Accuracy of delirium assessments in critically ill children: A prospective, observational study during routine care

Rebecca S. Paterson, PhD[1,2,3], Justin A. Kenardy, PhD[1,3], Belinda L. Dow, PhD[1,3], Alexandra C. De Young, PhD[3,4], Kylie Pearson, RN[2,5], Leanne M. Aitken, PhD[6,7,8], Debbie A. Long, PhD[2,5,6]

Affiliations: [1]School of Psychology, University of Queensland, Brisbane, AU; [2]Paediatric Critical Care Research Group, Centre for Children's Health Research, Brisbane, AU; [3]Children's Health Research Centre, University of Queensland, Brisbane, AU; [4]Centre for Children's Burn and Trauma Research, Centre for Children's Health Research, Brisbane, AU; [5]Paediatric Intensive Care Unit, Queensland Children's Hospital, Brisbane, AU; [6]Menzies Health Institute Queensland, Griffith University, AU; [7]School of Health Sciences, City, University of London, London, UK; [8]Princess Alexandra Hospital, Brisbane, AU

This study was conducted at the Queensland Children's Hospital, Brisbane, Queensland, Australia.

Address correspondence to: Rebecca Paterson, Level 7 Centre for Children's Health Research Building, 62 Graham Street, South Brisbane, QLD, 4101, Australia, r.paterson@uq.edu.au

Funding Source: All phases of this study were supported by the University of Queensland, Australia, Griffith University, Australia, and Children's Health Queensland, Australia. Rebecca Paterson is supported by an Australian Government Research Training Program (RTP) Scholarship (Full-time base RTP Stipend Scholarship); and the Children's Hospital Foundation (Children's Hospital Foundation PhD Scholarship Top-Up Grant RPCPHD0032017). Debbie Long was awarded a Queensland Health Nursing and Midwifery Research Fellowship as financial support for this project.

Financial Disclosure: The authors have indicated they have no financial relationships relevant to this article to disclose.

Conflict of Interest: The authors have indicated they have no potential conflicts of interest to disclose.

Acknowledgements: The authors gratefully acknowledge the contributions of Tara Williams, Jessica Schults, Kerry Johnson and Rebecca Doyle for their assistance as clinical nurse researchers; and Drs Ronan McKenna and Tavey Dorofaeff for their active roles as medical champions for the study. The authors additionally thank Dr. Anne Bernard for her review of the statistical analyses conducted in this manuscript, and Nicholas Gailer for his contribution to preparing raw medical data for analysis. The authors have no conflicts of interest to declare.

ABSTRACT

Objectives: To explore the accuracy of the Cornell Assessment for Pediatric Delirium (CAP-D), and the Pediatric- and Preschool- versions of the Confusion Assessment Method for the Intensive Care Unit (pCAM-ICU and psCAM-ICU) when implemented into routine care as a screening tool; and, assess patient characteristics and clinical variables which may affect their validity.
Design: A prospective observational study.
Setting: A 36-bed**,** mixed-PICU at an Australian tertiary hospital.
Patients: Critically ill children developmentally aged 6 months to 17 years with a PICU length of stay ≥ 18 hours.
Interventions: None.
Measurements and Main Results: Patients were screened for delirium by their bedside nurse (CAP-D and pCAM-ICU/psCAM-ICU) once daily, for up to five days. Delirium status identified using screening instruments was compared to delirium diagnosis using the diagnostic criteria for Delirium (DSM-5). In this sample, the CAP-D retained its high sensitivity (91.3%) and good specificity (75.2%), whereas the psCAM-ICU and pCAM-ICU had moderate sensitivity (58.8% and 75.0%, respectively) and excellent specificity (89.8% and 84.9%, respectively). There was moderate agreement between the CAP-D and the psCAM-ICU, $\kappa = .52$, $p < .001$; and good agreement between the CAP-D and the pCAM-ICU, $\kappa = .80$, $p < .01$.
Conclusions: Although the CAP-D, psCAM-ICU and pCAM-ICU all appear promising in their validation studies, when implemented into routine care, their performance can be variable. The CAP-D performed well in routine clinical practice, but follow-up diagnosis is required to confirm delirium. The ps/pCAM-ICU both provide valuable, objective assessments of delirium in critically ill children, however further evaluation of their implementation into routine clinical practice is needed.

Keywords: *Delirium; Intensive Care Units, Paediatric; Screening; Psychometrics*

INTRODUCTION

Delirium is a serious neuropsychiatric complication of critical illness characterised by acute and fluctuating disturbances in attention and awareness, and changes in baseline cognition. [1] Paediatric delirium is common in children admitted to the Paediatric Intensive Care Unit (PICU), occurring in 25% of patients, [2] and has been associated with increased health-care expenditure, length of hospitalisation, duration of mechanical ventilation and risk of mortality. [3-5] Emerging research in adults has linked the early prevention, identification and management of delirium with reduced burden on nursing workload. [6] Diagnosing delirium in children is challenging, however, as it requires clinicians to consider multiple factors (e.g., developmental regression, pain, anxiety, depression, iatrogenic withdrawal and psychosis) that may have a similar or comorbid presentation. [7, 8] The variation in developmental age and cognitive and language skills in children admitted to PICU, combined with the speculation that there is no lower age limit to delirium, further complicates paediatric delirium detection. Additionally, the availability of personnel, patient needs, and the feasibility of implementing time- and resource-intensive evaluation of symptoms against the Statistical Manual of Mental Disorders (DSM-5) diagnostic criteria, adds to the difficulty of delirium diagnosis. [9] As such, clinical decision-making relies on the use of delirium screening instruments that address all of the aforementioned challenges.

Rapid and reliable screening of paediatric delirium is important in aiding accurate identification, [10] ongoing assessment and re-assessment (including confirming delirium status with follow-up diagnosis), [8] and prevention and management strategies (e.g., implementing multi-component non-pharmacological interventions). [11] Accurate screening instruments are also necessary for better understanding the prevalence, risk factors, and outcomes of paediatric delirium. [12] In recognition that bedside nurses are uniquely placed to detect symptoms of delirium and instigate early prevention, identification and management strategies, [13, 14] there has been an increase in the availability screening instruments that have been developed and validated for use at the PICU bedside with minimal training, including: the Cornell Assessment for Pediatric Delirium (CAP-D), [15-17]

the Pediatric Confusion Assessment Method for the Intensive Care (pCAM-ICU), [18] and the Preschool Confusion Assessment Method for the Intensive Care Unit (psCAM-ICU) . [19] While all three instruments are reliable and rapid tools for detecting delirium, they have primarily been validated when employed by trained users; the CAP-D was initially validated with the attending paediatric intensivist and chief resident, [15] while the pCAM-ICU and psCAM-ICU were validated by a study team of trained paediatric anaesthesiologists, intensivists, nurse practitioners and registered nurses. [18, 19] Only the CAP-D has undergone validity testing with bedside nurses undertaking the screening assessment, however, this have yet to be replicated outside the developing institution. [17] To date, this has facilitated standardisation of these screening instruments but does not ensure accuracy when used for routine bedside screening.

Additionally, few studies have accounted for patient and clinical variables that may impact the accuracy of delirium screening instruments. Factors such as receiving sedation, mechanical ventilation, age and gender have been shown to impact the accuracy of the pCAM-ICU and its modified severity scale (sspCAM-ICU), [20] while others have also speculated that developmental delay, younger age and motoric subtype can all complicate the accuracy of delirium detection. [7, 9, 17] Therefore, this study aimed to i) evaluate the accuracy of each screening instrument, when implemented into routine care, against a reference standard assessment of delirium (clinical diagnosis using the DSM-5 criteria); [1] and, ii) assess patient characteristics and clinical variables which may affect their validity.

MATERIAL AND METHODS

Study Design and Participants

This prospective, observational study was conducted from November 2015 to April 2017 in a 36 bed, mixed medical and surgical PICU at an Australian tertiary hospital. The study was undertaken 12 months after commencing nursing education and the adoption of routine delirium screening using the ps/pCAM-ICU and CAP-D. This study was granted ethics approval by the Children's Health

Queensland Human Research Ethics Committee (HREC/13/QRCH/105/AM4) and the University of Queensland Human Research Ethics Committee (2015001425).

Inclusion Criteria

All patients developmentally aged 6 months to 17 years were eligible for inclusion. As the patient's age impacted the decision regarding which paediatric CAM-ICU used, developmental age was based on parent or bedside nurse report regarding the patient's capacity at baseline to complete the psCAM-ICU (6 months to 5 years) or the pCAM-ICU (≥6 years). Therefore, when (according to parent or bedside nurse report) a patient's developmental age was less than their chronical age, they were categorised as "developmentally delayed" but included in the study.

Exclusion Criteria

Patients were excluded if they had significantly impaired vision or hearing which prevented completion of the ps/pCAM-ICU; were comatose or deeply sedated (State Behavioral Scale, SBS, score <-1); had significant intellectual impairment or developmental delay (developmentally <6 months); were receiving end-of-life care; were admitted <18 hours; were under the care of the Department of Child Safety; or their parents were non-English speaking.

Recruitment, Consent and Study Procedures

On consecutive weekdays for 18 months, patients were screened daily for eligibility to participate in the study. Assessment of delirium could only be performed in patients with a SBS score ≥-1, [21] indicating the patient was sufficiently aware for delirium assessment. If a patient's SBS was too low for assessment (-2, -3), they were reviewed daily to determine their eligibility for inclusion in the study. In order to include patients who transited from deep sedation to awake states, all patients were evaluated for inclusion, even if they met exclusion criteria the previous day (e.g., due to being deeply sedated). Parents or guardians of eligible patients were approached to participate in the study. When informed consent was obtained, paired assessment commenced. Enrolled patients underwent paired assessments (a nurse evaluation using the CAP-D and the ps/pCAM-ICU and a

reference standard evaluation using the DSM-5 diagnostic criteria) at 12:00pm to align with routine screening. Consistent with Smith and colleagues, [18] paired assessments occurred within three hours. All assessors were blinded to the findings of each other's evaluations. Eligible patients were re-evaluated for delirium over consecutive days for the remainder of their PICU admission. To reduce potential response bias, and consistent with other studies, [18] assessments were capped at a maximum of five paired assessments per patient.

Nursing Education

Delirium screening education was provided through a mandatory online in-service that included a video series providing paediatric delirium education (e.g., symptoms, risk factors, management strategies), specific demonstrations of administering the CAP-D, pCAM-ICU and psCAM-ICU (including scenarios where assessment may be challenging, e.g., non-compliant, low arousal), and a test of their knowledge at the end of the online training. Overall, 83.4% of nursing staff (*n* = 191) completed the education series, with additional bedside training undertaken by a member of the research team. A convenience sample of ten paired CAP-D and p/psCAM-ICU assessments between the bedside nurse and research nurse providing the education session indicated acceptable interrater reliability (κ =0.62-1.00). [22] Further education was provided through bedside education maintenance and mandatory competency days.

Instruments

Index tests

To evaluate the accuracy of screening in routine clinical practice, each assessment was undertaken by the patient's bedside nurse. For children developmentally <6 years old, the nurse screened patients using the psCAM-ICU and the CAP-D, while children developmentally ≥6 years old, were screened for delirium using the pCAM-ICU and CAP-D.

The CAP-D is modified from the Pediatric Assessment of Emergence Delirium (PAED) scale to capture hypoactive, hyperactive and mixed motoric symptoms. [15, 23] The scale includes eight items which

assessors score based on the frequency of symptoms over the course of a shift ('Never', 'Rarely', 'Sometimes', 'Often', and 'Always'). For children ≤2 years old, the instrument is accompanied by a validated rubric of developmental anchor points. [16] Patients who score a total ≥9 are considered delirium positive.

The structure of the ps/pCAM-ICU is modified from the CAM-ICU to enable developmentally appropriate evaluation, [24, 25] and comprises four features: (A) acute onset or fluctuating course of altered mental status, (B) inattention, (C) altered level of consciousness, and (D) disorganised thinking. For very young children (psCAM-ICU), the assessment for inattention is completed using developmentally appropriate picture cards and a mirror. Similarly, picture cards can be used as an attentional test in older children, however assessment using the pCAM-ICU commonly incorporates a modified Vigilance 'A' test, [26] and evaluates disorganised thinking by asking children to respond to questions and commands, rather than looking for evidence of subtle neuropsychiatric symptoms (e.g., non-purposeful movement, inconsolability) and dysregulated systems and behaviour (psCAM-ICU). Both the psCAM-ICU and pCAM-ICU are considered delirium positive if the patient demonstrates disturbances in features A and B, and feature C and/or D.

Overall, the instruments either require (psCAM-ICU and pCAM-ICU) or do not require (CAP-D) patient interaction; are point-in-time (psCAM-ICU and pCAM-ICU) or require reflection of observations over several hours (CAP-D); all take <2 minutes to complete. The psychometric properties of the CAP-D, pCAM-ICU and psCAM-ICU are described in Table 1, with each instrument demonstrating acceptable sensitivity and specificity in previous validation studies. [17-19]

Reference Standard Assessment

The reference standard assessments were based on a clinical diagnosis by a graduate-level psychologist (author RP) with experience in PICU and hospitalised children more broadly, in accordance with the DSM-5 diagnostic criteria for delirium. [1] The reference standard assessor underwent training with a Clinical Psychologist and Neuropsychologist, and consulted with a

delirium assessment expert (Dr Gabrielle Silver), prior to commencing assessment. Information was collected from parents or guardians, bedside nursing staff, medical staff, and medical charts, and documented on a checklist of the five diagnostic criterion and associated features. Diagnosis was supported by a diagnostic aide (to be published), similar to the Vanderbilt Assessment for Delirium in Infants and Children (VADIC), [27] which included age- and developmentally-specific examples of paediatric disturbances in attention and awareness, cognition, sleep-wake cycle, emotional responsivity, and psychomotor activity. A clinical decision was made based on observations and collateral information, with consideration of alternative causes for changes from baseline attention and awareness. Reference standard assessments took approximately 10-30 minutes to complete. When an assessment was recorded as 'delirium present', a delirium subtype (hypoactive, hyperactive or mixed) was ascribed. There is no consensus on the definition of subsyndromal delirium (SSD), [28] so if an assessment was recorded as 'delirium absent', a classification of 'SSD present' was made if Criterion B (a disturbance "*which develops over a short period of time*", and which "*represents a change from baseline attention and awareness*") [1], and at least two other criteria were met.

Statistical Analysis

Statistical analysis for evaluating measurement properties was guided by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist [29] and is reported according to the Standards for Reporting Diagnostic accuracy studies (STARD) 2015 checklist. [30] Demographic characteristics are summarized using descriptive statistics. Statistical analyses were performed using the IBM SPSS Version 25 for Mac (IBM Statistics, Chicago, Illinois). [31]

Power calculations

In order to calculate the required sample size to determine the validity of each screening instrument, we adopted a H_0 of 50.0% sensitivity and specificity for each screening instrument. As the instruments were being evaluated as screening, not diagnostic, instruments, we prioritised

greater sensitivity over specificity when determining the minimum target sensitivity (H_a = 80.0) and specificity (H_a = 70.0) required, with α = .05, 1-β = .80. Based on the reported delirium incidence of 12.3% to 44.0% in previous validation studies, [15-19] we conservatively anticipated a 20% prevalence of delirium in our population. Therefore to achieve target sensitivity, we estimated a total sample size of 100 assessments were required. [32]

Predictive validity

The psychometric properties of the three screening instruments were determined by their predictive ability to classify patients into 'delirium present' or 'delirium absent' categories compared to the DSM-5 reference standard diagnosis. The number of true positives, false positives, true negatives and false negatives were calculated using 2 x 2 tables. Predictive validity was evaluated by calculating sensitivity (true positive/(true positive + false negative)); specificity (true negative/(false positive + true negative)); positive predictive value (PPV) (true positive/(true positive + false positive)); and negative predicative value (NPV) (true negative/(true negative + false negative)). Values greater than 0.70 (moderate) were pre-specified as acceptable. Accuracy of the three screening instruments was derived from the area under the curve (AUC) using Receiver Operator Characteristic (ROC) Curves, with scores $\geq$0.70 considered acceptable. [33]

Construct validity

Convergent validity was established by comparing delirium state (present/absent) between nurse assessments of the CAP-D and the ps/pCAM-ICU using Cohen's kappa statistics, as it can account for the variation in dichotomous outcomes due to chance alone.[34] Based on the assumption that the CAP-D and the ps/pCAM-ICU both measure delirium in critically ill children, it was hypothesised that the instruments would demonstrate κ > 0.65 (moderate) agreement. [22] Divergent validity by comparing nurse assessments of delirium using the CAP-D and the ps/pCAM-ICU to the child's concurrent Multidimensional Assessment of Pain Scale (MAPS) score, [35] using Cohen's kappa. It was assumed that the instruments measure distinct constructs (delirium versus pain), and hypothesised

that there would be $\kappa < 0.40$ (low) agreement. [22]

Covariate effects on screening accuracy

Disagreement between tests was determined by calculating the discordance between the index text outcome and the reference standard outcome ([index text outcome – reference test outcome]; disagreement = |1|, agreement = 0) and agreement evaluated using Cohen's kappa. Univariate and multivariate analysis using logistic regression was undertaken to identify patient- and treatment-related variables that contributed to the likelihood of disagreement. Based on a review of the literature, these included patient age, gender, developmental delay, illness severity (Paediatric index of mortality 3; PIM-3), mechanical ventilation status (mechanical ventilation required versus not required), sedation status (receiving sedation versus not receiving sedation), subsyndromal delirium (present versus absent), delirium subtype (hypoactive versus mixed/hyperactive subtypes), and time between paired assessments. Variables that were significant at the univariate level ($p < .10$) were included in the multivariate analysis. To control for responder bias in patients with prolonged admissions, the multivariate analyses included PICU length of stay.

RESULTS

Patient Characteristics

In total, 119 patients were enrolled and 186 reference standard assessments were conducted. Five patient assessments were excluded due to >180 minutes separating the nursing assessment from the reference standard, and 18 reference assessments were missing corresponding index tests (see Figure 1 for participant flow diagram). [30] Patient characteristics are described in Table 2. The majority of patients were female (*n* = 64; 54.0%) and had a primary diagnosis of 'Surgical, excluding cardiac' (*n* = 35, 29.4%). The median length of PICU stay was 2.5 (IQR = 1.8–6.6) days. The first paired assessment occurred on PICU admission day 2.0 (IQR = 1.0–5.0 days) and participants had a median of 1.0 (IQR = 1.0-2.0) study assessments. On average, the reference and index tests occurred within 30:33 (*Mdn* = 31:00, IQR = 15:00-60:00) minutes of each other. According to DSM-5 criteria, 22.6% of

all reference assessments (*n* = 42) were 'delirium present' (hypoactive *n* = 20, 47.6%; hyperactive *n* = 10, 23.8%; and mixed *n* = 12, 28.6%), and 38.3% assessments (*n* = 72) were 'SSD present'.

Criterion validity

The criterion validity of the screening instruments is summarised in Table 3 and included 176 CAP-D assessments (93.1%), 125 psCAM-ICU assessments (66.1%) and 38 pCAM-ICU assessments (21.6%). Positive CAP-D results show delirium was correctly in 21/23 assessments (sensitivity = 91.3%) and negative results show absence of delirium was correctly identified 115/117 assessments (specificity = 75.2). The PPV and NPV of the CAP-D were 35.6% and 98.3% respectively. Delirium positive scores on the psCAM-ICU indicated that delirium was correctly identified in 10/17 assessments, resulting in a sensitivity of 58.8% (NPV = 93.3%). Negative results demonstrated that absence of delirium was correctly identified by the psCAM-ICU in 97/104 patients, resulting in a specificity of 89.8% (PPV = 47.6%). Positive assessments using the pCAM-ICU showed that delirium was correctly detected in 3/4 assessments, yielding a sensitivity of 75.0% (PPV = 37.5%); negative assessments indicated that the absence of delirium was correctly identified 28/33 assessments, resulting in a specificity of 84.9% (NPV = 96.6%). The ROC Curve analysis is summarised in Figure 2 and shows an AUC of 0.74 for the psCAM-ICU ($p = .001$), 0.80 for the pCAM-ICU ($p = .051$), and 0.83 for the CAP-D ($p < .001$), when compared to the DSM-5 diagnostic criteria for delirium (Table 3).

Construct validity

Convergent validity was evaluated using paired assessments of the CAP-D and psCAM-ICU (*n* = 125) and the CAP-D and pCAM-ICU (*n* = 38). There was moderate to good agreement between the CAP-D and psCAM-ICU assessments, $\kappa = 0.52$ (95% CI, 0.37 to 0.67), $p < .001$; and the CAP-D and pCAM-ICU assessments, $\kappa = 0.80$ (95% CI, 0.58 to 1.0), $p < .001$, respectively.

Divergent validity was evaluated using paired assessments of the CAP-D and MAPS (*n* = 166), the psCAM-ICU and MAPS (*n* = 94) and the pCAM-ICU and MAPS (*n* = 35). There was very low agreement for both the CAP-D, $\kappa = 0.27$ (95% CI, -0.13 to 0.42), $p < .001$, and the psCAM-ICU, $\kappa = 0.23$ (95% CI, -

0.15 to 0.40), p = .009, when compared to the MAPS. For the pCAM-ICU, there was moderate agreement with the MAPS, κ = 0.41 (95% CI, -0.29 to 0.78), p = .001

Covariate effects on screening accuracy

A total of 40 CAP-D (22.7%; κ = .40 [95% CI, 0.26 to 0.54], p < .001), 18 psCAM-ICU (10.2%; κ = .44 [95% CI, 0.22 to 0.66], p < .001) and 6 pCAM-ICU (16.2%; κ = .42 [95% CI, 0.05 to 0.79], p < .01) assessments were discordant to the reference standard assessment. At the univariate level, CAP-D disagreement was associated with younger age (p = .015), receiving any sedation (p = .001), and SSD (p < .001). Disagreement with the psCAM-ICU was marginally associated with younger age (p = .052) and receiving sedation (p = .031). Factors such as time between paired assessments, mechanical ventilation status, developmental delay, and motoric subtype were not associated with likelihood of disagreement for the CAP-D and psCAM-ICU (p > .10). The number of discordant assessments for the pCAM-ICU were too small to determine any significant variables which may contribute to inaccuracy at the univariate level.

The multivariate model exploring disagreement between the reference standard and the CAP-D is summarised in Table 4. The regression model ($\chi^2(4) = 38.0$, p < .001) indicated an increased likelihood of an inaccurate result when patients were younger (OR = 0.9, p = .013), receiving any sedation (OR = 4.1, p = .006) or met SSD criteria (OR = 4.4, p = .001). The number of discordant assessments for the psCAM-ICU were too small to determine any significant variables which may contribute to inaccuracy at the multivariate level.

DISCUSSION

To the best of our knowledge, this current study is the first to evaluate the accuracy of paediatric delirium screening instruments implemented into routine care. This study found that the CAP-D was highly sensitive, with results similar to its validation studies, [15, 16] and had reasonable specificity. The pCAM-ICU had sensitivity scores slightly lower than previously reported, but maintained excellent specificity. [18, 20] Conversely, the psCAM-ICU was not as sensitive as previously reported [19], despite

excellent specificity.

Although the psCAM-ICU was not as sensitive as expected, it is likely that this is a reflection of lower than predicted delirium prevalence, [32] especially compared to its validation study. [19] The structure of the psCAM-ICU may have additionally contributed to its reduced sensitivity in this setting. In particular, the first step of the psCAM-ICU is to determine an "*acute change or fluctuating course of mental status*", which if absent, results in the assessment being discontinued and the patient scored as 'delirium negative'. Determining changes in mental status in critically ill infants and young children is difficult due to the large variations in language and cognitive development in this cohort, as highlighted by the univariate association between younger age and reduced psCAM-ICU accuracy. This association is only tenuous and does not account for co-occurring factors, however future implementation of the psCAM-ICU into routine clinical practice may benefit from the accompaniment of developmentally appropriate guidance for detecting altered mental status in this challenging cohort.

Overall, the corresponding PPV for each of the screening instruments was low compared to other studies. [12, 17-19] Again, this can be partially explained by the low prevalence of delirium (10.8% to 13.6%) which can have a significant impact on PPV and NPV values. [38] It is also possible that the determination of delirium "present" for each instrument could be improved, as false positives for each instrument were high compared to true positives. In considering how to determine delirium "present" however, consideration of the purpose of delirium screening and the context in which screening occurs is important. Ideally, diagnostic instruments should return high PPV values. [39] Yet, if the aim of a screening instrument is to identify as much delirium as possible (i.e., be highly sensitive), particularly when paediatric delirium is consistently described as under-recognised, a lower PPV balanced by a high NPV is appropriate. [40]

This study found acceptable convergent validity between each of the screening instruments, indicating that they are measuring similar constructs. However, this was lower than expected

between the CAP-D and the psCAM-ICU. This may be reflective of the challenges associated with detecting symptoms of delirium in young children, especially as both the CAP-D and psCAM-ICU appeared to be affected by this at the univariate level. It is also possible that disparity in the design of the two instruments, particularly in how both instruments evaluate attention and awareness, and disorganised thinking, contributed to differences observed. Conversely, both the CAP-D and the psCAM-ICU, and to a lesser degree the pCAM-ICU, were able to demonstrate acceptable ability in differentiating between delirium and pain as distinct constructs. That there is some small overlap between a delirium screening instrument and symptoms of pain is not surprising however, as similarities in symptom presentation between the two constructs has previously been discussed. [7, 8]

Delirium screening instruments are invaluable for detecting changes in attention, awareness and cognition, but are not diagnostic tools, and 'delirium positive' scores should always be followed up by clinical review. This was underscored by the acceptable but comparatively low indices of agreement with the reference standard, [15-19] potentially reflecting the variable performance of these instruments when not conducted by expert members of a research team. Alternatively, each of the screening instruments were compared to the DSM-5 diagnostic criteria for delirium while previous studies used the DSM-IV diagnostic criteria as the reference standard. The impact of differences in how delirium is measured between the two editions of the DSM therefore cannot be ruled out, especially as research in the adult literature indicates that the two editions can result in discordant assessments. [41]

This study is not without limitations. We made diligent attempts to consent every eligible patient to the study, but were unable to do so. As a result, our sample is skewed towards patients with lower acuity and the first assessment occurring later in admission, and consequently there is potential that those children who were missed had a higher prevalence of delirium. Although this may limit generalisability of the study's results to the entire PICU, arguably this sample is representative of the PICU population for which delirium assessment is most feasible. It is also possible that the concurrent screening using the CAP-D and ps/pCAM-ICU influenced and biased the results of both

nursing assessments. However, as the CAP-D is an indicator of delirium symptoms over a period of time (in this study, over 6 hours) while , conversely, the ps/pCAM-ICU measure delirium at a point-in-time, the risk of the results biasing the other is low. Additionally, we included up to five daily paired assessments, on the basis that delirium is a disorder characterised by rapidly developing and fluctuating symptoms. As the majority of participants only had one assessment per admission, we did not adjust for repeated measures, however the potential for repeated-measures bias is a limitation of the current study. Finally, the lower sample size of ps/pCAM-ICU assessments hindered concurrent evaluation of the screening instruments, which limits the scope of making direct comparisons between the instruments.

A major strength of this study was the inclusion of end-users in the validation of the three instruments in this population. Previously, validation of the CAP-D and the ps/pCAM-ICU has primarily relied on trained, expert users, restricting the interpretation of those results for routine clinical use. Future research is required to directly compare the accuracy and feasibility of each of these tools, evaluate their use in other medical settings (e.g., general hospital wards, oncology) and extend comparison to other promising screening instruments not available at the time of conducting this study (i.e., the SOS-PD). [42]

CONCLUSIONS

To date, limited research that evaluates paediatric screening instruments outside of their developing centres exists. This prospective observational study aimed to evaluate the psychometric performance of three delirium screening instruments: the CAP-D, the psCAM-ICU and the pCAM-ICU as part of routine clinical practice. As a screening instrument, the CAP-D performs well, though all patients with 'delirium positive' screens should be closely monitored, with follow-up diagnosis required to confirm delirium. The ps/pCAM-ICU both provide valuable, objective assessments of delirium in critically ill children, but may benefit from an accompanying diagnostic aide, similar to the CAP-D, to guide their use in very young children or children with developmental delay. Given the

variation of these results and previous research, further evaluation of their implementation into routine clinical practice is needed.

REFERENCES

1. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-V*. 5th ed. Washington, DC: American Psychiatric Publishing; 2013.
2. Traube C, Silver G, Reeder RW, Doyle H, Hegel E, Wolfe HA, et al. Delirium in Critically Ill Children: An International Point Prevalence Study. *Crit Care Med*. 2017;45(4):584-590.
3. Traube C, Silver G, Gerber LM, Kaur S, Mauer EA, Kerson A, et al. Delirium and Mortality in Critically Ill Children: Epidemiology and Outcomes of Pediatric Delirium. *Crit Care Med*. 2017.
4. Traube C, Mauer EA, Gerber LM, Kaur S, Joyce C, Kerson A, et al. Cost Associated With Pediatric Delirium in the ICU. *Crit Care Med*. 2016:1.
5. Smith HAB, Gangopadhyay M, Goben CM, Jacobowski NL, Chestnut MH, Thompson JL, et al. Delirium and Benzodiazepines Associated With Prolonged ICU Stay in Critically Ill Infants and Young Children. *Crit Care Med*. 2017.
6. Pretto M, Spirig R, Milisen K, Degeest S, Regazzoni P, Hasemann W. Effects of an interdisciplinary nurse-led Delirium Prevention and Management Program (DPMP) on nursing workload: a pilot study. *Int J Nurs Stud*. 2009;46(6):804-812.
7. Paterson RS, Kenardy JA, De Young AC, Dow BL, Long DA. Delirium in the Critically Ill Child: Assessment and Sequelae. *Dev Neuropsychol*. 2017;42(6):387-403.
8. Harris J, Ramelet AS, van Dijk M, Pokorna P, Wielenga J, Tume L, et al. Clinical recommendations for pain, sedation, withdrawal and delirium assessment in critically ill infants and children: an ESPNIC position statement for healthcare professionals. *Intensive Care Med*. 2016;42(6):972-986.
9. Smith HAB, Fuchs DC, Pandharipande PP, Barr FE, Ely EW. Delirium: An emerging frontier in the management of critically ill children. *Anesthesiol Clin*. 2011;29(4):729-750.
10. Gesin G, Russell BB, Lin AP, Norton HJ, Evans SL, Devlin JW. Impact of a delirium screening tool and multifaceted education on nurses' knowledge of delirium and ability to evaluate it correctly. *Am J Crit Care*. 2012;21(1):e1-11.
11. Hshieh TT, Yue J, Oh E, Puelle M, Dowal S, Travison T, et al. Effectiveness of multicomponent nonpharmacological delirium interventions: a meta-analysis. *JAMA Intern Med*. 2015;175(4):512-520.
12. Daoud A, Duff JP, Joffe AR, Alberta Sepsis N. Diagnostic accuracy of delirium diagnosis in pediatric intensive care: a systematic review. *Crit Care*. 2014;18(5):489.
13. Dammeyer JA, Mapili CD, Palleschi MT, Eagle S, Browning L, Heck K, et al. Nurse-led change: a statewide multidisciplinary collaboration targeting intensive care unit delirium. *Crit Care Nurs Q*. 2012;35(1):2-14.
14. Balas MC, Vasilevskis EE, Burke WJ, Boehm L, Pun BT, Olsen KM, et al. Critical care nurses' role in implementing the "ABCDE bundle" into practice. *Crit Care Nurse*. 2012;32(2):35-38, 40-37; quiz 48.
15. Silver G, Traube C, Kearney J, Kelly D, Yoon MJ, Nash Moyal W, et al. Detecting pediatric delirium: development of a rapid observational assessment tool. *Intensive Care Med*. 2012;38(6):1025-1031.
16. Silver G, Kearney J, Traube C, Hertzig M. Delirium screening anchored in child development: The Cornell Assessment for Pediatric Delirium. *Palliative and Supportive Care*. 2015;13(4):1005-1011.
17. Traube C, Silver G, Kearney J, Patel A, Atkinson TM, Yoon MJ, et al. Cornell Assessment of Pediatric Delirium: a valid, rapid, observational tool for screening delirium in the PICU. *Crit Care Med*. 2014;42(3):656-663.
18. Smith HAB, Boyd J, Fuchs DC, Melvin K, Berry P, Shintani A, et al. Diagnosing delirium in critically ill children: Validity and reliability of the Pediatric Confusion Assessment Method for the Intensive Care Unit. *Crit Care Med*. 2011;39(1):150-157.
19. Smith HAB, Gangopadhyay M, Goben CM, Jacobowski NL, Chestnut MH, Savage S, et al. The

Preschool Confusion Assessment Method for the ICU: Valid and Reliable Delirium Monitoring for Critically Ill Infants and Children. *Crit Care Med*. 2016;44(3):592-600.
20. Luetz A, Gensel D, Muller J, Weiss B, Martiny V, Heinz A, et al. Validity of Different Delirium Assessment Tools for Critically Ill Children: Covariates Matter. *Crit Care Med*. 2016.
21. Adamis D, Sharma N, Whelan PJ, Macdonald AJ. Delirium scales: A review of current evidence. *Aging Ment Health*. 2010;14(5):543-555.
22. Fleiss JL, Levin, B. and Paik, M. C. The Measurement of Interrater Agreement. In: Shewart WA, Wilks SS, editors. *Statistical Methods for Rates and Proportions*. 3rd Edition ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2003. p. 598-626.
23. Sikich N, Lerman J. Development and Psychometric Evaluation of the Pediatric Anesthesia Emergence Delirium Scale. *Anesthesiology*. 2004;100(5):1138-1145.
24. Ely EW, Inouye SK, Bernard GR, Gordon S, Francis J, May L, et al. Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *The Journal of the American Medical Association*. 2001;286(21):2703-2710.
25. Ely EW, Margolin R, Francis J, May L, Truman B, Dittus R, et al. Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *Crit Care Med*. 2001;29(7):1370-1379.
26. Hart RP, Best AM, Sessler CN, Levenson JL. Abbreviated cognitive test for delirium. *J Psychosom Res*. 1997;43(4):417-423.
27. Gangopadhyay M, Smith H, Pao M, Silver G, Deepmala D, De Souza C, et al. Development of the Vanderbilt Assessment for Delirium in Infants and Children to Standardize Pediatric Delirium Assessment By Psychiatrists. *Psychosomatics*. 2017;58(4):355-363.
28. Serafim RB, Soares M, Bozza FA, Lapa ESJR, Dal-Pizzol F, Paulino MC, et al. Outcomes of subsyndromal delirium in ICU: a systematic review and meta-analysis. *Crit Care*. 2017;21(1):179.
29. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-549.
30. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
31. IBM Corp. SPSS Statistics for Mac, Version 25.0. Armonk, NY: IBM Corp; 2017.
32. Bujang MA, Adnan TH. Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis. *J Clin Diagn Res*. 2016;10(10):YE01-YE06.
33. Hosmer DW, Jr., Lemeshow, S., & Sturdivant, R. X. . *Applied logistic regression*. 3rd Edition ed. Hoboken, NJ: Wiley; 2013.
34. Karras DJ. Statistical methodology: II. Reliability and validity assessment in study design, Part B. *Acad Emerg Med*. 1997;4(2):144-147.
35. Ramelet AS, Rees N, McDonald S, Bulsara M, Abu-Saad HH. Development and preliminary psychometric testing of the Multidimensional Assessment of Pain Scale: MAPS. *Paediatr Anaesth*. 2007;17(4):333-340.
36. Nunnally J, Bernstein L. *Psychometric theory*. New York: McGraw-Hill Higher, INC; 1994.
37. Bland JM, Altman DG. Cronbach's alpha. *BMJ*. 1997;314(7080):572.
38. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain*. 2008;8(6):221-223.
39. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Front Public Health*. 2017;5:307.
40. Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988;240(4857):1285-1293.
41. Meagher DJ, Morandi A, Inouye SK, Ely W, Adamis D, Maclullich AJ, et al. Concordance between DSM-IV and DSM-5 criteria for delirium diagnosis in a pooled database of 768

prospectively evaluated patients using the delirium rating scale-revised-98. *BMC Med*. 2014;12.
42. Ista E, van Beusekom B, van Rosmalen J, Kneyber MCJ, Lemson J, Brouwers A, et al. Validation of the SOS-PD scale for assessment of pediatric delirium: a multicenter study. *Crit Care*. 2018;22(1):309.

TABLE 1. Published properties of the CAP-D, psCAM-ICU and pCAM-ICU

Properties	CAP-D [16] (95% CI)	psCAM-ICU [18] (95% CI)	pCAM-ICU [17] (95% CI)
Sensitivity	94.1% (84.0-99.0)	78.0% (75.0-80.0)	78.0% (40.0-97.0)
Specificity	79.2% (74.0-85.0)	86.0% (84.0-88.0)	98.0% (91.0-99.9)
PPV	54.0% (44.0-64.0)	78.0% (75.0-80.0)	88.0% (51.0-99.9)
NPV	98.0% (94.0-99.6)	86.0% (85.0-88.0)	97.0% (88.0-99.8)

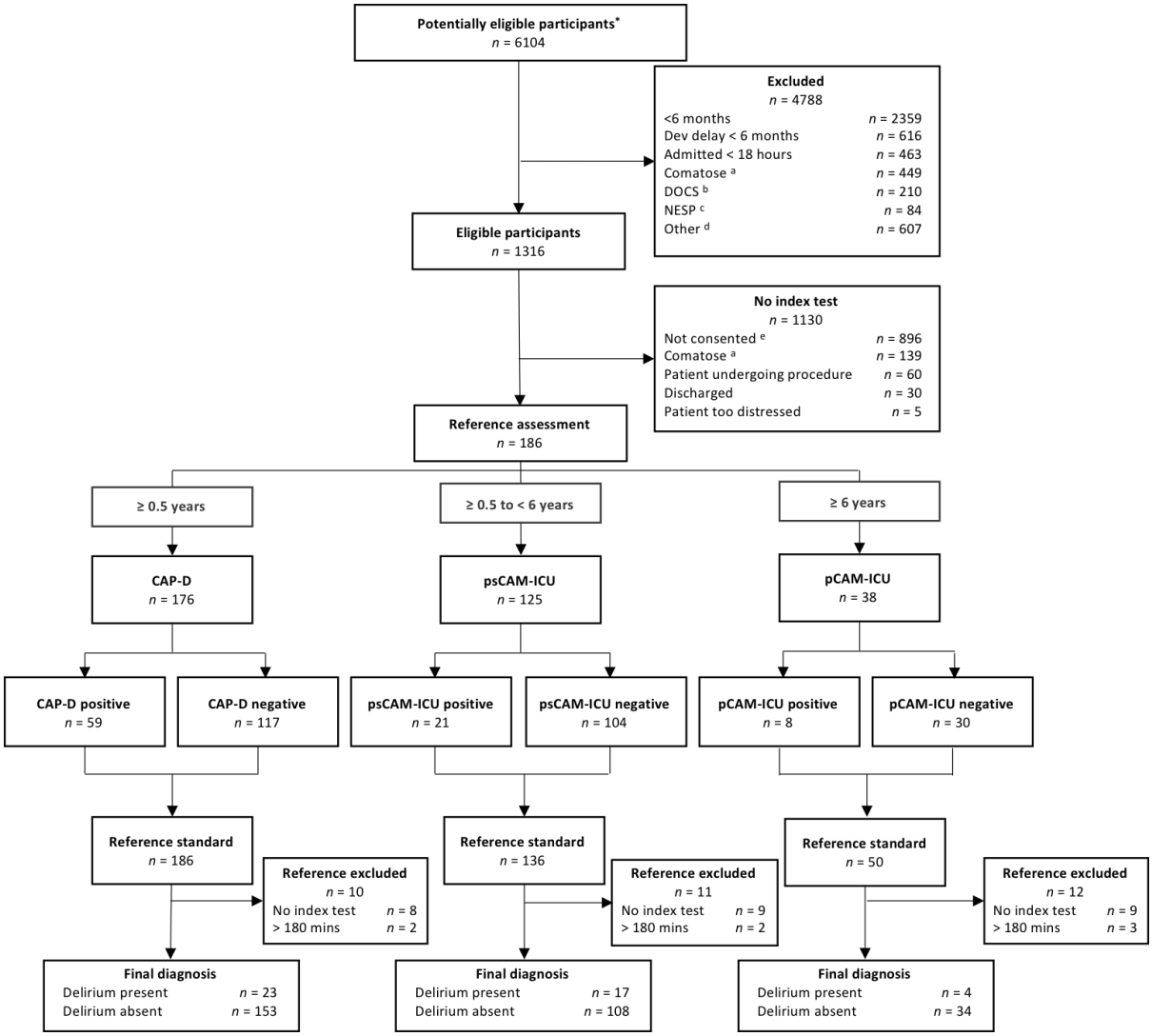

Figure 1. Participant flow diagram for all assessments included in analysis. *Potentially eligible accounts for all daily presentations over the 18 month study period; [a]Comatose refers to any patient with an SBS < -1 or with a GCS < 8 in patients who were not receiving sedation; [b]Under the care of the Department of Child Services; [c]NESP = Non-English speaking parent; [d]Other – Long-term patient previously declined; [e]Not consented – Primarily referred to patient discharged before approached for consent (81.5%).

TABLE 2. Demographic and Clinical Characteristics of Participants

Patient Characteristics (n =119)	*n* (% total)
Demographics	
Age at admission, years (mdn, IQR)	3.5 (1.2 – 10.9)
Gender (Female)	64 (54.0%)
Developmental Delay / Intellectual Impairment	13 (10.9%)
Socio-economic status (SES)[a] (mdn percentile rank, IQR)	59.0 (35.3 – 75.0)
Primary diagnosis	
Surgical (excl. cardiac)	35 (29.4%)
Respiratory	26 (21.8%)
Cardiovascular (incl. post-op)	21 (17.6%)
Gastrointestinal/Renal	21 (17.6%)
Neurologic	10 (8.4%)
Injury	3 (2.5%)
Other	21 (17.6%)
PICU Course	
Paediatric Index of Mortality (PIM-3) score, at discharge (mdn, IQR)	0.6 (0.2 – 1.1)
Length of mechanical Ventilation, hours (mdn, IQR)	5.25 (0.0 – 59.5)
Mechanical Ventilation (Yes)	71 (59.7%)
PICU Length of Stay, days (mdn, IQR)	2.5 (1.8 – 6.6)
Outcomes	
Overall delirium prevalence (reference standard)[b]	42 (22.6%)
Level of sedation, SBS Score (mdn, IQR)	0.0 (0.0 – 1.0)
Number of Assessments (mdn; IQR)	1.0 (1.0 – 2.0)
Day of admission at time of first assessment (mdn, IQR)[c]	2.0 (1.0-5.0)
Time between reference and index tests, minutes (mdn, IQR)	31:00 (15:00 – 60:00)
In hospital mortality	1 (0.8%)

[a]SES derived from the Australian Bureau of Statistics Index of Relative Socio-economic Advantage and Disadvantage percentile scores[26]; [b] Reported at assessment level for entire sample (Total *n* = 186); [c]Day 0 is first day of admission

TABLE 3. Predictive validity of delirium screening assessments by instrument

	CAP-D (*n* = 176) (95% CI)		psCAM-ICU (*n*= 125) (95% CI)		pCAM-ICU (*n* = 38) (95% CI)	
Properties	Absent	Present	Absent	Present	Absent	Present
Positive Reference	2	21	7	10	1	3
Negative Reference	115	38	97	11	28	5
Prevalence [a]	23 (13.1%)		17 (13.6%)		4 (10.8%)	
Sensitivity	91.3 (72.0-98.3)		58.8 (32.9-81.6)		75.0 (19.4-99.4)	
Specificity	75.2 (67.5-82.0)		89.8 (82.5-94.8)		84.9 (68.1-98.9)	
PPV	35.6 (29.0-42.8)		47.6(31.4-64.4)		37.5 (18.3-61.7)	
NPV	98.3 (93.8-99.5)		93.3 (88.7-96.1)		96.6 (83.6-99.4)	
AUC	83.2 (75.2-91.3)**		74.3 (59.6-89.0)*		80.1 (54.2-100.0)	

* = *p*-value < .01; ** = *p*-value < .001; [a] Reported prevalence rate based on subsample of reference standard assessments paired with a corresponding screening instrument.

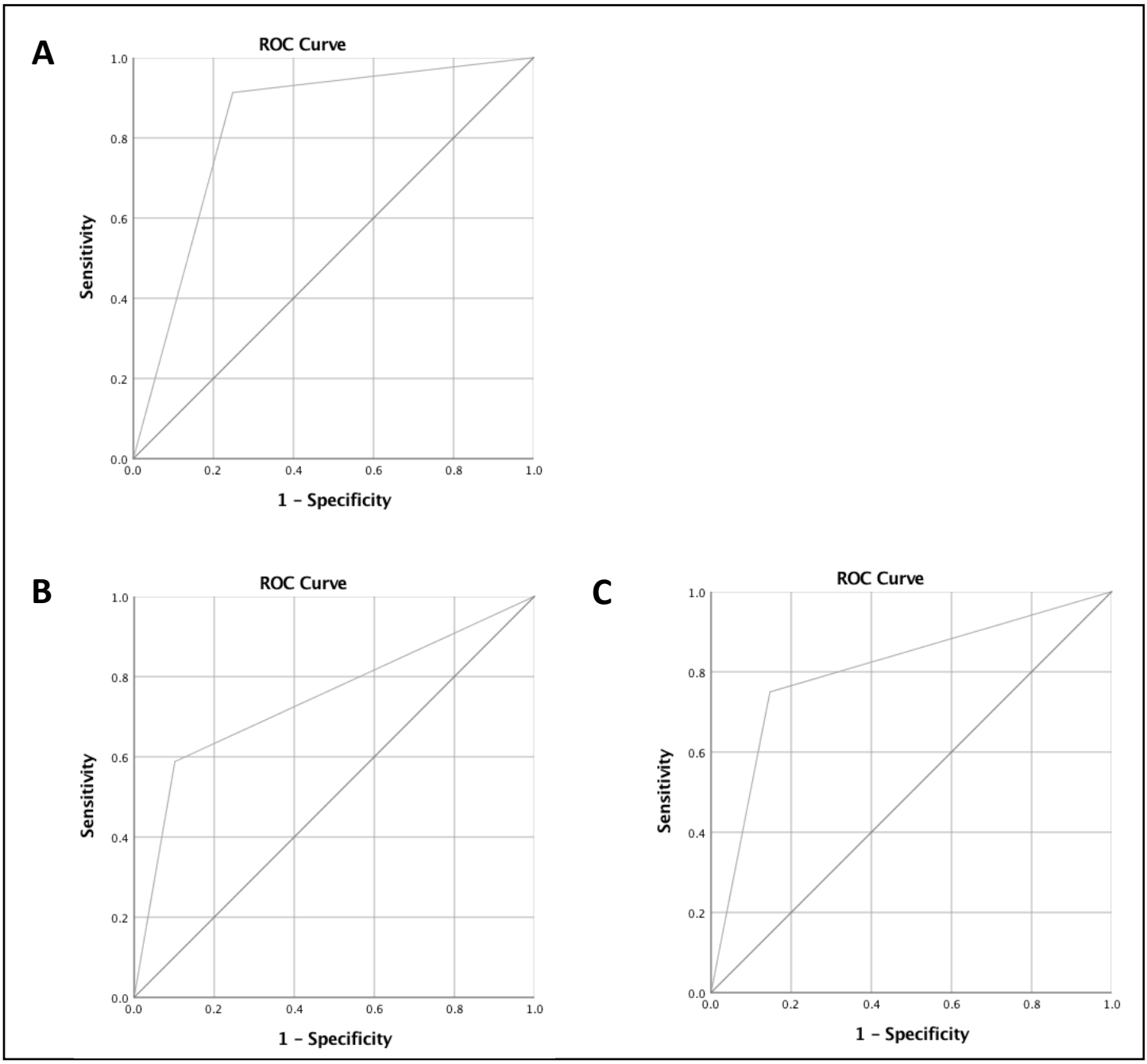

Figure 2. ROC Curve analysis plotting sensitivity and specificity of the screening instrument against the reference standard (the diagonal); (**A)** CAP-D (AUC = .83, $p < .001$); (**B**) psCAM-ICU (AUC = .74, $p = .001$); and **(C)** pCAM-ICU (AUC = .80, $p = .051$).

TABLE 4. Logistic Regression predicting discordance between delirium diagnosis and screening instrument outcome.

Variable	*B*	S.E.	Wald	*df*	*p*	Adjusted ORs	95% C.I. Lower	95% C.I. Upper
CAP-D[a]								
PICU LOS[b]	0.001	0.001	3.0	1	.102	1.0	1.0	1.0
Age	-0.1	0.04	6.2	1	.013	0.9	1.0	1.2
Sedation	-1.4	0.5	7.5	1	.006	4.1	1.5	11.5
SSD	-1.5	0.5	10.7	1	.001	4.4	1.8	10.5

Note. OR = odds ratio; SSD refers to Subsyndromal Delirium; sedation (currently receiving any sedation, yes/no) is absent compared to present; SSS is absent compared to present. [a]Nagelkerke R^2 =0.3, $\chi^2(4) = 38.0$, $p < .001$; [b]PICU length of stay (LOS) entered into the model as control variable.