British Journal of Anaesthesia **111** (4): 612–18 (2013)
Advance Access publication 8 May 2013 · doi:10.1093/bja/aet167

BJA

NEUROSCIENCES AND NEUROANAESTHESIA

Evaluation of two delirium screening tools for detecting post-operative delirium in the elderly

K. J. Neufeld[1*], J. S. Leoutsakos[1], F. E. Sieber[2], D. Joshi[3,6], B. L. Wanamaker[4], J. Rios-Robles[1,7] and D. M. Needham[5]

[1] Department of Psychiatry and Behavioral Sciences and [2] Department of Anesthesiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[3] Department of Internal Medicine, Good Samaritan Hospital, Baltimore, MD, USA
[4] Johns Hopkins University School of Medicine, Baltimore, MD, USA
[5] Division of Pulmonary and Critical Care Medicine, Department of Physical Medicine and Rehabilitation, School of Medicine, Johns Hopkins University, Baltimore, MD, USA
[6] Present address: Respiratory Institute, Cleveland Clinic Foundation, Cleveland, OH, USA
[7] Present address: Colegio de Médicos Cirujanos de Puerto Rico, San Juan, PR, USA

* Corresponding author. E-mail: kneufel2@jhmi.edu

Editor's key points

- The gold standard method for diagnosing delirium is psychiatric evaluation using the DSM-IV criteria.
- The CAM-ICU and NuDESC delirium screening tools perform well in adult ICU patients.
- The current study assessed their performance as post-anaesthesia screening tools in elderly patients.
- Compared with formal psychiatric evaluation, sensitivity was poor, while specificity was good.

Background. Postoperative delirium in the elderly is common and associated with poor outcomes, but often goes unrecognized. Delirium screening tools, validated in postoperative settings are lacking. This study compares two screening tools [Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) and Nursing Delirium Symptom Checklist (NuDESC)] with a DSM-IV-based diagnosis of delirium, conducted by neuropsychiatric examination in postoperative settings.

Methods. Consecutive English-speaking patients, ≥70 yr, undergoing surgery with general anaesthesia and capable of providing informed consent, were recruited. Diagnostic test characteristics were compared for each screening tool *vs* neuropsychiatric examination, both in the Post-Anaesthesia Care Unit (PACU), and daily during inpatient hospitalization, adjusting for repeated measures.

Results. Neuropsychiatric examination identified delirium in 45% of 91 patients evaluated in the PACU and in 32% of 166 subsequent delirium assessments on the ward in the 58 admitted patients. The sensitivity [95% confidence interval (CI)] of delirium detection of the CAM-ICU in the PACU, and in all repeated assessments was 28% (16–45%) and 28% (17–42%), respectively; for the NuDESC (scoring threshold ≥2), 32% (19–48%) and 29% (19–42%), respectively, and the NuDESC (threshold ≥1), 80% (65–91%) and 72% (60–82%), respectively. Specificity was >90% for both the CAM-ICU and the NuDESC (threshold ≥2); specificity for the NuDESC (threshold ≥1), in the PACU was 69% (54–80%) and 80% (73–85%) for all assessments.

Conclusions. While highly specific, neither CAM-ICU nor NuDESC (threshold ≥2) are adequately sensitive to identify delirium post-operatively; NuDESC (threshold ≥1) increases sensitivity, but reduces specificity.

Keywords: aged; delirium; neuropsychological tests; perioperative period; sensitivity and specificity

Accepted for publication: 21 March 2013

Delirium in the postoperative setting may occur in up to 50% of elderly patients, and is associated with increased mortality, longer hospital stays, and cognitive and functional decline.[1–3] Elderly patients are at high risk of delirium after surgical procedures,[4] especially when performed with general anaesthesia.[5] Delirium is often overlooked in hospitalized patients unless a routine screening programme is implemented.[6] Screening for delirium in the Post-Anaesthesia Care Unit (PACU) and on the postoperative hospital wards is generally not part of routine clinical practice; however, there is growing interest in implementing validated screening tools in these settings.[7]

A commonly used delirium screening tool is the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). The CAM-ICU has a large literature demonstrating its validity

in detecting delirium in mechanically ventilated ICU patients with two systematic reviews reporting pooled sensitivities of 76–80% and a pooled specificity of 96%.[8 9] This instrument is less sensitive in other patient populations, such as oncology in-patients[10] and elderly patients in the Emergency Department.[11 12] There are no studies documenting its performance in the PACU setting, but given its validation in patients receiving sedation and mechanical ventilation, evaluation for use in the PACU and postoperative inpatient setting is appropriate.

The Nursing Delirium Symptom Checklist (NuDESC) delirium tool was validated in 59 oncology/internal medicine inpatients, with a sensitivity and specificity of 86 and 87%, respectively, when using a scoring threshold of $\geq$2.[13 14] The NuDESC also has been used in the ICU, with a sensitivity and specificity of 83 and 81%, respectively.[15] The NuDESC was used to evaluate patients in the PACU where it had a sensitivity and specificity of 95 and 87%, respectively,[16] and in postoperative inpatients with a sensitivity and specificity of 98 and 92%, respectively.[7] In these latter two studies, the NuDESC was compared against the DSM-IV criteria for delirium, but without using an independent neuropsychiatric examination to determine the presence of the criteria.

Given the growing interest in early diagnosis of delirium, it is important to determine the best instruments for delirium detection in the PACU and postoperative in-patient settings. Thus, the objective of this study was to evaluate the diagnostic test characteristics of the CAM-ICU and NuDESC delirium screening instruments against a DSM-IV-based reference standard diagnosis of delirium conducted using a neuropsychiatric examination, among elderly postoperative patients in the PACU and postoperative in-patient settings.

Methods

The Johns Hopkins Institutional Review Board approved this study. The requirement for written informed consent was waived, with verbal informed consent obtained from all participants.

Patient population

Consecutive patients, undergoing surgery with general anaesthesia at one teaching hospital during July and August 2010, were recruited to participate in the study. Because standard medical practice at our institution limits exposure to intra-operative sedatives for elderly patients undergoing surgery with only regional anaesthesia, the study focused on patients receiving general anaesthesia. Inclusion criteria were: (i) $\geq$70 yr old, (ii) receipt of general anaesthesia during surgery, and (iii) English-speaking. Exclusion criteria were: (i) severe hearing impairment, (ii) cognitively incapable of providing informed consent, using an IRB-approved structured evaluation of decision-making capacity,[17] and (iii) prior enrolment in this study (for patients having repeat surgery).

Pre-surgical assessment

Patients were interviewed by telephone if undergoing elective surgery or in-person in the preoperative area if surgery was emergent or a telephone interview could not be completed. Patient interviews included baseline data (e.g. socio-demographics) and administration of the 30-item Mini-Mental State Examination (MMSE)[18] for in-person interviews, or a 26-item validated version of the MMSE when performing the telephone interview with conversion to the 30-item MMSE scoring system.[19] Data were abstracted from the medical records to determine patients' Charlson Comorbidity Index,[20] surgery type, and the American Society of Anesthesiologists Physical Classification System score (ASA score).[21]

Confusion Assessment Method for the Intensive Care Unit

The CAM-ICU evaluates the following four 'features' of delirium: (i) an acute change in mental status or fluctuation in the level of consciousness over the prior 24 h, (ii) inattention, (iii) disorganized thinking, and (iv) an altered level of consciousness. Inattention and disorganized thinking are each evaluated using brief, standardized testing specified by the CAM-ICU tool. Administration of the CAM-ICU takes 1–2 min.[22] The CAM-ICU was administered in the PACU and on the wards by three research staff members who each received one-on-one training plus quality assurance review of 20 independent assessments before the start of the study by a CAM-ICU expert at our institution (D.M.N.). The κ-statistic for agreement between the expert and each of the assessors was 1.0 indicating perfect agreement.[23]

Nursing Delirium Screening Checklist

The NuDESC evaluates delirium based on observation of the following five features, as defined by the instrument: (i) disorientation, (ii) inappropriate behaviour, (iii) inappropriate communication, (iv) illusions/hallucinations, and (v) psychomotor retardation. Each item is scored based on its severity (0=absence, 1=mild, and 2=severe).[13] Original validation studies used a total score $\geq$2 for delirium detection with a subsequent study suggesting use of a threshold of $\geq$1.[24] The NuDESC takes 1–2 min to complete. The NuDESC was administered by three research staff who rated each of the items based on brief interaction and observation with the patient, and interview of the patient's nurse. Each independent evaluator underwent training by a board-certified psychiatrist (K.J.N.), followed by a quality assurance review of at least six independent co-rated assessments with excellent to perfect agreement (κ=0.93–1.0) of each evaluator.

DSM-IV delirium reference standard using a neuropsychiatric examination

The delirium reference standard diagnosis was performed by either a board-certified psychiatrist (K.J.N.), or a fourth year psychiatry resident (J.R.-R.) who had performed 25 delirium assessments under supervision of the board-certified psychiatrist before starting the study. The reference standard diagnosis of delirium was based on the DSM-IV criteria[25] after full mental status examination of the patient, including administration of the MMSE, review of the medical records, and

interview of the patient's nurse, and any available family members. These two reference raters demonstrated excellent agreement (κ=0.93) with co-rating 15 neuropsychiatric examinations during this study.

Timing of evaluations

The CAM-ICU, NuDESC, and neuropsychiatric examination were conducted independently by separate raters, who were blinded to the results of the other assessments. All three assessments were completed within 60 min of each other. Order of assessments was varied. Timing of initiation of delirium assessments in the PACU was standardized, occurring once the patient reached an Aldrete score $\geq$9 indicating an appropriate level of wakefulness, and haemodynamic and respiratory stability for transfer from the PACU.[26] For those patients admitted to the in-patient surgical wards and ICU, study staff continued independently conducting all three delirium assessments on weekdays for the duration of the patient's hospitalization.

Statistical analysis

κ-statistics were calculated to measure level of agreement between each of the screening tests *vs* the reference rater diagnosis, with qualitative interpretation of κ-statistics based on prior guidelines.[23] Diagnostic test characteristics (sensitivity, specificity, positive and negative predictive values, and likelihood ratios) were calculated separately for the CAM-ICU and NuDESC *vs* the reference standard, for the PACU evaluation and for all delirium assessments (i.e. both PACU and in-patient assessments) with adjustment for repeated assessments of individual patients conducted using generalized linear regression models with a random intercept for each patient.[27] Data analyses were performed using STATA v.11.[28]

Results

Patient characteristics

A total of 196 patients met inclusion criteria during the study period, with 74 (38%) meeting exclusion criteria, 27 (14%) declining to participate, and 4 excluded for other reasons, leaving 91 consenting participants (Fig. 1). The mean age [standard deviation (SD)] of the 91 participants was 79 (6) yr and 58% were women (Table 1). The majority (78%) of patients were living independently in their own homes before surgery. Anaesthetic technique was comparable among subjects, consisting of propofol induction followed by maintenance with isoflurane, narcotic, and muscle relaxation as needed. Forty-six per cent (n=42) of patients received midazolam. The doses of narcotic (administered both intra-operatively and in the PACU), and of intra-operative propofol and midazolam were not statistically significantly different between those with and without delirium in the PACU after recovery from anaesthesia (Table 2).

Delirium evaluations

A total of 91 patients had a neuropsychiatric reference rating for delirium diagnosis in the PACU, with 58 (63%) admitted to an inpatient surgical ward with a total of 166 subsequent delirium assessments (Fig. 1). The median [interquartile range (IQR)] number of minutes required for each delirium assessment was: neuropsychiatric examination 10 (9–13), CAM-ICU 1 (1–2), and NuDESC 2 (2–3). The median time (IQR) from operating room exit to starting the neuropsychiatric examination in the PACU was 48 (33–62) and 42 (28–53) min (P=0.53) for those with and without delirium in the PACU after recovery from anaesthesia, respectively. The median (IQR) absolute difference in time between the neuropsychiatric examination and each blinded administration of delirium screening tools was 14 (6–31) and 13 (5–31) min for the CAM-ICU and the NuDESC, respectively, with the order of administration of the three delirium assessments varied across patients. The median (IQR) time between independent administrations of the two delirium screening tools was 7 (3–20) min.

Delirium prevalence and diagnostic test characteristics of CAM-ICU and NuDESC

Figure 1 summarizes the prevalence of delirium, using each assessment method, in the PACU and postoperative surgical ward. The neuropsychiatric reference standard diagnosed delirium in 45% of assessments in the PACU, and in 32% of all assessments done in both the PACU and in-patient surgical ward combined. κ-statistics (98% CI) for agreement of each screening instrument *vs* the reference standard delirium diagnosis indicated 'poor' agreement for both the CAM-ICU and the NuDESC (scoring threshold $\geq$2) in the PACU and across all assessments (Table 3). For the NuDESC (threshold $\geq$1) agreement was 'intermediate to good' for both the PACU and all assessments (Table 3).

The diagnostic test characteristics of the CAM-ICU and NuDESC screening instruments *vs* the reference standard delirium diagnosis demonstrated low sensitivities (range: 28–32%) and high specificities (range: 82–100%) for CAM-ICU and NuDESC (threshold $\geq$2) (Table 3). For NuDESC (threshold $\geq$1), the sensitivity was higher (72–80%) but with lower specificities (69–80%) compared with using the original scoring threshold (Table 3).

Discussion

This prospective study evaluated the diagnostic test characteristics of two popular delirium screening instruments (CAM-ICU and NuDESC) in 91 elderly postoperative patients. The screening instruments were compared against a reference standard diagnosis of delirium conducted by a psychiatry-trained physician using a neuropsychiatric examination. Both the CAM-ICU and NuDESC (cut $\geq$2) demonstrated high specificity, but inadequate sensitivity for screening purposes and thus, frequently may not detect true episodes of delirium in the PACU after recovery from anaesthesia and surgical ward. The NuDESC (threshold $\geq$1) had improved sensitivity, but reduced specificity.

The observation that the CAM-ICU is a sensitive and highly specific screening instrument useful for delirium detection among sedated, mechanically ventilated patients in the ICU,[8 9] created interest in its application in elderly postoperative patients who had recently recovered from general

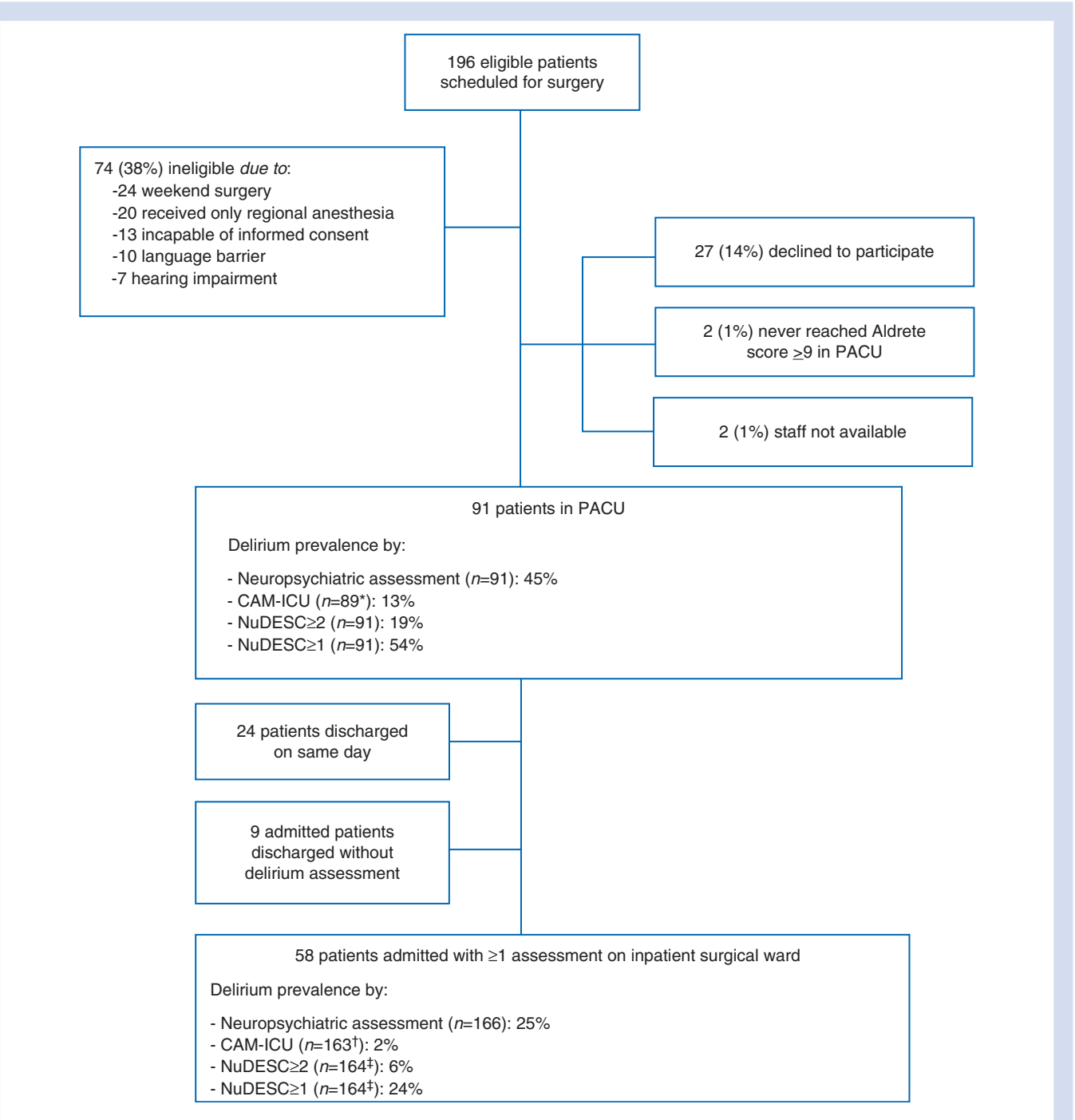

Fig 1 Consort flow diagram and prevalence of delirium. (*) Two CAM-ICU assessments missing because of unavailability of blinded staff member. (†) Three CAM-ICU assessments missing because of unavailability of blinded staff member. (‡) Two NuDESC assessments missing because of unavailability of blinded staff member.

anaesthesia and mechanical ventilation. In this study's population of elderly postoperative patients without critical illness, the CAM-ICU was highly specific with a high positive predictive value, indicating that a positive CAM-ICU did represent true delirium. However, given its low sensitivity and negative predictive value, the CAM-ICU did not perform well when used once as a screening tool for delirium detection in the PACU post-anaesthetic recovery, or once daily in the postoperative inpatient setting where the objective is to detect the vast majority of patients with delirium.[29] This finding was similar to that demonstrated in a sample of 139 oncology ward patients (delirium prevalence 26%) where sensitivity and specificity of the CAM-ICU was 18 and 99%, respectively.[10] Some may propose that having a screening instrument with high specificity is better than having no instrument. However, this proposal has not yet been empirically demonstrated in this population as

the outcomes of delirium in the PACU after recovery from anaesthesia and very early in postoperative inpatient setting have not yet been fully studied. Thus, future work should focus on both the short- and long-term implications of delirium in the PACU early in the recovery course, in addition to evaluating the best approaches to delirium screening in this setting.

The sensitivity of the NuDESC (threshold ≥ 2) in the PACU (32%) and surgical wards (29%) is lower than previously reported sensitivities of $\geq 95\%$ in PACU and inpatient surgical settings.[7 16] In the current study, the DSM-IV diagnosis of delirium included a neuropsychiatric examination which may differ from the prior study that did not describe a neuropsychiatric examination as a part of the delirium reference standard. This issue of delirium diagnostic techniques used for the reference standard is important because daily mental status testing and application of a validated diagnostic algorithm greatly increase estimates of delirium prevalence compared with methods that are less rigorous.[30] It is possible that as a result, the reference rater may have detected more subtle mental status changes and included more mild cases of delirium, therefore resulting in poorer sensitivity when comparing the other screening tools to this diagnostic standard. The current analysis does not include a severity measure of delirium and future work should include this comparison. Other potential sources of discordance in diagnosis include the NuDESC relying solely on the rater's observations of the patient and offers little standardization of how to rate the presence and severity of its five features of delirium; hence, variation in NuDESC ratings may have contributed to these differences in findings compared with prior studies. Further standardization of the instrument's administration and evaluation of inter-rater reliability are important areas for development.

Strengths of our study include a rigorous design in which the screening tools and neuropsychiatric examination were administered with each result blinded from the others, yet all performed close in time, with examiners receiving extensive training and demonstrating high agreement on quality assurance evaluations with expert raters. In addition, the neuropsychiatric examination, which provided the basis for the reference rater diagnosis, included direct cognitive testing and daily mental status examination of the patient, and also review of collateral information from the medical chart, nursing staff and family regarding the patient's mental state.

The study also has potential limitations. First, the NuDESC was administered by trained research staff rather than the nurse caring for the patient; hence, it is unclear if the results would have differed if administered directly by the nurses. Given the desire for very high-inter-rater reliability, the necessity for blinding among assessors, and completing all three delirium evaluations within a short-time frame, it was not feasible to rigorously train 120 nurses working in PACU and surgical wards in use of NuDESC for this study. However, in the current study, the research staff observed the patient and interviewed the patient's nurse for scoring of each of the five NuDESC features, as done in prior studies that reported sensitivities of 87–95% in the PACU[16] and inpatient surgical wards.[7] These studies may have detected the more noticeable hyperactive patients with mainly emergence delirium (generally referred to as agitated emergence from anaesthesia and brief in

Table 1 Characteristics of the patient population. Percentages may not add to 100% because of rounding. *Evaluated before surgery; 3 scores missing. †At time of surgery

	Total ($n=91$)
Age in years, mean (SD)	79 (6)
Women, *n* (%)	53 (58)
Race, *n* (%)	
White	81 (89)
Black	8 (9)
Asian	2 (2)
Education, *n* (%)	
Less than high school	25 (28)
High school diploma, or equivalent	25 (28)
Any college	30 (33)
Post-graduate	11 (12)
Living independently in own home, *n* (%)	71 (78)
Mini-Mental Status Exam (MMSE) score, mean (SD)*	25 (3)
Charlson Comorbidity Index, mean (SD)	2.2 (2.2)
ASA score† >2, *n* (%)	58 (64)
Type of surgery, *n* (%)	
Urinary and gynaecologic	25 (28)
Gastrointestinal	10 (11)
Orthopaedics	31 (34)
Other	25 (28)

Table 2 Characteristics of the anaesthesia drugs and administration dosages. *Student's *t*-test. †Two patients received no narcotics intra-operatively and neither developed delirium. ‡Nine patients received no propofol: five developed delirium, four did not. ¶Forty-seven patients received no midazolam: 22 developed delirium, 23 did not. §Fifty-four patients received no narcotics in the PACU: 26 developed delirium, 29 did not

Drug, mean dose (SD)	All patients ($n=91$)	No PACU delirium ($n=50$)	PACU delirium ($n=41$)	*P*-value*
Narcotic (intra-operative)†, mg of morphine equivalents per kg	0.23 (0.02)	0.23 (0.02)	0.24 (0.02)	0.86
Propofol (intra-operative)‡, mg per kg	2.0 (1.9)	1.9 (1.0)	2.2 (2.5)	0.43
Midazolam (intra-operative)¶, mg per kg	0.01 (0.02)	0.01 (0.01)	0.01 (0.02)	0.61
Narcotic (in PACU)§, mg of morphine equivalents per kg	0.006 (0.002)	0.008 (0.004)	0.004 (0.002)	0.23

Table 3 Diagnostic test characteristics of CAM-ICU and NuDESC delirium screening tools *vs* a DSM-IV-based delirium diagnosis using NPE. NPE, neuropsychiatric examination; CI, confidence interval. *Two assessments missing in PACU and three assessments missing on the surgical ward because of unavailability of blinded staff member. †Two assessments missing on the surgical ward because of unavailability of blinded staff member. ‡Adjusted for the repeated assessments of individual patients using generalized linear mixed models with a random effect for each patient

	CAM-ICU*		NuDESC† (scoring threshold ≥2)		NuDESC† (scoring threshold ≥1)	
	Evaluation in recovery room	Repeated evaluations in recovery room and inpatient ward‡	Evaluation in recovery room	Repeated evaluations in recovery room and inpatient ward‡	Evaluation in recovery room	Repeated evaluations in recovery room and inpatient ward‡
Patient days of evaluation	89	252	91	255	91	255
Delirium prevalence per NPE, %	45	32	45	32	45	32
κ-statistic, (95% CI)	0.28 (0.12–0.44)	0.24 (0.16–0.32)	0.25 (0.07–0.43)	0.23 (0.13–0.33)	0.48 (0.28–0.68)	0.42 (0.30–0.54)
Sensitivity, % (95% CI)	28 (16–45)	28 (17–42)	32 (19–48)	29 (19–42)	80 (65–91)	72 (60–82)
Specificity, % (95% CI)	98 (88–100)	100 (97–100)	92 (80–97)	96 (92–98)	69 (54–80)	80 (73–85)
Positive predictive value, % (95% CI)	92 (60–100)	98 (73–100)	76 (50–92)	73 (52–87)	67 (52–80)	58 (47–68)
Negative predictive value, % (95% CI)	64 (52–74)	83 (78–88)	62 (50–73)	79 (73–84)	81 (66–91)	88 (82–93)
Positive likelihood ratio, (95% CI)	14 (2–105)	33 (4–244)	4 (1–11)	5 (2–11)	3 (2–4)	3 (2–4)
Negative likelihood ratio, (95% CI)	0.7 (0.6–0.9)	0.8 (0.7–0.9)	0.7 (0.6–0.9)	0.8 (0.7–0.9)	0.3 (0.2–0.5)	0.5 (0.3–0.6)

duration) and hyperactive delirium subtypes lasting well into the postoperative period instead of both hyperactive and hypoactive subtypes.[31] Further work is required to compare the validity of this instrument, administered directly by PACU and surgical ward nurses, against a DSM-IV-based delirium reference standard that includes neuropsychiatric examination.

Secondly, the study focused only on elderly patients and was conducted at a single hospital, which may limit generalizabilty of the findings. However, our sample included all types of surgeries requiring general anaesthesia compared with prior studies that focused on delirium after a single type of surgery, such as hip fracture or cardiac procedures. Hence, these findings are more generalizable to the elderly surgical population at large. Restricting our study to elderly patients helped ensure that we were studying individuals at high risk of delirium in the postoperative setting and increased power of this study.

In conclusion, in our single-site study of elderly patients in the PACU and wards after general anaesthesia and surgery, the CAM-ICU and NuDESC screening instruments were not appropriate for routinely detecting delirium. Refinement of these instruments' content and scoring, and evaluation of new tools are appropriate given the high prevalence of delirium in postoperative elderly patients and its negative impact on patient outcomes.

Authors' contributions

All authors were involved in the concept and design of the study, the acquisition of data, analysis and interpretation of data, or all. All authors gave final approval of the version to be published.

Acknowledgements

The authors wish to thank Ms Veena Rao and Ms Jennifer Gibson Chambers for their assistance in helping to complete this research.

Declaration of interest

None declared.

Funding

This work was supported by internal funding from the Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, School of Medicine. The Johns Hopkins Predoctoral Clinical Research Training Program grant number 1TL1RR-025007 from the NIH and a Walker Award for Research in Psychiatry and Behavioral Sciences, Johns Hopkins University, School of Medicine supported B.L.W.'s efforts. The American Federation on Aging Research, Medical Student Training in Aging Research Program provided support of research personnel.

References

1 Nightingale S, Holmes J, Mason J, House A. Psychiatric illness and mortality after hip fracture. *Lancet* 2001; **357**: 1264–5

2 Bickel H, Gradinger R, Kochs E, Forstl H. High risk of cognitive and functional decline after postoperative delirium. *Dement Geriatr Cogn Disord* 2008; **26**: 26–31

3 Sanders RD, Pandharipande PP, Davidson AJ, Ma D, Maze M. Anticipating and managing postoperative delirium and cognitive decline in adults. *Br Med J* 2011; **343**: d4331

4 Inouye SK. Delirium in older persons. *N Engl J Med* 2006; **354**: 1157–65

5 Sieber FE, Barnett SR. Preventing postoperative complications in the elderly. *Anesthesiol Clin* 2011; **29**: 83–97

6 Spronk PE, Riekerk B, Hofhuis J, Rommes JH. Occurrence of delirium is severely underestimated in the ICU during daily care. *Intensive Care Med* 2009; **35**: 1276–80

7 Radtke FM, Franck M, Schust S, *et al.* A comparison of three scores to screen for delirium on the surgical ward. *World J Surg* 2010; **34**: 487–94

8 Gusmao-Flores D, Salluh JI, Chalhub RA, Quarantini LC. The Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) and Intensive Care Delirium Screening Checklist (ICDSC) for the diagnosis of delirium: a systematic review and meta-analysis of clinical studies. *Crit Care* 2012; **16**: R115

9 Neto AS, Nassar AP Jr, Cardoso SO, *et al.* Delirium screening in critically ill patients: a systematic review and meta-analysis. *Crit Care Med* 2012; **40**: 1946–51

10 Neufeld KJ, Hayat MJ, Coughlin JM, *et al.* Evaluation of two intensive care delirium screening tools for non-critically ill hospitalized patients. *Psychosomatics* 2011; **52**: 133–40

11 Han JH, Eden S, Shintani A, *et al.* Delirium in older emergency department patients is an independent predictor of hospital length of stay. *Acad Emerg Med* 2011; **18**: 451–7

12 Carpenter CR, Bassett ER, Fischer GM, Shirshekan J, Galvin JE, Morris JC. Four sensitive screening tools to detect cognitive dysfunction in geriatric emergency department patients: brief Alzheimer's Screen, Short Blessed Test, Ottawa 3DY, and the caregiver-completed AD8. *Acad Emerg Med* 2011; **18**: 374–84

13 Gaudreau JD, Gagnon P, Harel F, Tremblay A, Roy MA. Fast, systematic, and continuous delirium assessment in hospitalized patients: the nursing delirium screening scale. *J Pain Symptom Manage* 2005; **29**: 368–75

14 Gaudreau JD, Gagnon P, Harel F, Roy MA. Impact on delirium detection of using a sensitive instrument integrated into clinical practice. *Gen Hosp Psychiatry* 2005; **27**: 194–9

15 Luetz A, Heymann A, Radtke FM, *et al.* Different assessment tools for intensive care unit delirium: which score to use? *Crit Care Med* 2010; **38**: 409–18

16 Radtke FM, Franck M, Schneider M, *et al.* Comparison of three scores to screen for delirium in the recovery room. *Br J Anaesth* 2008; **101**: 338–43

17 Appelbaum PS. Clinical practice. Assessment of patients' competence to consent to treatment. *N Engl J Med* 2007; **357**: 1834–40

18 Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; **12**: 189–98

19 Newkirk LA, Kim JM, Thompson JM, Tinklenberg JR, Yesavage JA, Taylor JL. Validation of a 26-Point Telephone Version of the Mini-Mental State Examination. *J Geriatr Psychiatry Neurol* 2004; **17**: 81–7

20 Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *J Clin Epidemiol* 1994; **47**: 1245–51

21 ASA Physical Status Classification System. Available from http://www.asahq.org/Home/For-Members/Clinical-Information/ASA-Physical-Status-Classification-System (accessed 22 March 2012)

22 Pun BT, Gordon SM, Peterson JF, *et al.* Large-scale implementation of sedation and delirium monitoring in the intensive care unit: a report from two medical centers. *Crit Care Med* 2005; **33**: 1199–205

23 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74

24 Leung JM, Leung VW, Leung CM, Pan PC. Clinical utility and validation of two instruments (the Confusion Assessment Method Algorithm and the Chinese version of Nursing Delirium Screening Scale) to detect delirium in geriatric inpatients. *Gen Hosp Psychiatry* 2008; **30**: 171–6

25 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*, 4th Edn. Washington, DC: American Psychiatric Association, 1994

26 Aldrete JA. Modifications to the postanesthesia score for use in ambulatory surgery. *J Perianesth Nurs* 1998; **13**: 148–55

27 Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley-Interscience, 2004

28 StataCorp. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP, 2009

29 Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. *Br Med J* 2012; **345**: e3999

30 Rudolph JL, Marcantonio ER. Postoperative delirium: acute change with long-term implications. *Anesth Analg* 2011; **112**: 1202–11

31 Radtke MF, Hagemann L, Seeling M, Wernecke KD, Spies CD. Risk factors for inadequate emergence after anesthesia: emergence delirium and hypoactive emergence. *Minerva Anestesiol* 2010; **76**: 394–403

Handling editor: A. R. Absalom