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

Qiyun Shi1,2
Laura Warren3
Gustavo Saposnik2
Joy C MacDermid1

1Health and Rehabilitation Sciences, Western University, London, Ontario, Canada; 2Stroke Outcomes Research Center, Department of Medicine, St Michael’s Hospital, University of Toronto, Toronto, Ontario, Canada; 3Dalla 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 criteria3 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),3 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@uwow.ca

This article was published in the following Dove Press journal:
Neuropsychiatric Disease and Treatment
18 September 2013
Number of times this article has been viewed
Scale\textsuperscript{a} have been developed. Among them, CAM is one of the most widely used diagnostic instruments for clinical and research purposes with proven psychometric properties.\textsuperscript{9} It was developed by Inouye et al based on DSM III for the purpose of enabling nonpsychiatric trained clinicians to identify delirium.\textsuperscript{5,10} Ely et al developed the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU),\textsuperscript{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\textsuperscript{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.\textsuperscript{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\textsuperscript{14} and compared to the Moses-Littenberg method.\textsuperscript{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 $2 \times 2$ data table, in this meta-analysis, we either retrieved the data directly from the study or generated numbers\textsuperscript{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.\textsuperscript{17} Statistical significance was set a priori at an alpha level of 0.05.
Results

Characteristics of studies
A total of 22 studies 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% to 87%. The most common study populations were ventilated intensive care unit patients, geriatric inpatients, or postoperative patients. 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 focused on delirium which developed in the early stage of an acute disease while others monitored the patients from admission to discharge.

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 1 Characteristics of studies included in systematic review

<table>
<thead>
<tr>
<th>Author</th>
<th>Place of study</th>
<th>Target population</th>
<th>Screening test*</th>
<th>Assessor of screening test</th>
<th>Length of screening test administration (mean)</th>
<th>Assessor of reference test</th>
<th>When test was administrated</th>
<th>Sample size</th>
<th>Mean/median age</th>
<th>Incidence of delirium (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ely1,12</td>
<td>US</td>
<td>Ventilated ICU patients</td>
<td>CAM-ICU</td>
<td>Nurse</td>
<td>2 min</td>
<td>Psychiatrist, neurologist, geriatrician</td>
<td>Daily after admission</td>
<td>96</td>
<td>55.3</td>
<td>83</td>
</tr>
<tr>
<td>Ely1,12</td>
<td>US</td>
<td>Ventilated ICU patients</td>
<td>CAM-ICU</td>
<td>Nurse or intensivist</td>
<td>Not reported</td>
<td>Psychiatrist, geriatrician</td>
<td>Daily after admission</td>
<td>38</td>
<td>60</td>
<td>87</td>
</tr>
<tr>
<td>Fabbri18</td>
<td>Brazil</td>
<td>Inpatients &gt;60 admitted to emergency department</td>
<td>CAM, Nu-DESC</td>
<td>Geriatrician</td>
<td>Not reported</td>
<td>Psychiatrist</td>
<td>Within 24 h of admission</td>
<td>100</td>
<td>73.8</td>
<td>17</td>
</tr>
<tr>
<td>Laurila21</td>
<td>Finland</td>
<td>Inpatients &gt;70 admitted to acute geriatric wards</td>
<td>Finnish version of CAM</td>
<td>Geriatrician</td>
<td>Not reported</td>
<td>Geriatrician</td>
<td>Within 6 h of admission in work days</td>
<td>81</td>
<td>85</td>
<td>39.5</td>
</tr>
<tr>
<td>Lin23</td>
<td>Taiwan</td>
<td>Ventilated ICU patients</td>
<td>Chinese version of CAM-ICU</td>
<td>Research assistant</td>
<td>Not reported</td>
<td>Psychiatrist</td>
<td>Within 5 days of admission</td>
<td>109</td>
<td>D: 73.9</td>
<td>22.4</td>
</tr>
<tr>
<td>Gonzalez19</td>
<td>Spain</td>
<td>Inpatients &gt;65</td>
<td>CAM, Nu-DESC</td>
<td>CAM: 10 min</td>
<td>Psychiatrist</td>
<td>Within 48 h of admission</td>
<td>100</td>
<td>83.4</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Leung22</td>
<td>Hong Kong</td>
<td>Inpatients &gt;65 admitted to rehabilitation unit</td>
<td>CAM, Nu-DESC</td>
<td>Same as screening test</td>
<td>Psychiatrist</td>
<td>Patients ready for discharge from recovery room</td>
<td>154</td>
<td>53.8</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Radtke25</td>
<td>Germany</td>
<td>Postoperative patients in recovery room</td>
<td>CAM, Nu-DESC</td>
<td>Not reported</td>
<td>Same as screening test</td>
<td>Supervised by a psychiatrist</td>
<td>Not reported</td>
<td>154</td>
<td>74</td>
<td>46</td>
</tr>
<tr>
<td>Ryan27</td>
<td>Ireland</td>
<td>Patients admitted to palliative care unit</td>
<td>CAM</td>
<td>Non consultant hospital doctor</td>
<td>Psychiatrist</td>
<td>Within 24 h of admission</td>
<td>52</td>
<td>67.3</td>
<td>31.5</td>
<td></td>
</tr>
<tr>
<td>Guenther25</td>
<td>Germany</td>
<td>Patients admitted to surgical ICU</td>
<td>CAM-ICU</td>
<td>Intensivist and medical student</td>
<td>Psychiatrist</td>
<td>Within 4 h of admission</td>
<td>44</td>
<td>D: 74</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Luetz8</td>
<td>Germany</td>
<td>Postoperative patients admitted to ICU</td>
<td>CAM-ICU</td>
<td>Not reported</td>
<td>Psychiatrist or intensivist</td>
<td>1st postoperative day</td>
<td>156</td>
<td>ND: 61</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Radtke26</td>
<td>Germany</td>
<td>Postoperative patients</td>
<td>CAM, Nu-DESC</td>
<td>CAM: 5 min Nu-DESC: 1.27 min</td>
<td>Same as screening test</td>
<td>12 h before surgery and ended at 6th postoperative day or at hospital discharge</td>
<td>88</td>
<td>D: 68.7</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Gusmao-</td>
<td>Brazil</td>
<td>Patients admitted to medical and surgical ICU</td>
<td>Portuguese version of CAM-ICU</td>
<td>Not reported</td>
<td>Psychiatrist, neurologist, geriatrician</td>
<td>Not reported</td>
<td>119</td>
<td>57</td>
<td>38.6</td>
<td></td>
</tr>
<tr>
<td>Flores20</td>
<td>The Netherlands</td>
<td>Patients admitted to ICU</td>
<td>Dutch version of CAM-ICU</td>
<td>Nurse</td>
<td>Psychiatrist, geriatricists, neurologist</td>
<td>Not reported</td>
<td>749</td>
<td>59</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

*CAM-ICU, CAM, Nu-DESC, CAM-iCU, Brazilian version of CAM-iCU, Finnish version of CAM, Spanish version of CAM, Portuguese version of CAM-iCU, Dutch version of CAM-ICU.
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.

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).
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 administered 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 which 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.

<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>Reliability (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fabbri</td>
<td>94.1</td>
<td>96.4</td>
<td>84.2</td>
<td>98.7</td>
<td>86</td>
</tr>
<tr>
<td>Laurila</td>
<td>81.2</td>
<td>83.7</td>
<td>76.5</td>
<td>87.2</td>
<td>Not reported</td>
</tr>
<tr>
<td>Gonzalez</td>
<td>90</td>
<td>100</td>
<td>100</td>
<td>96.9</td>
<td>89</td>
</tr>
<tr>
<td>Radtke</td>
<td>42.9</td>
<td>98.5</td>
<td>81.8</td>
<td>91.6</td>
<td>Not reported</td>
</tr>
<tr>
<td>Leung</td>
<td>76</td>
<td>100</td>
<td>100</td>
<td>92.6</td>
<td>94</td>
</tr>
<tr>
<td>Ryan</td>
<td>88.2</td>
<td>100</td>
<td>100</td>
<td>94.9</td>
<td>Not reported</td>
</tr>
<tr>
<td>Radtke</td>
<td>74.9</td>
<td>100</td>
<td>Not reported</td>
<td>Not reported</td>
<td>100</td>
</tr>
<tr>
<td>Wongpakaran</td>
<td>91.9</td>
<td>100</td>
<td>100</td>
<td>90.6</td>
<td>91</td>
</tr>
<tr>
<td>Adams</td>
<td>Rater 1: 87.5</td>
<td>Rater 1: 91.5</td>
<td>Rater 1: 84</td>
<td>Rater 1: 94</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Rater 2: 79.2</td>
<td>Rater 2: 87.2</td>
<td>Rater 2: 76</td>
<td>Rater 2: 89</td>
<td></td>
</tr>
<tr>
<td>Ryan</td>
<td>71</td>
<td>83</td>
<td>75</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Thomas</td>
<td>74</td>
<td>100</td>
<td>100</td>
<td>91</td>
<td>Not reported</td>
</tr>
<tr>
<td><strong>Pooled CAM</strong></td>
<td><strong>82 (69, 91)</strong></td>
<td><strong>99 (87, 100)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAM-ICU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ely</td>
<td>Nurse 1: 100</td>
<td>Nurse 1: 98</td>
<td>Nurse 1: 92</td>
<td>Nurse 1: 100</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Nurse 2: 93</td>
<td>Nurse 2: 100</td>
<td>Nurse 2: 100</td>
<td>Nurse 2: 98</td>
<td></td>
</tr>
<tr>
<td>Ely</td>
<td>Nurse 1: 95</td>
<td>Nurse 1: 93</td>
<td>Not reported</td>
<td>Not reported</td>
<td>79–95</td>
</tr>
<tr>
<td></td>
<td>Nurse 2: 96</td>
<td>Nurse 2: 93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin</td>
<td>Rater 1: 91</td>
<td>Rater 1: 98</td>
<td>Rater 1: 91</td>
<td>Rater 1: 98</td>
<td>91</td>
</tr>
<tr>
<td>Luetz</td>
<td>81</td>
<td>96</td>
<td>Not reported</td>
<td>Not reported</td>
<td>89</td>
</tr>
<tr>
<td>Guenther</td>
<td>Rater 1: 88</td>
<td>Rater 1: 100</td>
<td>Rater 1: 100</td>
<td>Rater 1: 91</td>
<td>Rater 1: 94</td>
</tr>
<tr>
<td>Guzmao-Flores</td>
<td>72.5</td>
<td>96.2</td>
<td>90.6</td>
<td>87.4</td>
<td>Not reported</td>
</tr>
<tr>
<td>Heo</td>
<td>Nurse 1: 89.8</td>
<td>Nurse 1: 72.4</td>
<td>Not reported</td>
<td>Not reported</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Nurse 2: 77.4</td>
<td>Nurse 2: 75.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Eijk</td>
<td>46.7</td>
<td>98.1</td>
<td>94.6</td>
<td>72.2</td>
<td>63</td>
</tr>
<tr>
<td>Neufeld</td>
<td>18</td>
<td>99</td>
<td>88</td>
<td>85</td>
<td>20</td>
</tr>
<tr>
<td>Mitasova</td>
<td>76</td>
<td>98</td>
<td>91</td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Wang</td>
<td>Nurse 1: 91.8</td>
<td>Nurse 1: 90.8</td>
<td>Nurse 1: 90.3</td>
<td>Nurse 1: 92.2</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Nurse 2: 93.4</td>
<td>Nurse 2: 87.7</td>
<td>Nurse 2: 87.7</td>
<td>Nurse 2: 93.4</td>
<td></td>
</tr>
<tr>
<td><strong>Pooled CAM-ICU</strong></td>
<td><strong>81 (57, 93)</strong></td>
<td><strong>98 (86, 100)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 administered 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 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 administered by health professionals with appropriate training to obtain the reliable results. In contrast to Wei et al’s finding, 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.
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 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.\(^3\)

We also observed a wide range of reliability across studies, especially for CAM-ICU. It is plausible that sedatives\(^1\) widely used in critical care units were causing the fluctuation in delirium.\(^2\) 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.\(^2\)

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 QUA-DAS-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.

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

The authors thank Michael Manno for assistance of figures preparation.

Disclosure

The authors report no conflict of interest in this work.

References


Appendix 1: Search strategy

Database: Ovid MEDLINE

Searches March 11, 2013
1946 to March week 2 2013

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

Validated diagnosis filter
4. (exp sensitivity/ and specificity/) or sensitiv*.ti,ab.
5. *diagnosis/ or diagnos*.ti,ab.
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. delir$.ti,ab.
3. 1 or 2
4. exp “SENSITIVITY AND SPECIFICITY”/
5. sensitivity.tw.
6. specificity.tw.
7. ((pre-test or pretest) adj probability).tw.
8. post-test probability.tw.
9. predictive value$.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