Reliability and validity assessment of the Japanese version of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU)

Yuji Koga\textsuperscript{a,}\textsuperscript{*}, Ryosuke Tsuruta\textsuperscript{b}, Hiroaki Murata\textsuperscript{c}, Koji Matsuo\textsuperscript{d}, Tatsuhiko Ito\textsuperscript{e}, E. Wesley Ely\textsuperscript{f,g}, Ayumi Shintani\textsuperscript{h,i,j}, Hiroya Wakamatsu\textsuperscript{k}, Masamitsu Sanui\textsuperscript{l}, Hiroaki Yamase\textsuperscript{a}

\textsuperscript{a} Department of Nursing, Yamaguchi University Hospital, Yamaguchi, Japan
\textsuperscript{b} Advanced Medical Emergency and Critical Care Center, Yamaguchi University, Yamaguchi, Japan
\textsuperscript{c} Department of Nursing, Jikei University School of Medicine, Tokyo, Japan
\textsuperscript{d} Division of Neuropsychiatry, Department of Neuroscience, Yamaguchi University Graduate School of Medicine, Yamaguchi, Japan
\textsuperscript{e} Department of Psychiatry, Jikei University School of Medicine, Tokyo, Japan
\textsuperscript{f} Department of Medicine, Division of Allergy, Pulmonary and Critical Care Medicine Center for Health Services Research, Vanderbilt School of Medicine, Nashville, TN, USA
\textsuperscript{g} Geriatric Research, Education and Clinical Center (GRECC), Department of Veterans Affairs Medical Center, Tennessee Valley Healthcare System, Nashville, USA
\textsuperscript{h} Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, USA
\textsuperscript{i} Center for Health Services Research, Vanderbilt University School of Medicine, Nashville, USA
\textsuperscript{j} Center for Quantitative Sciences, Vanderbilt University School of Medicine, Nashville, USA
\textsuperscript{k} Department of Intensive Care, Yamaguchi University, Yamaguchi, Japan
\textsuperscript{l} Department of Anesthesiology and Critical Care Medicine, Jichi Medical University Saitama Medical Center, Saitama, Japan

Accepted 20 October 2014

KEYWORDS
CAM-ICU; Delirium; ICU; Japan;

Summary
Objective: Delirium may lead to adverse outcomes in patients with serious conditions, but is often under-diagnosed due to inadequate screening. The Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) is an established method for assessing delirium in the ICU. The validity and reliability of the Japanese version of the CAM-ICU has not, however, been verified, and we undertook this study to verify these parameters.

\textsuperscript{*} Corresponding author at: Department of Nursing, Yamaguchi University Hospital, 1-1-1, Minami-Kogushi, Ube 755-8505, Yamaguchi, Japan. Tel.: +81 836 22 2111; fax: +81 836 22 2729.
E-mail address: yuukoga@yamaguchi-u.ac.jp (Y. Koga).

http://dx.doi.org/10.1016/j.iccn.2014.10.002
0964-3397/© 2014 Elsevier Ltd. All rights reserved.

Reliability; Validity

Research methodology: CAM-ICU validity and reliability were assessed in two Japanese ICUs. Using the evaluation of the DSM-IV-TR in the psychiatrists group as the standard criteria for delirium diagnosis, we compared the evaluation of the Japanese version of the CAM-ICU between the research nurses group and the staff nurses group.

Results: According to DSM-IV-TR criteria, the prevalence of delirium was 22.0%, and according to CAM-ICU delirium was found in 22.0% with Research Nurses and 19.5% with Staff Nurses. CAM-ICU sensitivity ratings were 83% and 78%, while their specificity ratings were 95% and 97%, respectively. The Kappa inter-rater reliability was good ($\kappa = 0.85$), and Cronbach’s alpha coefficient was 0.69 (95% CI: 0.57–0.79). Mean rating time for the CAM-ICU was 2.5–2.8 minutes for Research Nurses and Staff Nurses, respectively.

Conclusion: The Japanese version of the CAM-ICU has comparable validity and reliability as a delirium assessment tool in surgical patients in two Japanese ICUs. With training, CAM-ICU can be incorporated into daily clinical practice.

Implications for Clinical Practice

- The risk of delirium in seriously ill ICU patients is high and proper evaluation of ICU delirium is important.
- Japanese ICUs, do not follow a standardised method for diagnosing delirium, and the diagnosis is solely made by psychiatrists.
- Psychiatrists, who are not often present in the ICU.
- We evaluated the validity and reliability of the Japanese version of the confusion assessment method for the ICU (CAM-ICU) in order to standardise critical care in Japan.
- Using a valid and reliable instrument may contribute to developing a national database that could contribute internationally to epidemiology on this important form of organ dysfunction in critical illness.

Introduction

Delirium is highly prevalent in intensive care units (ICUs) (Morandi et al., 2008; Ouimet et al., 2007), and up to 80% of intubated patients may develop its symptoms (Ely et al., 2001a). Risk factors for delirium include host, severe disease, and iatrogenic factors (Smith et al., 2009). The condition is associated with increased mortality, prolonged hospitalisation, and long-term cognitive impairment (Ely et al., 2004; Pisani et al., 2009; van Rompaey et al., 2009). In addition, it is reported that delirium is associated with longer ICU stay duration and higher medical expenses (Ely et al., 2001a; Milbrandt et al., 2004).

Over the past decade, the assessment standard for delirium has been the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR). Although the DSM-IV-TR has been supplanted by the DSM-5, the current edition does not substantially alter the delirium diagnosis. The manual lists the following constellation of symptoms with a diagnosis of delirium: changes and acute variations of mental state, upset mental state, attention deficit disorders, altered level of consciousness, hallucinations, delusions, illusions, and thought disorders (APA, 2000, 2013; Morandi et al., 2008). Because delirium has, by definition, a fluctuating course, it is often underdiagnosed in many clinical cases (Girard et al., 2008). Without the use of screening tools, research has shown that delirium is overlooked in 75% of ICU admissions (Spronk et al., 2009).

The Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) is a widely accepted measure—alongside the Intensive Care Delirium Screening Checklist—as a standard method of evaluating delirium in intensive care facilities. The current pain, agitation, and delirium (PAD) guideline recommends that these measures be used as delirium monitoring tools (Barr et al., 2013; Bergeron et al., 2001; Ely et al., 2001b). The CAM-ICU, a modified version of the Confusion Assessment Method (CAM), was designed for healthcare professionals other than psychiatrists (Inouye et al., 1990), and it has both high sensitivity and specificity when compared to the DSM-IV (Ely et al., 2001b). Furthermore, research has found it to have excellent inter-rater reliability when used by staff nurses (Pun et al., 2005).

In Japan, for both research and clinical purposes, a valid and reliable delirium diagnostic-scale is also required. It is true that a Japanese translation of the CAM-ICU and a training manual endorsed by the original authors (Tsuruta et al., 2002) has become available, but its validity and reliability have not been evaluated. Moreover, the Japanese version contains changes made to two items concerning attention deficits. The original version employs an alphabet test (grasp hand when saying “A” in “SAVEAHART”), which the Japanese version substitutes for a number test (grasp hand when saying “1” in “6153191124”) because the English alphabet is not common in Japan (Tsuruta et al., 2002).

Research question

The purpose of this research is to verify the validity and reliability of the Japanese version of the CAM-ICU.
Therefore, the sensitivity and specificity of the CAM-ICU in relation to the DSM-IV-TR are calculated. Moreover, the reliability between evaluators of CAM-ICU and inter-rater reliability (the kappa coefficient) and internal consistency (Cronbach’s $\alpha$ coefficient) of the CAM-ICU are calculated.

Methods

Sample size

Sample size calculations were performed based on sensitivity. We computed the sample size for this study to obtain the necessary number of patients to yield a lower confidence interval (CI) of 70%. Assuming a prevalence of 30% for delirium, and a point estimate for sensitivity of 85%, we found that a sample size of 80 patients was required. The lower 95% CI for specificity was estimated to be 75% for a point estimate of 85% specificity.

Participants

This study was conducted at two university-affiliated, tertiary care center ICUs—one in Yamaguchi Prefecture (with 12 beds) and the other in Tokyo (with 10 beds).

The clinical research ethics committee of each facility approved the study, and this study conforms to the guidelines set by the Declaration of Helsinki. Elective surgical patients who were 20 years of age or older and were scheduled for an ICU admission duration of 24 hours or more were included in this study. Participants gave their written consent during the lead-up to their surgeries.

Patients’ level of sedation was assessed by intensive care medical specialists using the RASS. Those with a score of $-3$ to $+4$ were then evaluated for delirium by the research team comprising psychiatrists, research nurses, and staff nurses.

We excluded patients who had histories of mental illness, a poor grasp of Japanese, significant hearing or visual impairments (in order to ensure appropriate communication), a necessity to use muscle relaxants, and RASS scores lower than $-3$. We further excluded those who developed a cerebral infarction during or after surgery.

Measurements

We noted each patient’s demographic data, medical history (including history of mental illness), in addition to their reasons for, and duration of, ICU admission. Additionally, we employed the following measures:

(a) The Japanese version of the CAM-ICU for a dichotomous classification of delirium vs. non-delirium. The Japanese version of the CAM-ICU (Tsuruta et al., 2002) was created using the original version of the CAM-ICU (Ely et al., 2001b), which was translated into Japanese, and then was back-translated into English by one of the original measure’s Japanese-fluent authors. The original CAM-ICU creators then checked any discrepancies between the original and the back-translated versions.

(b) The Richmond Agitation-Sedation Scale (RASS) to evaluate participants’ abilities to complete the CAM-ICU (Sessler et al., 2002).

(c) APACHE II, to evaluate participants’ illness severity (Knaus et al., 1985).

(d) DSM-IV-TR diagnostic criteria for delirium, which enabled a licensed psychiatrist to make a dichotomous classification (delirium/no delirium) (APA, 2000).

Procedure

Three groups performed the delirium diagnosis, and the group of psychiatrists diagnosed patients daily. The group of research nurses all had master’s degrees in nursing science, and they received CAM-ICU evaluation training by one of the authors of the Japanese version of the CAM-ICU. The group of staff nurses all had more than four years of ICU clinical experience, and they received CAM-ICU evaluation training from the research nurse group. CAM-ICU evaluation training was performed using the Japanese version of the CAM-ICU training manual (Tsuruta et al., 2002).

Psychiatrists performed a delirium assessment using DSM-IV-TR criteria within an arbitrarily chosen time frame (10 AM to 5 PM). After the psychiatrist’s evaluation, the research nurse performed a delirium assessment using the CAM-ICU within a window of two hours. Following this assessment, the staff nurse performed a second CAM-ICU evaluation.

Delirium is, by definition, a fluctuating condition; therefore, all delirium evaluations were completed within two hours to get the most accurate possible assessment. Each subsequent delirium evaluator was unaware of previous evaluation results, and the assessment was performed once for each patient.

Statistical analysis

To determine the validity of the Japanese version of CAM-ICU, we calculated its sensitivity, specificity, positive predictive value, and negative predictive value in relation to the DSM-IV-TR. In addition, as an index of reliability, we calculated inter-rater reliability between the research nurse and the staff nurse for CAM-ICU through the Kappa coefficient ($\kappa$). Internal consistency was evaluated by Cronbach’s $\alpha$. We performed the analysis using R version 3.0.1 (www.r-project.org), with continuous variables expressed as mean ± standard deviation. Probability and coefficients were expressed using a 95% CI.

Results

Participants

From the two ICUs, a total of six psychiatrists, three research nurses, and five staff nurses participated in this study. All participants understood the purpose of this research and consented to participate in the study.

Ninety-nine patients were admitted to the two ICUs between September 2012 and June 2013. Of these, we excluded two (2%) due to post-surgical cerebral infarction,
Meanwhile, 20 psychiatrist Japanese delirium 18.4% patient According Delirium-related and negatively correlated with the diagnosis of delirium (r = 0.59—0.96), and the evaluation interval between the psychiatrist and the research nurse was 20.7 ± 21.4 min, and the evaluation interval between the research and staff nurses was 24.1 ± 26.5 min.

Validity

The results of the comparison between patients’ DSM-IV-TR results (psychiatrist) and CAM-ICU results (nurses) are available in Table 2. In comparison with the DSM-IV-TR, the research nurse and staff nurse evaluation results demonstrated the CAM-ICU to possess satisfactory sensitivity of 83% (59—96%) and 78% (52—94%), specificity of 95% (87—99%) and 97% (89—100%), positive predictive value of 83% (59—96%) and 88% (62—98%), and negative predictive value of 97% (89—100%) and 94% (85—98%), respectively.

Our analyses examined all instances of included data (82 participants) from the two facilities. Although we performed a secondary analysis of the first ICU’s data (62 patients) in order to consider the differences between the facilities, no meaningful differences emerged.

Reliability

We performed a pairwise comparison of inter-rater reliability for the research/staff nurses’ CAM-ICU evaluations. Both evaluation results were matched for the 78 patients, and \( \kappa \) was 0.85 (0.64–1.00). We also calculated each CAM-ICU result’s \( \kappa \) coefficient (see Table 3). Cronbach’s \( \alpha \) coefficient in the research nurse evaluation was 0.69 (0.57–0.579) and 0.71 (0.59–0.80) in the staff nurse evaluation, indicating good internal consistency.

As was the case for reliability, our primary analyses employed all data from the two facilities. Although we performed secondary analyses on the first ICU’s data (62 patients) for comparative purposes, no significant differences emerged.

### Table 1: Primary patient characteristics.

<table>
<thead>
<tr>
<th>Gender</th>
<th>( n ) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>48 (58.5%)</td>
</tr>
<tr>
<td>Female</td>
<td>34 (41.5%)</td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td>68.5 ± 10.3</td>
</tr>
<tr>
<td>APACHE II (mean ± SD)</td>
<td>14.7 ± 4.7</td>
</tr>
</tbody>
</table>

### Table 2: Validity analysis of the Japanese version of the CAM-ICU.

<table>
<thead>
<tr>
<th>Evaluator</th>
<th>Sensitivity (%) (95% CI)</th>
<th>Specificity (%) (95% CI)</th>
<th>PPV (%) (95% CI)</th>
<th>NPV (%) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research nurse (( n = 3 ))</td>
<td>83 (59—96)</td>
<td>95 (87—99)</td>
<td>83 (59—96)</td>
<td>97 (89—100)</td>
</tr>
<tr>
<td>Staff nurse (( n = 5 ))</td>
<td>78 (52—94)</td>
<td>97 (89—100)</td>
<td>88 (62—98)</td>
<td>94 (85—98)</td>
</tr>
</tbody>
</table>

PPV, positive predictive value; NPV, negative predictive value.

### Table 3: Reliability analysis of the Japanese version of the CAM-ICU.

<table>
<thead>
<tr>
<th>Evaluator</th>
<th>( \kappa ) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>0.85 (0.64–1.00)</td>
</tr>
<tr>
<td>FEATURE 1</td>
<td>0.22 (0.03–0.41)</td>
</tr>
<tr>
<td>FEATURE 2</td>
<td>0.90 (0.68–1.11)</td>
</tr>
<tr>
<td>FEATURE 3</td>
<td>0.93 (0.32–0.75)</td>
</tr>
<tr>
<td>FEATURE 4</td>
<td>0.69 (0.47–0.90)</td>
</tr>
</tbody>
</table>

Delirium-related findings

According to the psychiatrists’ evaluations (i.e., DSM-IV-TR), 18 patients (22.0%) experienced delirium. Using the Japanese version of the CAM-ICU, delirium prevalence was 19.7% (15 patients) when scored by the research nurses and 18.4% (14 patients) by the staff nurses. The mean time the psychiatrist spent on delirium assessment was 9.3 ± 2.7 min. Meanwhile, the mean time the research nurses spent on evaluation was 2.5 ± 3.6 min, with the staff nurses spending 2.8 ± 4.5 min. The mean period between ICU admission and delirium evaluation lasted 28.9 ± 33.0 hours. The evaluation

A total of 82 patients—62 from the first institution and 20 from the second—were subject to our analysis. The mean patient age was 68.5 ± 10.3 years. Eleven patients (11.8%) were intubated. Patient demographic variables are summarized in Table 1.
Discussion

When compared to the reference standard (DSM-IV-TR), our data demonstrate that the CAM-ICU’s Japanese version possesses robust validity and reliability. These results are consistent with numerous other publications regarding the use of the CAM-ICU in other languages.

Validity

According to meta-analytic research, the pooled sensitivity of the CAM-ICU is 80.0% (77.1—82.6%), with pooled specificity being 95.9% (94.8—96.8%) (Gusmao-Flores et al., 2012), and these findings are identical to the present results. Therefore, the Japanese-language version of the CAM-ICU yielded high sensitivity and specificity, demonstrating criterion-related validity.

Reliability

Between the research and staff nurses, the Japanese CAM-ICU displayed high inter-rater reliability (κ = 0.85). Other language versions (original English, Spanish, Korean, and Greek versions) have yielded similar results (Adamis et al., 2012; Ely et al., 2001b; Heo et al., 2011; Tobar et al., 2010). This shows that the CAM-ICU evaluation training was effective.

With regard to the CAM-ICU findings’ κ coefficients, Feature 1’s (acute onset or fluctuating course) degree of coincidence (κ = 0.22) was low, compared with versions of this measure in other languages (Adamis et al., 2012; Ely et al., 2001b). Feature 1 was judged not only from the patient’s mental status at a certain evaluation point, but from various information obtained over the preceding 24 hours, including nurse’s records. Therefore, the quality of information affecting Feature 1’s results might be suboptimal. It is possible that either there was a lack of motivation in some of the staff nurses who participated in the study or that other nurses that did not participate in this study were less sensitive to the patient’s mental status, preventing them from providing precise documentation in the medical records. More rigorous quality-assurance measurements should have been taken to improve this feature.

The Japanese version of the CAM-ICU uses not the alphabet but the number for the evaluation of Feature 2 (inattention). With respect to Feature 2, we found a high degree of agreement (κ = 0.90), which was superior to that witnessed in other language versions (Adamis et al., 2012; Heo et al., 2011). This shows that the Japanese version’s use of numbers rather than the audio attention screening test (ASE) using the alphabet that is included in the all other language versions may be distinctly advantageous in terms of reliability.

In the present study, sedation level was maintained at the RASS of −0.33 ± 2.5, which is in the context of the light sedation management recommended by the PAD guideline (Barr et al., 2013). In most previous CAM-ICU validation studies, the sedation ranges were not clearly described (Adamis et al., 2012; Ely et al., 2001b; Heo et al., 2011; Tobar et al., 2010), but they were assumingly heavier than what is allowed for by the current standard practice. Considering the similarity of the sensitivities and specificities between these studies and ours, it is suggested that the CAM-ICU is valid for use within a wide range of sedation values.

Limitations

Although the validity and reliability of the Japanese version of the CAM-ICU have been verified in this study, data collections were performed only in the daytime. Therefore, the true incidence of delirium in the Japanese ICUs remains to be investigated.

Conclusion

This study suggests that the Japanese version of the CAM-ICU is a valid, reliable tool for screening delirium. With proper training, it will become a relevant tool in the daily clinical practice of Japanese ICUs. Future studies are warranted to investigate the precise incidence of delirium and the validation of this screening tool with a large cohort.

Acknowledgments

We would like to thank the psychiatrists who provided their expertise in consultation-liaison psychiatry (Drs. Toshio Matsubara, Koji Otsuki, Teruyuki Hobara, Aya Fujioka and Fumihiro Higuchi), Ms. Yoko Yamaguchi of the Jikei University School of Medicine Hospital ICU, and the ICU nurses who assessed patient delirium.

Funding: The authors have no sources of funding to declare.

Conflicts of interest: The authors have no conflict of interest to declare.

References


R version 3.0.1 (Reference: www.r-project.org).


