Journal Pre-proof

Diagnostic accuracy of the CAM-ICU and ICDSC in detecting intensive care unit delirium: A bivariate meta-analysis

Ting-Jhen Chen , Yi-Wei Chung , Hui-Chen (Rita) Chang , Pin-Yuan Chen , Chia-Rung Wu , Shu-Hua Hsieh , Hsiao-Yean Chiu

Please cite this article as: Ting-Jhen Chen , Yi-Wei Chung , Hui-Chen (Rita) Chang , Pin-Yuan Chen , Chia-Rung Wu , Shu-Hua Hsieh , Hsiao-Yean Chiu , Diagnostic accuracy of the CAM-ICU and ICDSC in detecting intensive care unit delirium: A bivariate meta-analysis, *International Journal of Nursing Studies* (2020), doi: https://doi.org/10.1016/j.ijnurstu.2020.103782

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Diagnostic accuracy of the CAM-ICU and ICDSC in detecting intensive care unit delirium: A bivariate meta-analysis

Ting-Jhen Chen [a,b], Yi-Wei Chung [c], Hui-Chen (Rita) Chang [b,d], Pin-Yuan Chen [e,f], Chia-Rung Wu [a,g], Shu-Hua Hsieh [a,g], Hsiao-Yean Chiu [a,h]

AFFILIATIONS:

[a]School of Nursing, College of Nursing, Taipei Medical University, Taipei, Taiwan

[b]School of Nursing, Faculty of Science, Medicine and Health, University of Wollongong, Northfields Ave, Wollongong, NSW 2522, Australia

[c]Department of Cardiology, Taipei Medical University Shuang Ho Hospital, New Taipei City, Taiwan

[d]Illawarra Health and Medical Research Institute, Northfields Ave, Wollongong, NSW 2522, Australia

[e]School of Medicine, Chang Gung University, Taoyuan, Taiwan

[f]Neurosurgical Department, Chang Gung Memorial Hospital, Keelung, Taiwan.

[g]Department of Nursing, Far Eastern Memorial Hospital, New Taipei, Taiwan

[h]Research Center of Sleep Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

Corresponding authors

- Hsiao-Yean Chiu

Address: 250 Wu-Hsing St., Taipei, 110 Taiwan

Abstract

Background: Delirium is a critical and highly prevalent problem among critically ill patients. The Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) and the Intensive Care Delirium Screening Checklist (ICDSC) are the most recommended assessment tools for detecting intensive care unit (ICU) delirium.

Objectives: To synthesize the current evidence and compared the diagnostic accuracy of the two tools in the detection of delirium in adults in ICUs.

Design: Systematic review and meta-analysis.

Data source: A comprehensive search of the following electronic databases was performed using PubMed, Embase, CINAHL and ProQuest Dissertations and Theses A&I. The date range searched was from database inception to April 26, 2019.

Review methods: Two researchers independently identified articles, systematically abstracted data and evaluated the sensitivity and specificity of the CAM-ICU or the ICDSC against standard references. Bivariate diagnostic statistical analysis with a random-effects model was performed to summarize the pooled sensitivity and specificity of the two tools.

Results: In total, 29 CAM-ICU and 12 ICDSC studies were identified. The pooled sensitivity was 0.84 and 0.83 and pooled specificity was 0.95 and 0.87 for the CAM-ICU and the ICDSC, respectively. The CAM-ICU had higher summary specificity than the ICDSC did ($p = 0.04$). The percentage of hypoactive delirium, ICU type, use of mechanical ventilation, number of participants, and female percentage moderated the accuracy of the tools. Most of the domains of patient selection,

2

index test, reference standards, and flow and timing were rated as having a low or unclear risk of

bias.

Conclusions: Although both the CAM-ICU and the ICDSC are accurate assessment tools for

screening delirium in critically ill patients, the CAM-ICU is superior in ruling out patients without

ICU delirium and detecting delirium in patients in the medical ICU and those receiving mechanical

ventilation. Further investigations are warranted to validate our findings. The study protocol is

registered at PROSPERO (CRD42020133544).

Keywords: Confusion Assessment Method for the Intensive Care Unit, critical care, delirium,

Intensive Care Delirium Screening Checklist, intensive care unit

What is already known about the topic?

- Critically ill patients have a high risk of delirium, which affects patients' physical and

  psychological prognoses.

- The Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) and the Intensive

  Care Delirium Screening Checklist (ICDSC) are valid and reliable assessment tools for the daily

  screening of intensive care unit delirium, and they are commonly used in clinical settings.

What this paper adds

- The CAM-ICU and the ICDSC had comparable pooled sensitivity (84% vs. 83%), but the

  CAM-ICU had higher pooled specificity than the ICDSC did (95% vs. 87%, $p = 0.04$).

- The CAM-ICU is superior in detecting delirium in patients in the medical ICU and those

  receiving mechanical ventilation.

1. Introduction

Delirium, a common type of acute brain dysfunction in critically ill patients, is characterized by the acute disturbance of consciousness, disorientation, inattention, and fluctuating mental status. It is divided into three subtypes according to psychomotor activity levels: hyperactive, hypoactive, and mixed (Lipowski, 1989; Liptzin and Levkoff, 1992); hypoactive delirium is the most frequent but the most difficult to observe (Krewulak et al., 2018). Factors including age, sex, disease severity, mechanical ventilator use, and intensive care unit (ICU) type have been linked with the occurrence of ICU delirium (Ely et al., 2001; Pandharipande et al., 2008; Zaal et al., 2015). Such acute psychiatric changes may lead to prolonged mechanical ventilation and ICU or hospital stay, increased mortality, and impaired long-term cognitive function (Lin et al., 2004; Salluh et al., 2015; Van den Boogaard et al., 2012). Therefore, Society of Critical Care Medicine (SCCM) guidelines recommend that delirium be detected in patients in the ICU by using valid assessment tools (Barr et al., 2013).

The Confusion Assessment Method for the Intensive Care Unit (CAM-ICU, Ely et al., 2001) and Intensive Care Delirium Screening Checklist (ICDSC, Bergeron et al., 2001) are the most widely used tools for detecting delirium in critically ill patients. Both the CAM-ICU and the ICDSC were established in 2001 on the basis of *Diagnostic and Statistical Manual of Mental Disorders (DSM)* criteria. The four-feature CAM-ICU was initially applied for nonverbal patients receiving mechanical ventilation. A flow sheet form was subsequently developed for convenient use in clinical practice (Ely et al., 2001). The eight-domain ICDSC is used to evaluate the features of

delirium and can be rapidly applied at bedside for all patients in an ICU. The two scales have been

translated into different languages with high reliability and validity (Danzeng et al., 2019; Larsen et

al, 2019; Nishimura et al., 2016). The National Institute for Health and Care Excellence (NICE) and

the SCCM have produced inconsistent recommendations; the NICE recommends the CAM-ICU for

detecting ICU delirium (Young et al., 2010), whereas the SCCM reports that both the tools can be

used to assess ICU delirium (Barr et al., 2013). This makes the section of the optimal tool for

screening ICU delirium difficult. In addition, the diagnostic accuracy of the two scales has varied in

relevant studies.

To date, two meta-analyses comparing the CAM-ICU and the ICDSC have been published (both

in 2012); however, the inclusion of different studies contributed to inconsistent accuracy results

(Neto et al., 2012; Gusmao-Flores et al., 2012). Additionally, neither publication compared the

diagnostic accuracy or examined the moderating effects of study features and patient characteristics

on the diagnostic accuracy of the two scales. Moreover, because more studies have been published

since the two meta-analyses were conducted, new data may address whether the CAM-ICU or the

ICDSC is a superior screening tool. Therefore, an updated and comprehensive diagnostic

meta-analysis is of clinical importance.

We compared the diagnostic accuracy of the CAM-ICU and the ICDSC against reference

standards (i.e., various editions of the *DSM* or *International Classification of Diseases* [*ICD*]) in

detecting ICU delirium. In addition, potential moderating effects on diagnostic accuracy were

investigated.

6

2. Materials and methods

The study protocol is registered at PROSPERO (CRD42020133544).

2.1. Study identification

We conducted a comprehensive search of electronic databases, including PubMed, Embase,

CINAHL, and ProQuest Dissertations and Theses A&I. Each database was searched from its

inception to April 26, 2019. The following keyword combinations developed with the librarian's

assistance were used: (ICU OR critical) AND (sensitivity OR specificity OR valid*) AND

(CAM-ICU or "confusion assessment method for the intensive care unit") and (ICDSC OR

"Intensive Care Delirium Screening Checklist"). An example of a search string is presented in Table

S1. In addition, we manually searched the reference lists of primary articles included in the

meta-analysis to identify more relevant articles for review.

2.2. Inclusion criteria and study selection

We included full-text studies assessing the sensitivity and specificity of the CAM-ICU or the

ICDSC against reference standards (i.e., various editions of the *DSM* or *ICD*) in adult patients (aged

≥18 years) who were admitted to an ICU. Incomplete published theses and dissertations were also

included if they met the aforementioned criteria. No language restrictions were applied. Studies

published in conference proceedings or book chapters without full text were excluded. Full-text

articles were reviewed after duplicate and irrelevant studies were excluded on the basis of title and

abstract screening by two independent reviewers (TJC and CRW). Any disagreement was resolved

in discussion with a third reviewer (HYC).

2.3. Scales

The CAM-ICU can be applied for verbal and nonverbal patients in the ICU. Four features are assessed in this scale, namely (1) acute change or fluctuating course of mental status, (2) inattention, (3) disorganized thinking, and (4) altered level of consciousness. The CAM-ICU returns a dichotomous value of either delirium or no delirium. If features 1 and 2 and are preset with either feature 3 or 4, the patient has a delirious status (Ely et al., 2001). The ICDSC is used for evaluating delirium on the basis of eight domains (i.e., level of consciousness, inattention, disorientation, hallucination, psychomotor activity, speech or mood disturbance, sleep disturbance, and symptom fluctuation). Patients are considered to have delirium if their ICDSC score is ≥4 (Bergeron et al., 2001). Other details regarding the two tools are presented in Table 1.

2.4. Data extraction

Relevant data, namely author, publication year, study country, study design, ICU type, number of participants, participant age, percentage of female participants, Acute Physiology and Chronic Health Evaluation II score, standard reference, duration between index and reference tests, and the cutoff point of the index test, were independently extracted by the two reviewers (TJC and CRW) by using a predesigned extraction form. If more than one pair of sensitivity and specificity values were reported in an included study, we selected the highest Youden's index value (Fluss et al., 2005). All disagreements were resolved through discussion or consultation with the third reviewer (HYC) for consensus.

8

2.5. *Assessments of methodological quality*

The risk of bias for each study was independently assessed in accordance with the Quality

Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2, Whiting et al., 2011) by the two

reviewers (TJC and CRW). The QUADAS-2 consists of four domains: patient selection, index test,

reference standard, and flow and timing. The first three domains concurrently evaluate applicability.

The risk of bias in each domain is judged as "high," "low," or "unclear" on the basis of a set of

signaling questions. The risk of bias was rated "low" if all signaling questions yielded "yes"

answers. If any domain had at least one "no" response, the study was rated as having a "high" risk

of bias. Studies had an "unclear" risk of bias if insufficient information was reported.

2.6. *Statistical analysis*

Data analyses were conducted using Stata version 14 with the metandi and midas user-written

commands (StataCorp LP, College Station, TX, USA). A bivariate diagnostic statistical analysis

with a random-effects model was conducted to produce pooled sensitivity and specificity. If any of

the cells in the 2×2 table had values of 0, the value of 0.5 was added to the cell to prevent

sensitivity and specificity from equaling 1 (Walter, 2002). The I^2 statistic, representing the total

variance percentage across the included studies, was used to assess heterogeneity. An I^2 value of

approximately 0% represents homogeneity, and that greater than 50% indicates substantial

between-study heterogeneity (Higgins and Thompson, 2002). Univariate moderator and

metaregression analyses were further performed to explore the causes of heterogeneity. The prior

causes were examined by entering predefined covariates into the moderator and metaregression

analyses on the basis of the aforementioned predesigned extraction form. To ensure sufficient data

for analyses, at least two studies were included in each group of moderator analyses. Deeks' funnel

plot was used to detect potential publication bias, with p values of >0.1 indicating no publication

bias (Deeks et al., 2005). We compared the two instruments by using the PROC MIXED module in

SAS version 3.9 (SAS Institute Inc., Cary, NC, USA), and the restricted maximum likelihood

estimation method was used to estimate covariance parameters.

3. Results

3.1. Search results

As presented in Figures 1 and 2, the comprehensive review initially included 663 studies that

assessed the CAM-ICU and 352 studies that assessed the ICDSC. After duplicates were removed

using Endnote and titles and abstracts were screened, 36 studies on the CAM-ICU and 13 studies on

the ICDSC were considered potentially suitable. No study was excluded because of unavailable full

text. After the full text of the studies was reviewed, seven studies on the CAM-ICU were excluded.

Among them, six did not employ the *DSM* or *ICD* as the reference standard and one used the same

sample for data estimations. For the ICDSC, one study was excluded because it used the same

sample for estimations of diagnostic accuracy. The lists of excluded studies after the full-text review

are presented in Table S2. Finally, we identified 29 studies on the CAM-ICU (Adamis et al., 2012;

Akinci et al., 2005; Aljuaid et al., 2018; Barman et al., 2018; Boettger et al., 2017; Bui et al., 2017;

Chanques et al., 2018; Chuang et al., 2007; Danzeng et al., 2019; Ely et al., 2001; Ely et al., 2001;

Gusmao-Flores et al., 2011; Heo et al., 2011; Karlicic et al., 2016; Koga et al., 2015; Larsen et al.,

2019; Lin et al., 2004; Luetz et al., 2010; Mitášová et al., 2010; Mitasova et al., 2012; Nishimura et al., 2016; Pipanmekaporn et al., 2014; Selim et al., 2018; Tobar et al., 2010; Toro et al., 2009; van Eijk et al., 2011; Van Eijk et al., 2009; Vreeswijk et al., 2009; Wang et al., 2013) and 12 studies assessing the ICDSC (Barman et al., 2018; Boettger et al., 2017; Chanques et al., 2018; Domenico and Federica, 2012; George et al., 2011; Gusmao-Flores et al., 2011; Kose et al., 2016; Larsen et al., 2019; Nishimura et al., 2016; Radtke et al., 2009; Van Eijk et al., 2009; Bergeron et al., 2001) for further analyses (Figures 1 and 2).

3.2. *Study characteristics*

The details of the study characteristics are presented in Tables S3 and S4.

3.2.1. CAM-ICU

Of the 29 studies that used the CAM-ICU as the index test, 1 enrolled 5 different group of patients: acute stroke, mechanical ventilation, dementia, depression, and schizophrenia (Mitášová et al., 2010). However, we included data for only patients with acute stroke and mechanical ventilation because the other three study populations did not report the occurrence of delirium. Therefore, 30 pairs of diagnostic sensitivity and specificity values were used in the meta-analysis. The meta-analysis included a total of 4002 participants with a mean age of 63.3 years. Prospective and cross-sectional study designs were used in 22 and 7 trials, respectively. Of the studies, 8, 7, and 15 were conducted in medical, surgical, and mixed ICUs, respectively. The *DSM* was the most commonly used reference for delirium diagnosis (n = 28).

3.2.2. ICDSC

As presented in Table S4, the 12 studies on the ICDSC included 1326 participants with a mean age of 62.7 years. Eight and four studies used prospective and cross-sectional designs, respectively. Three, three, and six studies were conducted in medical, surgical, and mixed ICUs, respectively. An ICDSC score of 4 was used as the cutoff point in nine studies. One study used 3 as the cutoff value on the basis of the highest Youden's index value (George et al., 2011). The two remaining studies did not report a cutoff value for the index test. The *DSM* (i.e., fourth edition [*DSM-IV*], *DSM-IV-Text Revision* [*TR*], or fifth edition [*DSM-5*]) was the most frequently used reference standard for delirium diagnosis (n = 9).

3.3. Study quality

For the patient selection domain, one study that simultaneously investigated the CAM-ICU and the ICDSC had an unclear risk of bias because it had insufficient information to determine whether patient enrollment was consecutive or random. For the index test and reference standard domains, 19 and 8 studies respectively on the CAM-ICU and the ICDSC were rated as having an unclear risk of bias because inadequate information was available to determine whether the index test results were interpreted without knowledge of reference test findings or vice versa. For the ICDSC, one study was rated as having a high risk of bias in the index test domain because it did not use a predefined threshold; the optimal cutoff value was selected post hoc. In the flow and timing domains, five and six studies on the CAM-ICU and the ICDSC, respectively, had an unclear risk of bias because they reported unclear time intervals between the index test and reference standard. One study on the CAM-ICU and one study on the ICDSC were rated as having a high risk of bias in the

flow and timing domains because they reported inappropriate time intervals between the index test and reference standard.

Most of the studies on the CAM-ICU and the ICDSC were rated as having a "low" risk of bias in applicability (Tables S5 and S6), with the exception of one study with high risk because a cutoff value of 3 rather than 4 was used to determine the diagnostic accuracy of the ICDSC; this cutoff value might have affected the estimated diagnostic accuracy.

3.4. *Summary of the CAM-ICU and the ICDSC diagnostic accuracy*

Pooled sensitivity and specificity values with 95% confidence intervals (CIs) are listed in Table 2. For the CAM-ICU, the pooled sensitivity and specificity were 0.84 and 0.95 (95% CIs = 0.77 to 0.88 and 0.91 to 0.97), respectively. For the ICDSC, the pooled sensitivity and specificity were 0.83 and 0.87 (95% CIs = 0.74 to 0.90 and 0.78 to 0.93), respectively.

A significant difference was observed in pooled specificity but not pooled sensitivity between the CAM-ICU and the ICDSC ($p = 0.04$, degrees of freedom = 1), indicating that the CAM-ICU yielded a higher pooled specificity than did the ICDSC.

3.5. *Heterogeneity of the CAM-ICU and the ICDSC diagnostic accuracy*

As presented in Table S7, substantial between-study heterogeneity for pooled sensitivity and specificity was observed for both the CAM-ICU and the ICDSC. The I^2 value for the pooled sensitivity and specificity of the CAM-ICU were 87.99% and 96.64%, respectively ($Q = 241.53$ and 863.78, respectively, both $p < 0.001$). For the ICDSC, the I^2 value was 82.70% and 87.93% for pooled sensitivity and specificity, respectively ($Q = 63.57$ and 91.10 respectively, both $p < 0.001$).

Therefore, moderator and metaregression analyses were performed to further explore the causes of heterogeneity.

3.6. Moderator and metaregression analyses

The results of the moderator and metaregression analyses are presented in Tables 3 and 4. For the CAM-ICU, the percentage of hypoactive delirium detection was positively correlated with pooled specificity (B = 3.50, p = 0.04). Studies conducted in medical ICUs had higher pooled sensitivity values compared with those conducted in nonmedical ICUs (93% vs. 79%, p = 0.01). Studies that included patients undergoing mechanical ventilation at enrollment yielded greater sensitivity and specificity compared with those that did not include all patients using mechanical ventilators (sensitivity: 95% vs. 81%, p = 0.02; specificity: 99% vs. 94%, p = 0.04).

For the ICDSC, the number of participants was positively associated with pooled specificity, and the percentage of female participants was positively associated with pooled sensitivity. This indicated that more participants yielded greater specificity (B = 0.007, p = 0.04) and that a larger percentage of female participants yielded greater sensitivity (B = 7.66, p = 0.001).

3.7. Publication bias

No significant publication bias was observed in the CAM-ICU or the ICDSC by using Deeks' funnel plot asymmetry test (all p > 0.01).

4. Discussion

The current diagnostic meta-analysis examined and compared the accuracy of the CAM-ICU and the ICDSC against reference standards. Both the tools had comparable pooled sensitivity (84%

14

vs. 83%). However, the CAM-ICU had higher pooled specificity than the ICDSC did (95% vs. 87%, $p = 0.04$). Our findings provide updated information because several studies have been conducted since the publication of the two previous meta-analyses (Neto et al., 2012; Gusmao-Flores et al., 2012). Because the current meta-analysis included more studies with numerous participants, our findings can be considered credible.

Hypoactive delirium is manifested as lethargy, decreased activity, and an inability to concentrate, and it occurs more frequently than the other two subtypes do (Krewulak et al., 2018); however, hypoactive delirium is also the most under-recognized and undiagnosed subtype in critically ill patients (Pandharipande et al., 2007). We demonstrated that the higher percentage of hypoactive delirium was associated with greater specificity when the CAM-ICU is used to assess ICU delirium. The assessment of patients' attention and thinking ability might enhance the identification of critically ill patients without hypoactive delirium. Features 2 and 3 (inattention and disorganized thinking, respectively) of the CAM-ICU require patients to cooperate with assessors (Ely et al., 2001). This may partially explain why the CAM-ICU can accurately exclude patients without hypoactive delirium. Because some of the included studies provided insufficient information regarding the hyperactive and mixed types of delirium, further investigations of the diagnostic accuracy of these tools for these two delirium types are warranted.

The CAM-ICU yielded higher sensitivity (93%) when administered in medical ICUs. Discrepancies in study populations when the instruments were initially developed may explain this difference. The CAM-ICU was initially developed to assess delirium in medical ICU patients, and

sedative use does not preclude the application of the CAM-ICU, even when patients are deeply

sedated (Ely et al., 2001). By contrast, the ICDSC was established based on a group of both medical

and surgical ICU patients, and deeply sedated patients were classified as "not possible to assess"

(Bergeron et al., 2001). A previous study demonstrated that agreement rates between the two tools

vary between patients in medical and surgical ICUs, suggesting that disease severity partially

affects the diagnostic accuracy of the tools (de Oliveira Fagundes et al. 2012). Therefore, we

inferred that differences in characteristics between the two study populations but not the properties

of the scales led to such results. However, because we did not observe a similar phenomenon in the

ICDSC use, our findings should be interpreted with caution. Further research on the effects of

administering the two assessment tools in different ICU types is required.

In the present study, the CAM-ICU administration in patients receiving mechanical ventilation

at enrollment had higher sensitivity and specificity compared with those without mechanical

ventilation. The high prevalence of ICU delirium in patients receiving mechanical ventilation and

the assessment content of the CAM-ICU may explain this finding. Mechanical ventilator use is

commonly accompanied by infusion sedation and physical restraints to manage agitation and

prevent self-removal of catheters (De Jonghe et al., 2013; Rose et al., 2016). Such circumstances

consequently increase the risk of delirium, resulting in delirium being easier to detect by using the

tool (Ouimet et al., 2007). The CAM-ICU features of inattention and disorganized thinking require

patient interaction with assessors; this interaction can effectively reveal the presence of delirium

even in patients receiving mechanical ventilation. However, the relationship between the ICDSC

and mechanical ventilation could not be validated in the current study because the included studies

provided insufficient information regarding mechanical ventilation use for further analysis.

We demonstrated that the number of participants was positively associated with the pooled

specificity of the ICDSC. An adequate number of participants is a critical component of diagnostic

test evaluations because large numbers of participants can prevent imprecise accuracy estimations

(Joneset al., 2003; Leeflang et al., 2008). Future studies on diagnostic tests should consider the

effect of the number of participants on the diagnostic accuracy of instruments.

The metaregression analysis indicated that a higher percentage of female participants in studies

assessing the ICDSC increased the sensitivity and reduced the specificity of the tool. Female

patients who are older (Ahmed et al., 2014), have a lower body mass index (Lee et al., 2011), have

mental disorders (Gottlieb et al., 2004; Van de Velde et al., 2010), or have preoperative dementia

(Azadet al., 2007; Kim et al., 2018) may have an increased risk of delirium. Nonetheless, because

the current study could not ascertain potential confounders or exact reasons, further investigations

are warranted to explore the effect of gender on the diagnostic accuracy of the ICDSC.

For study quality, we noted numerous instances of unclear bias risk in the index test and

reference standard domains. Insufficient information on whether the assessor was blinded to the

results of the index text and reference standard was the main reason for unclear bias risk. This

unclear risk may have produced inflated estimates of diagnostic accuracy (Ruitenberg et al., 2001;

Lijmer et al., 1999; Whiting et al., 2013). On the basis of these findings, our results should be

interpreted conservatively. In addition, we suggest a more rigorous study design involving assessor

blinding to the results of the index test and reference standard in future research.

An ideal instrument in terms of the detection of ICU delirium, diagnostic properties, delirium features, and feasibility should be considered. For diagnostic properties, both the CAM-ICU and the ICDSC yield comparably high sensitivity, indicating that the two tools can properly detect ICU delirium. Our findings further reveal that the specificity of the ICDSC was lower than that of the CAM-ICU; however, nurses should focus on sensitivity rather than specificity because high sensitivity can enable the accurate diagnosis of more true-positive cases and minimize the consequences of undiagnosed conditions (Lalkhen and McCluskey, 2008). Concerning delirium features, although both tools were established on basis of *DSM* criteria, the four features of the CAM-ICU and eight domains of the ICDSC were developed for clinical assessment. With respect to feasibility, the two assessments for ICU delirium can be completed within 2–3 minutes if the assessor is familiar with them. A qualitative study reported that nurses believe that the CAM-ICU is difficult to use in some situations, namely when patients are uncooperative or refuse the test and when nurses have time constraints or heavy workloads. In addition, nurses who are unfamiliar with the CAM-ICU initially require more time to complete it (Jung et al., 2013). The assessment approach requiring more patients' cooperation might explain why nurses believe the CAM-ICU is difficult to administer. The CAM-ICU is based on observation and interaction with patients at a single time point and requires additional attention screening tasks and organized thinking tests. By contrast, the ICDSC is based on continual observation during routine care, which may be more convenient for ICU nurses. Although the CAM-ICU and the ICDSC yield comparable sensitivity

and delirium features, the CAM-ICU requires additional tasks apart from routine care that might

increase the difficulty of clinical implementation among nurses who do not receive comprehensive

pretraining. We suggest that clinicians and nurses with comprehensive training can use the

CAM-ICU and the ICDSC in clinical practice as screening tools for ICU delirium and ICU

delirium–related treatment outcomes.

5. Limitations

The present diagnostic meta-analysis has several limitations. First, considerable between-study

heterogeneity and unequal sizes in the numbers of participants between the two tools were observed

in the included studies; therefore, our findings should be interpreted with caution. Second,

approximately one-third to one-half of the studies had an unclear risk of bias regarding assessor

blinding to the index test results or reference standards because insufficient information was

provided. This unclear risk may have affected the validity of our findings. Nonetheless, because

additional studies with numerous participants from various geographic areas were included in our

analysis, the external validity of the study findings is acceptable.

6. Conclusion

In summary, both the CAM-ICU and the ICDSC are accurate tools for detecting ICU delirium,

but compared with the ICDSC, the CAM-ICU is superior for ruling out patients without ICU

delirium. Considering the diagnostic properties, delirium features, and feasibility of the two tools,

both the CAM-ICU and the ICDSC are suitable for use by ICU nurses and clinicians who have

received comprehensive pretraining. Furthermore, the CAM-ICU is an adequately accurate

instrument for detecting delirium in patients in medical ICUs and those receiving mechanical

ventilation. Because of the small number of the ICDSC studies, more investigations of its

diagnostic accuracy in specific populations (i.e., patients receiving mechanical ventilation or those

in medical ICUs) are required. To accurately estimate the diagnostic accuracy of tools for the

detection of ICU delirium, future studies should include consecutively or randomly selected patients

and blind the assessor to the results of the index test and reference standard.

References

Adamis, D., Chara, D., Stella, A., Alexandros, Z., Irene, A., Athanasios, A., et al., 2012. Validation

of the Greek version of confusion assessment method for the intensive care unit (CAM-ICU).

Intensive Crit. Care Nurs. 28 (6), 337-343.

Ahmed, S., Baptiste, L., Elizabeth, L. S., 2014. Risk factors for incident delirium among older

people in acute hospital medical units: a systematic review and meta-analysis. Age and

ageing 43 (3), 326-333.

Akinci, S.B., Rezaki M., Ozdemir H., Çelikcan A., Kanbak M., Yorganci K., et al., 2005. Validity

and reliability of confusion assessment method for the intensive care unit. Türk. Anest. Rean.

Der. Dergisi. 33 (4), 333-341.

Aljuaid, M.H., Deeb A.M., Dbsawy, M., Alsayegh, D., Alotaibi, M., Arabi, Y.M., 2018.

Psychometric properties of the Arabic version of the confusion assessment method for the

intensive care unit (CAM-ICU). BMC psychiatry 18 (1),91.

Azad, N.A., Muneerah, A.B., Inge L.-E., 2007. Gender differences in dementia risk factors. Gender

medicine 4 (2), 120-129.

Barman, A., Pradhan, D., Bhattacharyya, P., Dey, S., Bhattacharjee, A., Tesia, S.S., Mitra, J.K., 2018.

Diagnostic accuracy of delirium assessment methods in critical care patients.   J. Crit. Care

44, 82-86.

Barr, J., Fraser, G.L., Puntillo, K., Ely, E.W., Gélinas, C., Dasta, J.F., et al., 2013. Clinical practice

guidelines for the management of pain, agitation, and delirium in adult patients in the

intensive care unit. Crit. Care Med. 41 (1), 263-306.

Bergeron, N., Dubois, M.J., Dumont, M., Dial, S., Skrobik, Y., 2001. Intensive Care Delirium

Screening Checklist: evaluation of a new screening tool. Intensive Care Med. 27

(5),859-864.

Boettger, S., Nuñez, D.G., Meyer, R., Richter, A., Fernandez, S.F., Rudiger, A., et al., 2017.

Delirium in the intensive care setting: A reevaluation of the validity of the CAM-ICU and

ICDSC versus the DSM-IV-TR in determining a diagnosis of delirium as part of the daily

clinical routine. Palliat. Support. Care 15 (6), 675-683.

Bui, L.N., Pham, V.P., Shirkey, B.A., Swan, T.J., 2017. Effect of delirium motoric subtypes on

administrative documentation of delirium in the surgical intensive care unit. J. Clin. Monit.

Comput. 31 (3), 631-640.

Chanques, G., Ely, E.W., Garnier, O., Perrigault, F., Eloi, A., Carr, J., et al., 2018. The 2014 updated

version of the Confusion Assessment Method for the Intensive Care Unit compared to the

5th version of the Diagnostic and Statistical Manual of Mental Disorders and other current

methods used by intensivists. Ann. Intensive Care 8 (1), 33.

Chuang, W.L., Lin, C.H., Hsu, W.C., Ting, Y.J., Lin, K.C., Ma, S.C., 2007. Evaluation of the

reliability and validity of the Chinese version of the confusion assessment method for the

intensive care unit. Hu Li Za Zhi 54 (4),45-52. [in Chinese]

Danzeng, Q.Z., Cui, N., Wang, H., Pan, W.J., Long, Y., Deji, Y.Z., et al., 2019. Translation and

validation of the Tibetan confusion assessment method for the intensive care unit. Chin.

Med. J. 132 (10), 1154.

De Jonghe, B., Constantin, J.M., Chanques, G., Capdevila, X., Lefrant, J.Y., Outin, H., et al., 2013.

Physical restraint in mechanically ventilated ICU patients: a survey of French practice.

Intensive Care Med. 39 (1), 31-37.

de Oliveira Fagundes, J.A., Tomasi, C.D., Giombelli, V.R., Alves, S.C., de Macedo, R.C., Topanotti,

M.F.L., et al., 2012. CAM-ICU and ICDSC agreement in medical and surgical ICU patients

is influenced by disease severity. PLoS One 7 (11),e51010.

Deeks, J.J., Petra M., Les I., 2005. The performance of tests of publication bias and other sample

size effects in systematic reviews of diagnostic test accuracy was assessed. J. Clin.

Epidemiol. 58 (9), 882-893.

Domenico, G.G., Piergentili, F., 2012. Cultural and linguistic validation of the Italian version of the

intensive care delirium screening checklist. Dimens. Crit. Care Nurs. 31 (4), 246-251.

Ely, E.W., Sharon, K.I., Gordon, R.B., Sharon, G., Joseph, F., Lisa, M., et al., 2001. Delirium in

mechanically ventilated patients: validity and reliability of the confusion assessment method

for the intensive care unit (CAM-ICU). JAMA 286 (21), 2703-2710.

Ely, E.W., Margolin, R., Francis, J., May, L., Truman, B., Dittus, R., et al., 2001. Evaluation of

delirium in critically ill patients: validation of the Confusion Assessment Method for the

Intensive Care Unit (CAM-ICU). Crit. Care Med. 29 (7), 1370-1379.

Fluss, R., Faraggi, D., Reiser, B., 2005. Estimation of the Youden Index and its associated cutoff

point. Biom. J. 47 (4), 458-472.

George, C., Nair, J.S., Ebenezer, J.A., Gangadharan, A., ChristuDas, A., Gnanaseelan, L.K., et al., 2011. Validation of the Intensive Care Delirium Screening Checklist in nonintubated intensive care unit patients in a resource-poor medical intensive care setting in South India. J. Crit. Care 26 (2), 138-143.

Gottlieb, S.S., Khatta, M., Friedmann, E., Einbinder, L., Katzen, S., Baker, B., et al., 2004. The influence of age, gender, and race on the prevalence of depression in heart failure patients. J. Am. Coll. Cardiol. 43 (9), 1542-1549.

Gusmao-Flores, D., Salluh, J.I., Chalhub, R.A., Quarantini, L.C., 2012. 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 16 (4), R115.

Gusmao-Flores, D., Salluh, J.I., Dal-Pizzol, F., Ritter, C., Tomasi, C.D., de Lima, M.A., et al., 2011. The validity and reliability of the Portuguese versions of three tools used to diagnose delirium in critically ill patients. Clin. 66 (11), 1917-1922.

Heo, E.Y., Lee, B.J., Hahm, B.J., Song, E.H., Lee, H.A., Yoo, C.G., et al., 2011. Translation and validation of the Korean confusion assessment method for the intensive care unit. BMC psychiatry 11 (1), 94.

Higgins, J.P.T, Thompson, S.G. 2002. Quantifying heterogeneity in a meta- analysis. Stat. Med. 21 (11), 1539-1558.

Jones, S., Carley, S., Harrison, M., 2003. An introduction to power and sample size estimation.

Emerg. Med. J. 20 (5), 453-458.

Jung, J.H., Lim, J.H., Kim, E.J., An, H.C., Kang, M.K., Lee, J., et al., 2013. The experience of

delirium care and clinical feasibility of the CAM-ICU in a Korean ICU. Clin. Nurs. Res. 22

(1), 95-111.

Karlicic, I.S., Stašević, M., Janković, S., Dejanović, S.D., Dutina, A., Grbić, I., 2016. Translation to

serbian, transcultural adaptation and validation of the Confusion Assessment Method for the

Intensive Care Unit (CAM-ICU). Acta Medica Int. 3 (2), 68-71.

Kim, M.Y., Kim, K., Hong, C.H., Lee, S.Y., Jung, Y.S., 2018. Sex differences in cardiovascular risk

factors for dementia. Biomol. Ther. 26 (6), 521-523.

Koga, Y., Tsuruta, R., Murata, H., Matsuo, K., Ito, T., Ely, E.W., et al., 2015. Reliability and validity

assessment of the Japanese version of the Confusion Assessment Method for the Intensive

Care Unit (CAM-ICU). Intensive Crit. Care Nurs. 31 (3), 165-170.

Kose, G., Bolu, A., Ozdemir, L., Acikel, C., Hatipolu, S., 2016. Reliability and Validity of the

Intensive Care Delirium Screening Checklist in Turkish. International Journal of Nursing

Knowledge 27 (2):119-124.

Krewulak, Karla D, Henry T Stelfox, Jeanna Parsons Leigh, E Wesley Ely, and Kirsten M Fiest.

2018. "Incidence and prevalence of delirium subtypes in an adult ICU: a systematic review

and meta-analysis." Critical care medicine 46 (12):2029-2035.

Lalkhen, Abdul Ghaaliq, and Anthony McCluskey. 2008. "Clinical tests: sensitivity and specificity."

Continuing Education in Anaesthesia Critical Care & Pain 8 (6):221-223.

Larsen, Laura Krone, Vibe G Frøkjær, Jette Stub Nielsen, Yoanna Skrobik, Yvonne Winkler, Kirsten Møller, Marian Petersen, and Ingrid Egerod. 2019. "Delirium assessment in neuro- critically ill patients: A validation study." Acta Anaesthesiologica Scandinavica 63 (3):352-359.

Lee, Hochang B, Simon C Mears, Paul B Rosenberg, Jeannie- Marie S Leoutsakos, Allan Gottschalk, and Frederick E Sieber. 2011. "Predisposing factors for postoperative delirium after hip fracture repair in individuals with and without dementia." Journal of the American Geriatrics Society 59 (12):2306-2313.

Leeflang, Mariska MG, Karel GM Moons, Johannes B Reitsma, and Aielko H Zwinderman. 2008. "Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions." Clinical chemistry 54 (4):729-737.

Lijmer, Jeroen G, Ben Willem Mol, Siem Heisterkamp, Gouke J Bonsel, Martin H Prins, Jan HP Van Der Meulen, and Patrick MM Bossuyt. 1999. "Empirical evidence of design-related bias in studies of diagnostic tests." JAMA 282 (11):1061-1066.

Lin, Shu-Min, Chien-Ying Liu, Chun-Hua Wang, Horng-Chyuan Lin, Chien-Da Huang, Pei-Yao Huang, Yueh-Fu Fang, Meng-Heng Shieh, and Han-Pin Kuo. 2004. "The impact of delirium on the survival of mechanically ventilated patients." Critical care medicine 32 (11):2254-2259.

Lipowski, Zbigniew J. 1989. "Delirium in the elderly patient." New England Journal of Medicine 320 (9):578-582.

Liptzin, Benjamin, and Sue E Levkoff. 1992. "An empirical study of delirium subtypes." The

British Journal of Psychiatry 161 (6):843-845.

Luetz, Alawi, Anja Heymann, Finn M Radtke, Chokri Chenitir, Ulrike Neuhaus, Irit Nachtigall,

Vera von Dossow, Susanne Marz, Verena Eggers, and Andreas Heinz. 2010. "Different

assessment tools for intensive care unit delirium: which score to use?"   Critical care

medicine 38 (2):409-418.

Mitasova, Adela, Milena Kostalova, Josef Bednarik, Radka Michalcakova, Tomas Kasparek, Petra

Balabanova, Ladislav Dusek, Stanislav Vohanka, and E Wesley Ely. 2012. "Poststroke

delirium incidence and outcomes: validation of the Confusion Assessment Method for the

Intensive Care Unit (CAM-ICU)."   Critical care medicine 40 (2):484-490.

Mitášová, Adéla, Josef Bednařík, Milena Košťálová, Radka Michalčáková, Martina Ježková, Tomáš

Kašpárek, Světlana Skutilová, Eva Straževská, Petra Šályová, and Veronika Šikolová. 2010.

"Standardizace české verze The Confusion Assessment Method for the Intensive Care Unit

(CAM-ICUcz)."   Cesk Slov Neurol (2010):73.

Neto, A Serpa, AP Nassar Júnior, Sérgio O Cardoso, José A Manetta, VG Pereira, Daniel C Espósito,

MC Damasceno, and Arjen J Slooter. 2012. "Delirium screening in critically ill patients: a

systematic review and meta-analysis."   Critical Care 16 (S1):P337.

Nishimura, Katsuji, Kanako Yokoyama, Noriko Yamauchi, Masako Koizumi, Nozomi Harasawa,

Taeko Yasuda, Chizuru Mimura, Hazuki Igita, Eriko Suzuki, and Yoko Uchiide. 2016.

"Sensitivity and specificity of the Confusion Assessment Method for the Intensive Care Unit

(CAM-ICU) and the Intensive Care Delirium Screening Checklist (ICDSC) for detecting

post-cardiac surgery delirium: a single-center study in Japan." Heart & Lung 45 (1):15-20.

Ouimet, Sébastien, Brian P Kavanagh, Stewart B Gottfried, and Yoanna Skrobik. 2007. "Incidence,

risk factors and consequences of ICU delirium." Intensive care medicine 33 (1):66-73.

Pandharipande, Pratik, Bryan A Cotton, Ayumi Shintani, Jennifer Thompson, Sean Costabile,

Brenda Truman Pun, Robert Dittus, and E Wesley Ely. 2007. "Motoric subtypes of delirium

in mechanically ventilated surgical and trauma intensive care unit patients." Intensive care

medicine 33 (10):1726-1731.

Pandharipande, Pratik, Bryan A Cotton, Ayumi Shintani, Jennifer Thompson, Brenda Truman Pun,

John A Morris Jr, Robert Dittus, and E Wesley Ely. 2008. "Prevalence and risk factors for

development of delirium in surgical and trauma ICU patients." The Journal of trauma 65

(1):34.

Pipanmekaporn, Tanyong, Nahathai Wongpakaran, Sirirat Mueankwan, Piyawat Dendumrongkul,

Kaweesak Chittawatanarat, Nantiya Khongpheng, and Nongnut Duangsoy. 2014. "Validity

and reliability of the Thai version of the Confusion Assessment Method for the Intensive

Care Unit (CAM-ICU)." Clinical interventions in aging 9:879.

Radtke, Finn M, Martin Franck, Stefan Oppermann, Alawi Lütz, Matthes Seeling, Anja Heymann,

Robin Kleinwächter, Felix Kork, Yoanna Skrobik, and Claudia D Spies. 2009. "The

Intensive Care Delirium Screening Checklist (ICDSC)--translation and validation of

intensive care delirium checklist in accordance with guidelines." Anasthesiologie,

Intensivmedizin, Notfallmedizin, Schmerztherapie: AINS 44 (2):80-86.

Rose, Louise, Lisa Burry, Ranjeeta Mallick, Elena Luk, Deborah Cook, Dean Fergusson, Peter

Dodek, Karen Burns, John Granton, and Niall Ferguson. 2016. "Prevalence, risk factors, and

outcomes associated with physical restraint use in mechanically ventilated adults." Journal

of critical care 31 (1):31-35.

Ruitenberg, Annemieke, Alewijn Ott, John C van Swieten, Albert Hofman, and Monique MB

Breteler. 2001. "Incidence of dementia: does gender make a difference?" Neurobiology of

aging 22 (4):575-580.

Salluh, Jorge IF, Han Wang, Eric B Schneider, Neeraja Nagaraja, Gayane Yenokyan, Abdulla

Damluji, Rodrigo B Serafim, and Robert D Stevens. 2015. "Outcome of delirium in

critically ill patients: systematic review and meta-analysis." bmj 350.

Selim, Abeer, Nahed Kandeel, Mohamed Elokl, Mohamed Shawky Khater, Ashraf Nabil Saleh,

Rami Bustami, and E Wesley Ely. 2018. "The validity and reliability of the Arabic version of

the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU): A prospective

cohort study." International journal of nursing studies 80:83-89.

Tobar, E, C Romero, T Galleguillos, P Fuentes, R Cornejo, MT Lira, L de la Barrera, JE Sanchez, F

Bozan, and G Bugedo. 2010. "Confusion Assessment Method for diagnosing delirium in

ICU patients (CAM-ICU): cultural adaptation and validation of the Spanish version."

Medicina intensiva 34 (1):4-13.

Toro, AC, LM Escobar, JG Franco, JL Diaz-Gomez, JF Munoz, F Molina, J Bejarano, D Yepes, E

Navarro, and A Garcia. 2009. "Spanish version of the CAM-ICU (Confusion Assessment

Method for the Intensive Care Unit). Pilot study of validation."   Medicina intensiva 34

(1):14-21.

Van de Velde, Sarah, Piet Bracke, and Katia Levecque. 2010. "Gender differences in depression in

23 European countries. Cross-national variation in the gender gap in depression."   Social

science & medicine 71 (2):305-313.

Van den Boogaard, Mark, Lisette Schoonhoven, Johannes G Van der Hoeven, Theo Van Achterberg,

and Peter Pickkers. 2012. "Incidence and short-term consequences of delirium in critically

ill patients: a prospective observational cohort study."   International journal of nursing

studies 49 (7):775-783.

van Eijk, Maarten M, Mark van den Boogaard, Rob J van Marum, Paul Benner, Piet Eikelenboom,

Marina L Honing, Ben van der Hoven, Janneke Horn, Gerbrand J Izaks, and Annette Kalf.

2011. "Routine use of the confusion assessment method for the intensive care unit: a

multicenter study."   American journal of respiratory and critical care medicine 184

(3):340-344.

Van Eijk, Maarten MJ, Rob J Van Marum, Ine AM Klijn, Nelleke De Wit, Jozef Kesecioglu, and

Arjen JC Slooter. 2009. "Comparison of delirium assessment tools in a mixed intensive care

unit."   Critical care medicine 37 (6):1881-1885.

Vreeswijk, R, A Toornvliet, MLH Honing, K Bakker, T De Man, G Daas, PE Spronk, JFM de

Jonghe, and KJ Kalisvaart. 2009. "Validation of the Dutch version of the Confusion

Assessment Method (CAM-ICU) for delirium screening in the Intensive Care Unit."

Netherlands Journal of Critical Care 13 (2):73-78.

Walter, SD. 2002. "Properties of the summary receiver operating characteristic (SROC) curve for

diagnostic test data." Statistics in medicine 21 (9):1237-1256.

Wang, Chunli, Ying Wu, Peng Yue, E Wesley Ely, Jie Huang, Xin Yang, and Yisi Liu. 2013.

"Delirium assessment using Confusion Assessment Method for the Intensive Care Unit in

Chinese critically ill patients." Journal of Critical Care 28 (3):223-229.

Whiting, Penny F, Anne WS Rutjes, Marie E Westwood, Susan Mallett, Jonathan J Deeks, Johannes

B Reitsma, Mariska MG Leeflang, Jonathan AC Sterne, and Patrick MM Bossuyt. 2011.

"QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies."

Annals of internal medicine 155 (8):529-536.

Whiting, Penny F, Anne WS Rutjes, Marie E Westwood, Susan Mallett, and QUADAS-2 Steering

Group. 2013. "A systematic review classifies sources of bias and variation in diagnostic test

accuracy studies." Journal of clinical epidemiology 66 (10):1093-1104.

Young, John, Lakshmi Murthy, Maggie Westby, Anayo Akunne, and Rachel O'Mahony. 2010.

"Diagnosis, prevention, and management of delirium: summary of NICE guidance." Bmj

341.

Zaal, Irene J, John W Devlin, Linda M Peelen, and Arjen JC Slooter. 2015. "A systematic review of

risk factors for delirium in the ICU." Critical care medicine 43 (1):40-47.

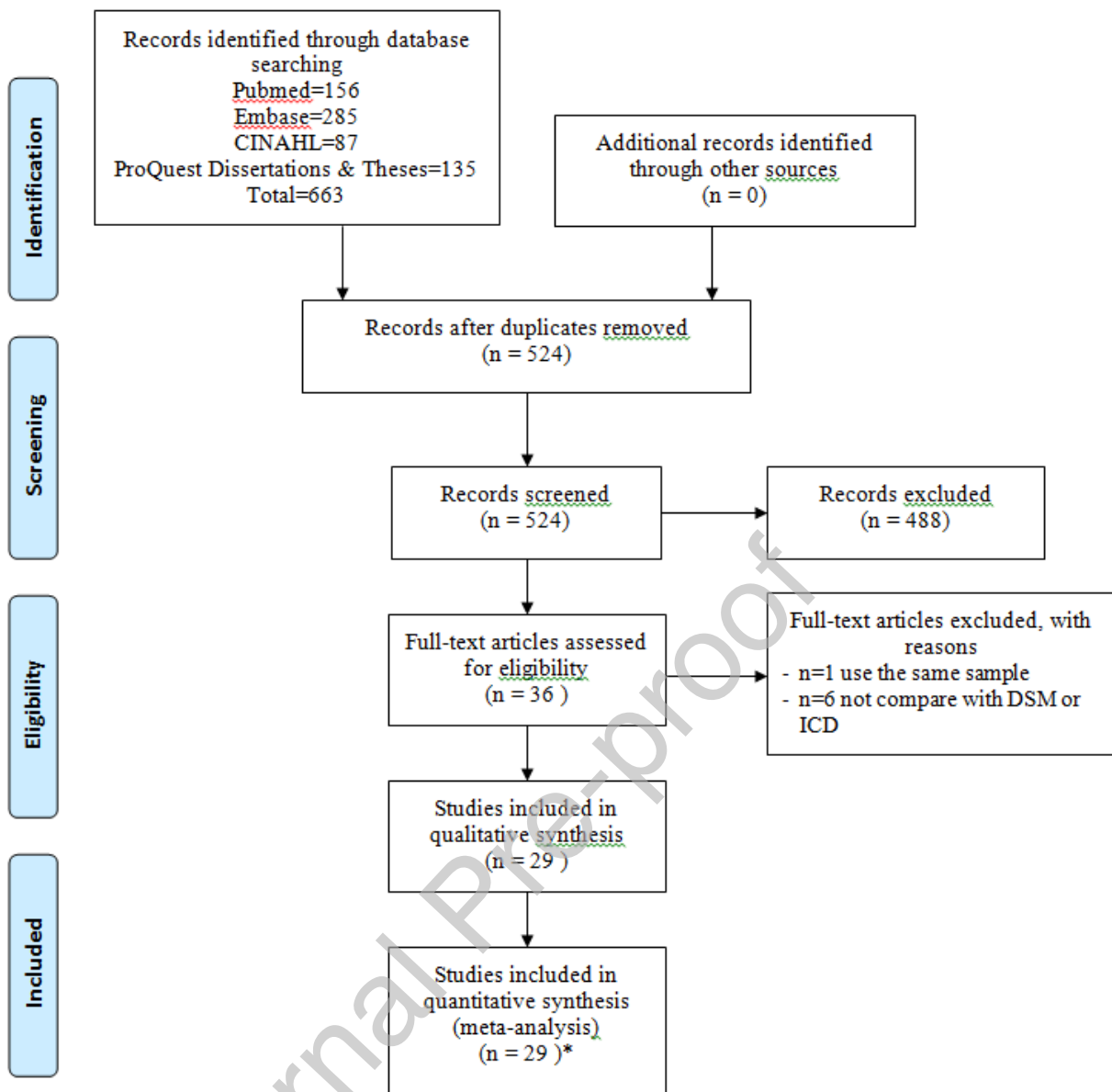**Figure 1** Preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2009 flow diagram for Confusion Assessment Method for the Intensive Care Unit

*One study separate participants into two group. Therefore, 30 effect sizes entered this analysis.
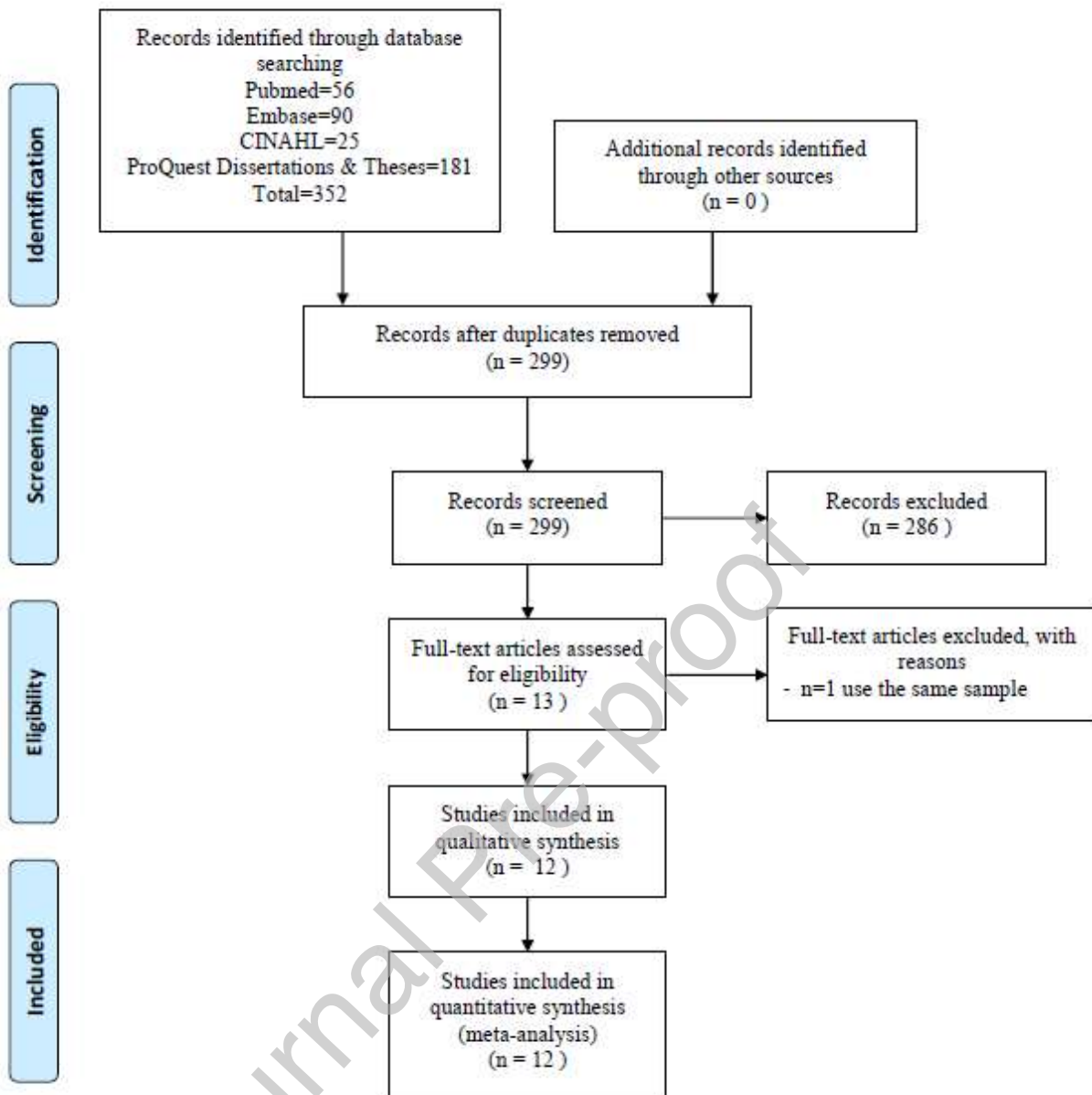
Figure 2 Preferred reporting items for systematic reviews and meta-analyses (PRISMA) 2009 flow diagram for Intensive Care Delirium Screening Checklist

Table 1 Descriptions of the CAM-ICU and ICDSC

Instruments	CAM-ICU				ICDSC		
Origin population	Adult medical ICU patients. Patients with a history of severe dementia, psychosis, or neurologic disease were excluded				Adult medical and surgical ICU patients. Patients admitted with a diagnosis of delirium were excluded.		
Features/ Items	1. Acute changes or fluctuation in mental status 2. Inattention 3. Disorganized thinking 4. Altered level of consciousness				1. Altered level of consciousness 2. Inattention 3. Disorientation 4. Psychosis	5. Altered psychomotor activity 6. Inappropriate speech/ mood 7. Sleep disturbance 8. Symptom fluctuation	
Reliability	Interrater reliability: 89 to 97%				Cronbach's alpha: 0.71 to 0.79 Interrater reliability: >94%		
Time needed	2 -3 mins in average, but may be up to 10 mins when users are unfamiliar with the content				2 mins in average		
Languages available	Arabic Czech Danish Dutch Egyptian English French Greek	German Hebrew Hindi India Italian Japanese Korean Marathi India Norwegian	Persian Polish Portuguese Romanian Russian Serbian Simplified Chinese	Spanish Swedish Turkish Thai Traditional Chinese Tunisian Zulu	Dutch English French German Italian	Japanese Persian Portuguese Swedish Turkish	
Scoring systems/ cutoff point	Check every feature is presented or not. Positive for delirium if Features 1 and 2 and *either* Feature 3 or Feature 4 are identified.				Give one point for each symptom manifested, zero point if symptom dose not manifest. A total score ≥4 indicates the presence of delirium.		

1

| Assessment approach | Based on the observation and interaction at one time-point. Need additional test (e.g. attention screening examination) | Based on the continuous observation of routine care. |

Table 2 Summary estimates of sensitivity and specificity

Scales	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)
CAM-ICU	0.84 (0.77 to 0.88)	0.95 (0.91 to 0.97)
ICDSC	0.83 (0.74 to 0.90)	0.87 (0.78 to 0.93)
F/P value CAM-ICU vs. ICDSC	0.03/0.85	4.32/0.04

Note: F/P value indicates the F value and P value of comparing the CAM-ICU and ICDSC

2

Table 3 Metaregression and moderator analysis of CAM-ICU (n=30)

Variables	n	Sensitivity		Specificity	
		Point estimates	P value	Point estimates	P value
Age	30	0.03	0.27	-0.03	0.47
Sample size	30	-0.0007	0.46	-0.001	0.24
Female	30	0.44	0.81	0.79	0.75
APACHE II	16	-0.03	0.60	-0.08	0.16
Hypoactive-delirium	10	-1.43	0.42	3.50	0.04
SICU					
Yes	7	0.75	0.10	0.95	0.72
No	23	0.86		0.94	
MICU					
Yes	8	0.93	0.01	0.93	0.62
No	22	0.79		0.95	
Mixed-ICU					
Yes	15	0.82	0.49	0.95	0.87
No	15	0.85		0.94	
SICU v.s. MICU					
SICU	7	0.75	0.02	0.95	0.71
MICU	8	0.94		0.94	
MV use at enrollment					
Yes	5	0.95	0.02	0.99	0.04
Not all	23	0.81		0.94	
Country-region					
Asia	11	0.87	0.35	0.93	0.34
Non-Asia	19	0.82		0.96	

Abbreviation: APACHE II = Acute Physiology and Chronic Health Evaluation II; MV= mechanical ventilation;

SICU = surgical intensive care unit; MICU = medical intensive care unit

Table 4 Metaregression and moderator analysis of ICDSC (n=12)

Variables	n	Sensitivity		Specificity	
		Point estimates (95%CI)	P value	Point estimates	P value
Age	12	-0.008	0.86	-0.04	0.38
Sample size	12	-0.003	0.34	0.007	0.04
Female	12	7.66	0.001	-8.77	0.003
APACHE II	7	-0.12	0.30	-0.08	0.50
Hypoactive-delirium	5	3.02	0.07	1.01	0.65
SICU					
Yes	3	0.84	0.92	0.91	0.48
No	9	0.83		0.86	
MICU					
Yes	3	0.87	0.58	0.73	0.07
No	9	0.82		0.90	
Mix-ICU					
Yes	6	0.80	0.49	0.90	0.36
No	6	0.86		0.84	
Country-region					
Asia	3	0.89	0.29	0.90	0.53
Non-Asia	9	0.81		0.86	

Abbreviation: APACHE II = Acute Physiology and Chronic Health Evaluation II; MV= mechanical ventilation;

SICU = surgical intensive care unit; MICU = medical intensive care unit