously retrieved articles were also searched. Details of search strategy can be found in Appendix 1.

Study selection
Research articles examining either CAM or CAM-ICU were included for review if they met the following inclusion criteria: (1) the study was designed as an observational, cross-sectional or case series study; (2) the reference test was DSM IV for delirium diagnosis; (3) diagnostic accuracy estimates were reported in the paper, including sensitivity, specificity, true positive, false positive, true negative, and false negative, or sufficient detail to derive these numbers; and (4) the study was written in English.

We excluded papers on prevalence of delirium without diagnostic accuracy data, reference test other than DSM IV, or if the definition of delirium was unclear in the original paper. Figure 1 shows a flow chart of the search.

Data extraction and quality assessment
Data were extracted to a form which included the following information: first author, year of publication, study population characteristics, name of tool, assessor of screening and reference test, diagnostic cut point used, and length of administration.

Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) guidelines[13] were employed for this systematic review to assess the study quality. For the purposes of this systematic review, we decided that a low risk of bias was assumed if all the questions were answered "yes." If an answer was either "no" or "unclear," a high risk of bias or "unclear" was assigned to the corresponding domain. In the patient selection domain, we considered the studies which included conditions related to mental disorders (ie, dementia, psychosis) which mimic delirium as an "appropriate inclusion" because we believe a good scale can discriminate between people who have delirium and people who have other mental disorders. For the index domain, we considered that a threshold was prespecified in translation versions. In domain 4, flow and timing, we considered 3 hours between the index test and reference standard as a reasonable time interval due to the fluctuations of delirium presentation. We appraised other items following guidelines.[13] For the applicability domain, studies were considered to be "high risk of bias" if no explicit description of inclusion and exclusion criteria were reported.

All data extraction and quality assessment were conducted by two reviewers (QS and LW). Discrepancies were resolved by discussion. A third reviewer (GS) was consulted if discrepancies remained.

Data synthesis and analysis
The estimates for sensitivity, specificity, and negative and positive likelihood ratios were computed using bivariate models[14] and compared to the Moses-Littenberg method.[15] We used a bivariate model which preserves the two-dimensional nature of sensitivity and specificity and treats these two estimates as a paired index, thus accounting for the possibility of correlation which other methods do not address. In contrast to a traditional funnel plot, which uses straight lines for pooled estimates, the bivariate model produces an ellipse with a pooled mean sensitivity/specificity, 95% confidence interval, and 95% prediction ellipse. As both the bivariate model and the Moses-Littenberg method require a 2×2 data table, in this meta-analysis, we either retrieved the data directly from the study or generated numbers[16] based on the information presented in the original paper. Bivariate analyses were performed with PROC NLMixed with SAS (version 9.2, SAS Institute Inc, Cary, NC, USA) and figures were plotted using R (version 2.14, Lucent Technologies, Murray Hill, NJ, USA). We also built a summary receiver operating characteristic (SROC) curve using RevMan.[17] Statistical significance was set a priori at an alpha level of 0.05.
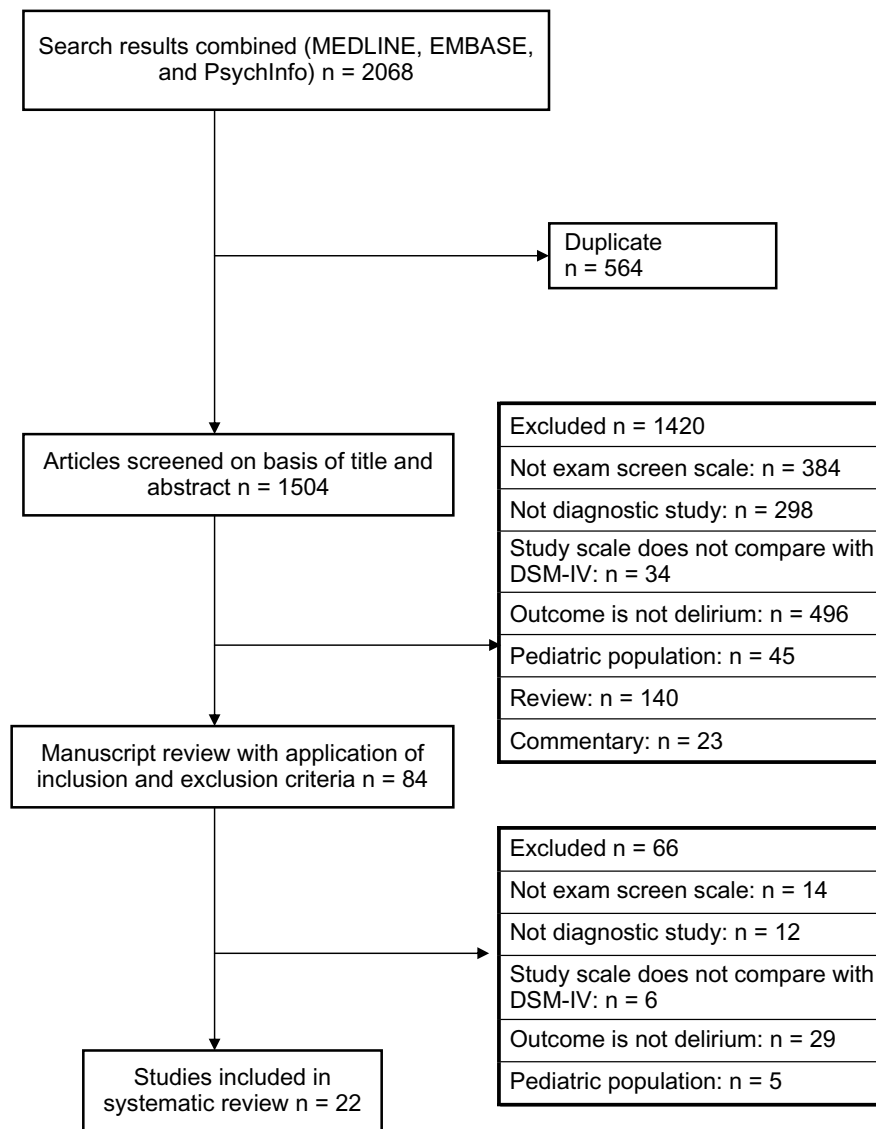
Figure 1 Result of literature search.
Abbreviation: DSM-IV, Diagnostic and Statistical Manual of Mental Disorders IV.

Results
Characteristics of studies

A total of 22 studies[8,11,12,18–36] were identified for this systematic review (Table 1). Nine studies examined CAM (n = 1,033) while 13 assessed for CAM-ICU (n = 1,409). Overall, the mean age of patients included in these studies ranged from 54 to 85 years old. The percentage of patients diagnosed with delirium varied from 14%[25] to 87%.[12] The most common study populations were ventilated intensive care unit patients,[11,12] geriatric inpatients,[19,21,22] or postoperative patients.[8,25,26] Most studies examined the accuracy of delirium diagnoses made by nondelirium health professionals such as general practitioners, nurses, or trained research assistants compared to delirium experts such as psychiatrists

or geriatricians. In these studies, both CAM and CAM-ICU could be administered as a quick questionnaire. A number of studies[18,21,27,29] focused on delirium which developed in the early stage of an acute disease while others monitored the patients from admission to discharge.[11,12,24]

Study quality

There were no disagreements in quality assessment between the reviewers that affected the categorization of studies as high or low risk of bias (Table 3). All 22 studies were considered to have a low risk of bias with respect to the reference test part. However, 13 studies had a high risk of bias for the patient selection criteria due to the inappropriate exclusion of demented or psychotic patients who were easily

Table I Characteristics of studies included in systematic review

Author	Place of study	Target population	Screening test*	Assessor of screening test	Length of screening test administration (mean)	Assessor of reference test	When test was administrated	Sample size	Mean/ median age	Incidence of delirium (%)
Ely[11,12]	US	Ventilated ICU patients	CAM-ICU	Nurse	2 min	Psychiatrist, neurologist, geriatrician	Daily after admission	96	55.3	83
Ely[11,12]	US	Ventilated ICU patients	CAM-ICU	Nurse or intensivist	Not reported	Psychiatrist, geriatrician	Daily after admission	38	60	87
Fabbri[18]	Brazil	Inpatients >60 admitted to emergency department	Brazilian version of CAM-ICU	Geriatrician	Not reported	Psychiatrist	Within 24 h of admission	100	73.8	17
Laurila[21]	Finland	Inpatients >70 admitted to acute geriatric wards	Finnish version of CAM	Geriatrician	Not reported	Geriatrician	Within 6 h of admission in work days	81	85	39.5
Lin[23]	Taiwan	Ventilated ICU patients	Chinese version of CAM-ICU	Research assistant	Not reported	Psychiatrist	Within 5 days of admission	109	D: 73.9 ND: 73.3	22.4
Gonzalez[19]	Spain	Inpatients >65	Spanish version of CAM	General physician or psychiatrist	7 min	Psychiatrist	Not reported	133	77.7	24
Leung[22]	Hong Kong	Inpatients >65 admitted to rehabilitation unit	CAM, Nu-DESC	Family physician	CAM: 10 min Nu-DESC: 1 min	Psychiatrist	Within 48 hours of admission	100	83.4	25
Radtke[25]	Germany	Postoperative patients in recovery room	CAM, Nu-DESC	Trained research assistant, supervised by a psychiatrist	Not reported	Same as screening test	Patients ready for discharge from recovery room	154	53.8	14
Ryan[27]	Ireland	Patients admitted to palliative care unit	CAM	Non consultant hospital doctor	Not reported	Psychiatrist	Within 24 h of admission	52	67.3	31.5
Guenther[32]	Germany	Patients admitted to surgical ICU	CAM-ICU	Intensivist and medical student	50 seconds	Psychiatrist	Within 4 h of admission	44	D: 74 ND: 61	46
Luetz[8]	Germany	Postoperative patients admitted to ICU	CAM-ICU Nu-DESC	Physician and nurse	Not reported	Psychiatrist or intensivist	1st postoperative day	156	Range 64–80	40
Radtke[26]	Germany	Postoperative patients	CAM, Nu-DESC	Trained research assistant, supervised by a psychiatrist	CAM: 5 min Nu-DESC: 1.27 min	Same as screening test	12 h before surgery and ended at 6th postoperative day or at hospital discharge	88	D: 68.7 ND: 64.7	19
Gusmao-Flores[30]	Brazil	Patients admitted to medical and surgical ICU	Portuguese version of CAM-ICU	Intensivist	Not reported	Psychiatrist, neurologist,	Not reported	119	57	38.6
van Eijk[28]	The Netherlands	Patients admitted to ICU	Dutch version of CAM-ICU	Nurse	Not reported	Psychiatrist, geriatricians, neurologist	Not reported	282	59	28

Study	Country	Patients	Test	Assessor	Time	Reference	Timing	N		
Heo[20]	Korea	Ventilated and non-ventilated patients in ICU	Korean version of CAM-ICU	Nurse	Not reported	Psychiatrist	Not reported	22	68	72.7
Neufeld[34]	US	Patients admitted in oncology unit	CAM-ICU	Trained evaluator	Not reported	Psychiatry resident	Daily assessment	117	57	26
Wongpakaran[29]	Thailand	Inpatients	Thai version of CAM	Family physician	7.77 min	Psychiatrist	Within 72 hours of admission	66	74.5	56.1
Adamis[33]	Greece	Patients admitted to ICU ≥24 hours	Greek version of CAM-ICU	ICU staff	6 min	Psychiatrist	Daily in the morning	71	61.5	33.8
Mitasova[24]	Czech Republic	Inpatient with acute stroke	CAM-ICU	Physician	A few min	Neurologist, neurophysiologist	1st day after stroke till at least 7 days	129	72.5	42.6
Ryan[35]	Ireland	Inpatients in general hospital	CAM	Medicine consultants or senior trainees	Not reported	Psychiatrist	Within 6 hours of admission	280	69	19.6
Thomas[31]	Germany	Inpatients ≥80 years old	CAM	Physician in training or gerontologist	Not reported	Psychologist or geriatrician	3rd day of admission	79	84.1	28
Wang[36]	China	Patients admitted to ICU	Chinese version of CAM-ICU	Nurse	Not reported	Neurologist	Daily assessment	126	74	48.4

Notes: *Indicates that if there is no specification, the English version was used in the assessment.
Abbreviations: CAM, Confusion Assessment Method; CAM-ICU, Confusion Assessment Method for the Intensive Care Unit; D, delirium group; Nu-DESC, the Nursing Delirium Screening Scale; ND, nondelirium group; ICU, intensive care unit.

confused with delirium patients. One quarter of studies did not explain whether the assessor was blinded either to the reference test or index test. Therefore, these potentially unblinded studies were assigned as "high risk" for bias. Approximately 20% of studies scored as unclear or a high risk of bias because the interval time between the two tests exceeded 3 hours. Most studies had high applicability. Only three studies had a high risk of bias for the patient selection criteria as a result of vague descriptions of inclusion and exclusion criteria.

Test accuracy of CAM and CAM-ICU

The psychometric properties of delirium screening tests are summarized in Table 2. Overall, CAM and CAM-ICU demonstrated similar performance characteristics when diagnosing delirium in both ventilated and nonventilated patients. The sensitivity of pooled CAM and CAM-ICU were 82% (95% confidence interval [CI]: 69%–91%) and 81% (95% CI: 57%–93%), respectively. The specificity of the two scales were 99% (95% CI: 87%–100%) and 98% (95% CI: 86%–100%), respectively. Figure 2 shows the bivariate summary estimates of sensitivity and specificity for CAM and CAM-ICU diagnostic accuracy. It also contains the 95% CI and prediction ellipses. The red dot shows the mean estimate of CAM-ICU, which was very close to CAM although the sensitivity was slightly lower. The solid line represents the 95% CI. We observed that CAM-ICU had a similar variance in comparison to CAM. The dotted lines represent the 95% prediction ellipses. As expected, CAM-ICU had a wider prediction range than CAM.

Figure 3 shows the SROC curves for all 22 studies. Six studies were identified to deviate from the average ROC curves. Three studies had high specificity but low sensitivity while another three had moderate to high sensitivity and specificity. Otherwise, all other studies showed high values for both sensitivity and specificity. The funnel plot (Figure 4) shows the distribution of sensitivity and specificity across studies. A greater degree of homogeneity in specificity was observed across studies in comparison to sensitivity.

Due to a lack of an adequate number of studies, we could not perform subgroup analysis to investigate potential factors, such as patient characteristics, type of disease, and difference in assessors influence on accuracy. On the other hand, CAM shows a good reliability across the studies, ranging from 86% to 94%. CAM-ICU demonstrated more variance in reliability, ranging from 20% to 96% (Table 2).

Table 2 Diagnostic accuracy of included screening test

Test	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Reliability (%)
CAM					
Fabbri[18]	94.1	96.4	84.2	98.7	86
Laurila[21]	81.2	83.7	76.5	87.2	Not reported
Gonzalez[19]	90	100	100	96.9	89
Radtke[25]	42.9	98.5	81.8	91.6	Not reported
Leung[22]	76	100	100	92.6	94
Ryan[27]	88.2	100	100	94.9	Not reported
Radtke[26]	74.9	100	Not reported	Not reported	100
Wongpakaran[29]	91.9	100	100	90.6	91
Adamis[33]	Rater 1: 87.5	Rater 1: 91.5	Rater 1: 84	Rater 1: 94	75
	Rater 2: 79.2	Rater 2: 87.2	Rater 2: 76	Rater 2: 89	
Ryan[35]	71	83	75	80	Not reported
Thomas[31]	74	100	100	91	Not reported
Pooled CAM	82 (69, 91)	99 (87, 100)			
CAM-ICU					
Ely[11,12]	Nurse 1: 100	Nurse 1: 98	Nurse 1: 92	Nurse 1: 100	96
	Nurse 2: 93	Nurse 2: 100	Nurse 2: 100	Nurse 2: 98	
Ely[11,12]	Nurse 1: 95	Nurse 1: 93	Not reported	Not reported	79–95
	Nurse 2: 96	Nurse 2: 93			
	Intensivist: 100	Intensivist: 89			
Lin[23]	Rater 1: 91	Rater 1: 98	Rater 1: 91	Rater 1: 98	91
	Rater 2: 95	Rater 2: 98	Rater 2: 91	Rater 2: 99	
Luetz[8]	81	96	Not reported	Not reported	89
Guenther[32]	Rater 1: 88	Rater 1: 100	Rater 1: 100	Rater 1: 91	Rater 1: 94
	Rater 2: 92	Rater 2: 100	Rater 2: 100	Rater 2: 94	Rater 2: 96
Gusmao-Flores[30]	72.5	96.2	90.6	87.4	Not reported
Heo[20]	Nurse 1: 89.8	Nurse 1: 72.4	Not reported	Not reported	81
	Nurse 2: 77.4	Nurse 2: 75.8			
van Eijk[28]	46.7	98.1	94.6	72.2	63
Neufeld[34]	18	99	88	85	20
Mitasova[24]	76	98	91	94	94
Wang[36]	Nurse 1: 91.8	Nurse 1: 90.8	Nurse 1: 90.3	Nurse 1: 92.2	92
	Nurse 2: 93.4	Nurse 2: 87.7	Nurse 2: 87.7	Nurse 2: 93.4	
Pooled CAM-ICU	81 (57, 93)	98 (86, 100)			

Note: Figures in brackets 95% CI.
Abbreviations: CAM, Confusion Assessment Method; CAM-ICU, Confusion Assessment Method for the Intensive Care Unit; NPV, negative predictive value; PPV, positive predictive value; CI, confidence interval.

Discussion

The diagnosis of delirium in different settings constitutes a challenge. Despite a variety of existing tools, limited information is available on the performance and psychometric properties of these tools. In the present systematic review and meta-analysis, we identified 22 papers with 2,449 patients which studied the accuracy of the two most commonly used tools (CAM and CAM-ICU) in the diagnosis of delirium over the past decade against the standard delirium diagnostic test (DSM-IV).

CAM and CAM-ICU are tools which can be administrated quickly, and have high sensitivity and specificity for early identification of delirium in a variety of hospitalized populations.

This review builds on previous work by Wei et al[37] in 2006. In addition to the incorporation of more recent studies,

we also employed QUADAS-2 for quality assessment and compared two statistical analysis methods when examining the accuracy performance of CAM and CAM-ICU, which had not previously been done. Overall, CAM and CAM-ICU showed moderate to high sensitivity, high specificity, and moderate to high reliability. Both tests can be administrated by health professionals with appropriate training to obtain the reliable results. In contrast to Wei et al's finding,[37] we observed higher specificity for both CAM and CAM-ICU in comparison to the sensitivity. This may be a result of differences in statistical methodology. However, the results shown in this paper support Wei et al's recommendation that given the relatively low values of sensitivity (approximately 80%–85%), clinicians should not base delirium diagnoses solely on either CAM or CAM-ICU assessments; rather they should use the tests in addition to their clinical judgment.[37]

Table 3 Quality assessment of included papers

Study	Risk of bias				Applicability		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Ely[11,12]	High	Low	Low	Unclear	Low	Low	Low
Ely[11,12]	High	Low	Low	Low	Low	Low	Low
Fabbri[18]	High	Unclear	Low	Low	High	Low	Low
Laurila[21]	Low	Low	Low	Low	Low	Low	Low
Lin[23]	High	Low	Low	High	Low	Low	Low
Gonzalez[19]	High	Low	Low	Low	High	Low	Low
Leung[22]	Low	Low	Low	Unclear	High	Low	Low
Ryan[27]	Low	Low	Low	Low	Low	Low	Low
Radtke[25]	High	Unclear	Unclear	Low	Low	Low	Low
Radtke[26]	High	Unclear	Unclear	Unclear	Low	Low	Low
Guenther[32]	High	Low	Low	High	Low	Low	Low
Luetz[8]	High	Low	Low	Low	Low	Low	Low
Wongpakaran[29]	Low	Low	Low	Low	Low	Low	Low
van Eijk[28]	Low	Low	Low	High	Low	Low	Low
Gusmao-Flores[30]	High	Low	Low	Low	Low	Low	Low
Heo[20]	High	Unclear	Unclear	Unclear	Low	Low	Low
Neufeld[34]	Low	Low	Low	Low	Low	Low	Low
Adamis[33]	Low	Low	Low	Low	Low	Low	Low
Mitasova[24]	High	Low	Low	High	Low	Low	Low
Ryan[35]	Low	Unclear	Unclear	High	Low	Low	Low
Thomas[31]	High	Low	Low	High	Low	Low	Low
Wang[36]	Low	Low	Low	Low	Low	Low	Low

It is not surprising that CAM and CAM-ICU demonstrated similar diagnostic accuracy as they are derived from the same algorithm. However, it is worth noticing that there are wider variances of sensitivity than specificity when applying these two screening scales. The CAM diagnostic
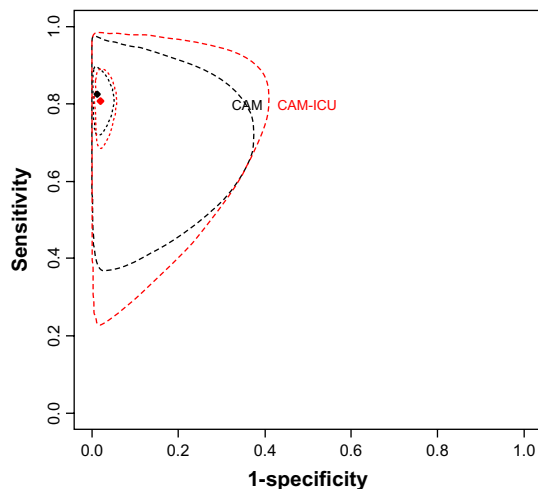
Figure 2 Bivariate estimate of sensitivity and specificity.
Notes: Bivariate summary estimates of sensitivity and specificity for delirium screening test with 95% confidence and prediction ellipses. Upper left dots represent sensitivity and specificity of the Confusion Assessment Method and the Confusion Assessment Method for the Intensive Care Unit. Smaller ellipses represent 95% confidence interval of sensitivity and specificity. Larger ellipses represent 95% confidence interval of prediction sensitivity and specificity. Red color represents the Confusion Assessment Method for the Intensive Care Unit and the black color represents the Confusion Assessment Method.
Abbreviations: CAM, Confusion Assessment Method; CAM-ICU, Confusion Assessment Method for the Intensive Care Unit.

algorithm is comprised of four components: (1) an acute onset of mental status changes of fluctuating course; (2) inattention; (3) disorganized thinking; and (4) an altered level of consciousness. The diagnosis of delirium is based on the presence of both component 1 and 2, and either 3 or 4. Missing any of these diagnostic criteria due to inadequate training in assessment may underestimate the percentage of delirium cases, especially hypoactive patients. Another possible reason for the variance in sensitivity is the difference in the rate of sedative drugs included in the studies which affects the diagnostic accuracy of delirium.[38]

We also observed a wide range of reliability across studies, especially for CAM-ICU. It is plausible that sedatives[39,40] widely used in critical care units were causing the fluctuation in delirium.[11,12,41] Therefore, it is important to adhere closely to the screening protocol with regard to the point of time assessment, observation period, as well as to provide sufficient training to medical or research staff when conducting delirium assessments.[42,43]

To our knowledge, this is the first systematic review and meta-analysis to evaluate CAM and CAM-ICU in comparison to DSM-IV using the bivariate model and Moses-Littenberg method. We appraised all studies with recently revised QUADAS-2 to assess quality. However, in the current review, we are only able to compare two of the most popular delirium screening tools. Further studies may be warranted to expand these findings.
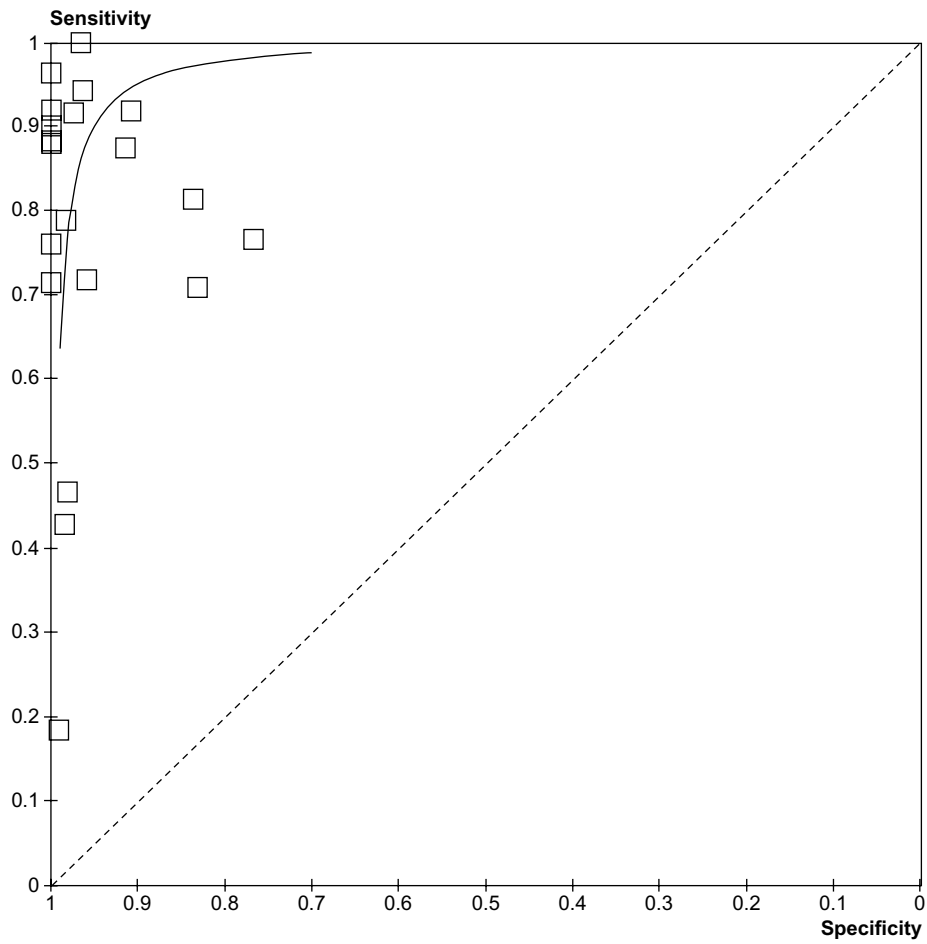
Figure 3 Summary receiver operating characteristic curve of delirium screen scales.

| Study	TP	FP	FN	TN	Sensitivity	Specificity	Sensitivity	Specificity
Adamis[33]	21	4	3	43	0.88 [0.68, 0.97]	0.91 [0.80, 0.98]		
van Eijk[28]	35	2	40	104	0.47 [0.35, 0.59]	0.98 [0.93, 1.00]		
Ely[11,12]	77	0	3	16	0.96 [0.89, 0.99]	1.00 [0.79, 1.00]		
Fabbri[18]	16	3	1	80	0.94 [0.71, 1.00]	0.96 [0.90, 0.99]		
Gonzalez[19]	27	0	3	93	0.90 [0.73, 0.98]	1.00 [0.96, 1.00]		
Guenther[32]	22	0	3	19	0.88 [0.69, 0.97]	1.00 [0.82, 1.00]		
Gusmao-Flores[30]	33	3	13	70	0.72 [0.57, 0.84]	0.96 [0.88, 0.99]		
Laurila[21]	26	8	6	41	0.81 [0.64, 0.93]	0.84 [0.70, 0.93]		
Leung[22]	19	0	6	75	0.76 [0.55, 0.91]	1.00 [0.95, 1.00]		
Lin[23]	22	2	2	76	0.92 [0.73, 0.99]	0.97 [0.91, 1.00]		
Luetz[8]	137	11	42	36	0.77 [0.70, 0.83]	0.77 [0.62, 0.88]		
Mitasova[24]	225	12	60	706	0.79 [0.74, 0.84]	0.98 [0.97, 0.99]		
Neufeld[34]	14	3	62	344	0.18 [0.10, 0.29]	0.99 [0.97, 1.00]		
Radtke[25]	9	2	12	131	0.43 [0.22, 0.66]	0.98 [0.95, 1.00]		
Radtke[26]	15	0	2	71	0.88 [0.64, 0.99]	1.00 [0.95, 1.00]		
Ryan[27]	15	0	2	37	0.88 [0.64, 0.99]	1.00 [0.91, 1.00]		
Ryan[35]	39	13	16	64	0.71 [0.57, 0.82]	0.83 [0.73, 0.91]		
Thomas[31]	15	0	6	58	0.71 [0.48, 0.89]	1.00 [0.94, 1.00]		
Wang[36]	56	6	5	59	0.92 [0.82, 0.97]	0.91 [0.81, 0.97]		
Wongpakaran[29]	34	0	3	29	0.92 [0.78, 0.98]	1.00 [0.88, 1.00]		
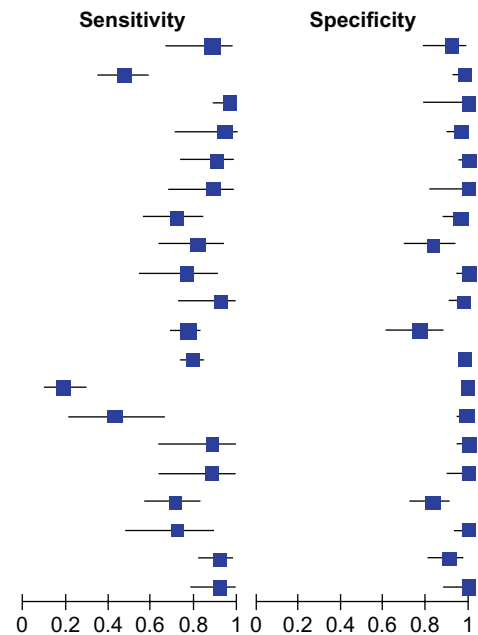
Figure 4 Funnel plot of delirium screen scales.
Abbreviations: TP, true positive; FP, false positive; FN, false negative; TN, true negative.

Conclusion

Both CAM and CAM-ICU are validated instruments in the diagnosis of delirium in a variety of medical settings, including the emergency department, the postoperative recovery room, in palliative care, the stroke unit, and the rehabilitation unit. Health professionals with appropriate training can achieve similar accuracy to experts specializing in psychiatric evaluation. CAM and CAM-ICU both present higher specificity than sensitivity. Health professionals should be cautiously interpreting these results as superior sensitivity is expected to a screening scale. The incidence of delirium may be underestimated due to the relatively high rate of false negatives in a low sensitivity scale. Therefore, CAM and CAM-ICU instruments should not be relied on alone for diagnosis, but that application of clinical judgment (presumably based on application of DSM) is essential to diagnose the presence and severity of delirium.

Authors' contributions

QS and JCM conceived the study. QS and LW performed the database searches. QS undertook the statistical analysis. QS contributed to the writing of the first draft of the manuscript. All of the authors contributed to and have approved the final manuscript. GS and JCM both contributed as senior authors.

Acknowledgment

Disclosure

The authors report no conflict of interest in this work.

References

1. Lipowski Z. Delirium in geriatric patients. In: *Delirium*: *Acute Confusional States*. New York: Oxford University Press; 1990.
2. Shi Q, Presutti R, Selchen D, Saposnik G. Delirium in acute stroke: a systematic review and meta-analysis. *Stroke*. 2012;43(3):645–649.
3. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, *Fourth Edition*: DSM-IV-TR. Washington. DC: American Psychiatric Association; 1994.
4. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization; 1992.
5. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Ann Intern Med*. 1990;113(12): 941–948.
6. Trzepacz PT, Baker RW, Greenhouse J. A symptom rating scale for delirium. *Psychiatry Res*. 1988;23(1):89–97.
7. Radtke FM, Franck M, Oppermann S, et al. [The Intensive Care Delirium Screening Checklist (ICDSC) – translation and validation of intensive care delirium checklist in accordance with guidelines.] *Anasthesiol Intensivmed Notfallmed Schmerzther*. 2009;44(2):80–86. German.
8. Luetz A, Heymann A, Radtke FM, et al. Different assessment tools for intensive care unit delirium: which score to use? *Crit Care Med*. 2010; 38(2):409–418.
9. Higgins JP, Altman DG, Gotzsche PC, et al; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
10. Inouye SK, Foreman MD, Mion LC, Katz KH, Cooney Jr LM. Nurses' recognition of delirium and its symptoms: comparison of nurse and researcher ratings. *Arch Intern Med*. 2001;161(20):2467–2473.
11. Ely EW, Inouye SK, Bernard GR, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA*. 2001;286(21): 2703–2710.
12. Ely EW, Margolin R, Francis J, et al. Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *Crit Care Med*. 2001;29(7): 1370–1379.
13. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536.
14. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982–990.
15. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med*. 1993;12(14):1293–1316.
16. Sackett DL, Haynes RB, Guyatt CH, Tugwell P. The selection of diagnostic tests. In: Sackett DL, Haynes RB, Guyatt GH, Tugwell P, editors. *Clinical Epidemiology*. *A Basic Science for Clinical Medicine*. 2nd ed. Boston: Little, Brown and Company; 1991.
17. Review Manager (RevMan) [computer program]. Version 5.0: Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2008.
18. Fabbri RM, Moreira MA, Garrido R, Almeida OP. Validity and reliability of the Portuguese version of the Confusion Assessment Method (CAM) for the detection of delirium in the elderly. *Arq Neuropsiquiatr*. 2001;59(2-A):175–179.
19. Gonzalez M, de Pablo J, Fuente E, et al. Instrument for detection of delirium in general hospitals: adaptation of the confusion assessment method. *Psychosomatics*. 2004;45(5):426–431.
20. Heo EY, Lee BJ, Hahm BJ, et al. Translation and validation of the Korean Confusion Assessment Method for the Intensive Care Unit. *BMC Psychiatry*. 2011;11:94.
21. Laurila JV, Pitkala KH, Strandberg TE, Tilvis RS. Confusion assessment method in the diagnostics of delirium among aged hospital patients: would it serve better in screening than as a diagnostic instrument? *Int J Geriatr Psychiatry*. 2002;17(12):1112–1119.
22. Leung JM, Leung VW, Leung CM, Pan PC. 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*. 2008;30(2):171–176.
23. Lin SM, Liu CY, Wang CH, et al. The impact of delirium on the survival of mechanically ventilated patients. *Crit Care Med*. 2004;32(11): 2254–2259.
24. Mitasova A, Kostalova M, Bednarik J, et al. Poststroke delirium incidence and outcomes: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *Crit Care Med*. 2012;40(2):484–490.
25. Radtke FM, Franck M, Schneider M, et al. Comparison of three scores to screen for delirium in the recovery room. *Br J Anaesth*. 2008;101(3): 338–343.
26. Radtke FM, Franck M, Schust S, et al. A comparison of three scores to screen for delirium on the surgical ward. *World J Surg*. 2010;34(3): 487–494.

27. Ryan K, Leonard M, Guerin S, Donnelly S, Conroy M, Meagher D. Validation of the confusion assessment method in the palliative care setting. *Palliat Med*. 2009;23(1):40–45.

28. van Eijk MM, van den Boogaard M, van Marum RJ, et al. Routine use of the confusion assessment method for the intensive care unit: a multicenter study. *Am J Respir Crit Care Med*. 2011;184(3):340–344.

29. Wongpakaran N, Wongpakaran T, Bookamana P, et al. Diagnosing delirium in elderly Thai patients: utilization of the CAM algorithm. *BMC Fam Prac*. 2011;12:65.

30. Gusmao-Flores D, Salluh JI, Dal-Pizzol F, et al. The validity and reliability of the Portuguese versions of three tools used to diagnose delirium in critically ill patients. *Clinics (Sao Paulo)*. 2011;66(11): 1917–1922.

31. Thomas C, Kreisel SH, Oster P, Driessen M, Arolt V, Inouye SK. Diagnosing delirium in older hospitalized adults with dementia: adapting the confusion assessment method to international classification of diseases, tenth revision, diagnostic criteria. *J Am Geriatr Soc*. 2012;60(8):1471–1477.

32. Guenther U, Popp J, Koecher L, et al. Validity and reliability of the CAM-ICU Flowsheet to diagnose delirium in surgical ICU patients. *J Crit Care*. 2010;25(1):144–151.

33. Adamis D, Dimitriou C, Anifantaki S, et al. Validation of the Greek version of confusion assessment method for the intensive care unit (CAM-ICU). *Intensive Crit Care Nurse*. 2012;28(6):337–343.

34. Neufeld KJ, Hayat MJ, Coughlin JM, et al. Evaluation of two intensive care delirium screening tools for non-critically ill hospitalized patients. *Psychosomatics*. 2011;52(2):133–140.

35. Ryan DJ, O'Regan NA, Caoimh RO, et al. Delirium in an adult acute hospital population: predictors, prevalence and detection. *BMJ Open*. 2013;3(1).

36. Wang C, Wu Y, Yue P, et al. Delirium assessment using Confusion Assessment Method for the Intensive Care Unit in Chinese critically ill patients. *J Crit Care*. 2013;28(3):223–229.

37. Wei LA, Fearing MA, Sternberg EJ, Inouye SK. The Confusion Assessment Method: a systematic review of current usage. *J Am Geriatr Soc*. 2008;56(5):823–830.

38. Gusmao-Flores D, Figueira Salluh JI, Chalhub RA, Quarantini LC. The confusion assessment method for the intensive care unit (CAM-ICU) and intensive care delirium screening checklist (ICDSC) for the diagnosis of delirium: a systematic review and meta-analysis of clinical studies. *Crit Care*. 2012;16(4):R115.

39. Ramsay M, Savege T, Simpson B, Goodwin R. Controlled sedation with alphaxalone-alphadolone. *Br Med J*. 1974;2(5920):656–659.

40. Riker RR, Picard JT, Fraser GL. Prospective evaluation of the Sedation-Agitation Scale for adult critically ill patients. *Crit Care Med*. 1999;27(7):1325–1329.

41. Bergeron N, Dubois M-J, Dumont M, Dial S, Skrobik Y. Intensive Care Delirium Screening Checklist: evaluation of a new screening tool. *Intensive Care Med*. 2001;27(5):859–864.

42. Vasilevskis EE, Pandharipande PP, Girard TD, Ely EW. A screening, prevention, and restoration model for saving the injured brain in intensive care unit survivors. *Crit Care Med*. 2010;38(10):S683–S691.

43. Spronk PE, Riekerk B, Hofhuis J, Rommes JH. Occurrence of delirium is severely underestimated in the ICU during daily care. *Intensive Care Med*. 2009;35(7):1276–1280.

Appendix 1: Search stategy

Database: Ovid MEDLINE
Searched March 11, 2013
1946 to March week 2 2013

1. exp Delirium/
2. deliri$.ti,ab.
3. or/1–2

Validated diagnosis filter

4. (exp sensitivity/ and specificity/) or sensitiv*.ti,ab.
5. *diagnosis/ or diagnos*.ti,ab.
6. *Diagnostic Tests, Routine/ or diagnostic.mp.
7. *diagnosis,differential/
8. di.fs.
9. (Test or assessment or scale or checklist or instrument).mp. [mp = title, abstract, original title, name of substance word, subject heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]
10. 3 and (or/4–8) and 9
11. limit 10 to English language

Embase Classic+Embase <1947 to 2013 week 11>
Search Strategy:

1. exp delirium/
2. deliri$.ti,ab.
3. 1 or 2
4. exp "SENSITIVITY AND SPECIFICITY"/
5. sensitivity.tw.
6. specificity.tw.
7. ((pre-test or pretest) adj probability).tw.
8. post-test probability.tw.
9. predictive value$.tw.
10. predictive value$.tw.
11. *Diagnostic Accuracy/
12. 3 and (or/4–11)

PsycINFO 1806 to March week 2 2013
Search Strategy:

1. (delirium or deliria).mp.
2. exp sensitivity/ and specificity/) or sensitiv*.ti,ab.
3. *diagnosis/ or diagnos*.ti,ab.
4. *Diagnostic Tests, Routine/ or diagnostic.mp.
5. (Test or assessment or scale or checklist or instrument).mp. [mp = title, abstract, heading word, table of contents, key concepts, original title, tests and measures]
6. 1 and (or/2–4) and 5