Recognition of Delirium in Postoperative Elderly Patients: A Multicenter Study

Tianne Numan, MSc,* Mark van den Boogaard, PhD,† Adriaan M. Kamper, MD, PhD,‡ Paul J.T. Rood, MSc,† Linda M. Peelen, PhD,*§ and Arjen J.C. Slooter, MD, PhD* for the Dutch Delirium Detection Study Group[1]

OBJECTIVES: To evaluate to what extent delirium experts agree on the diagnosis of delirium when independently assessing exactly the same information and to evaluate the sensitivity of delirium screening tools in routine daily practice of clinical nurses.

DESIGN: Prospective observational longitudinal study.

SETTING: Three medical centers in the Netherlands.

PARTICIPANTS: Elderly postoperative adults (n = 167).

MEASUREMENTS: A researcher examined participants daily (Postoperative Day 1–3) for delirium using a standardized cognitive assessment and interview including the Delirium Rating Scale Revised-98 as global impression without any cut-off values that was recorded on video. Two delirium experts independently evaluated the videos and clinical information from the last 24 hours in the participants' record and classified each assessment as delirious, possibly delirious, or not delirious. Interrater agreement between the delirium experts was determined using weighted Cohen's kappa. When there was no consensus, a third expert was consulted. Final classification was based on median score and compared with the results of the Confusion Assessment Method for Intensive Care Unit and Delirium Observation Scale that clinical nurses administered.

RESULTS: Four hundred twenty-four postoperative assessments of 167 participants were included. The overall kappa was 0.61 (95% confidence interval = 0.53–0.68). There was no agreement between the experts for 89 (21.0%) assessments and a third delirium expert was needed for the final classification. Delirium screening that nurses performed detected 32% of the assessments that the experts diagnosed as (possibly) delirious.

CONCLUSION: There was considerable disagreement in classification of delirium by experts who independently assessed exactly the same information, showing the difficulty of delirium diagnosis. Furthermore, the sensitivity of daily delirium screening by clinical nurses was poor. Future research should focus on development of objective instruments to diagnose delirium. **J Am Geriatr Soc 2017.**

Key words: delirium; encephalopathy; delirium screening tools; postoperative; diagnosis

From the *Department of Intensive Care Medicine and Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht; †Department of Intensive Care Medicine, Radboud University Medical Center Nijmegen, Nijmegen; ‡Section of Geriatrics, Department of Internal Medicine, Isala Hospital, Zwolle; and §Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands.

[1]Members of the Dutch Delirium Detection Study group are listed in Appendix S1.

Address correspondence to Tianne Numan, Department of Intensive Care Medicine, University Medical Center Utrecht, F06.149, PO Box 85500, 3508 GA Utrecht, The Netherlands. E-mail: tiannenuman@gmail.com

DOI: 10.1111/jgs.14933

Delirium is characterized by an acute disturbance of attention and awareness, with additional changes in cognition that tend to fluctuate over time.[1] The incidence of delirium during hospital stay is high, especially in elderly postoperative and critically ill adults.[2–4] Consequences of delirium are longer intensive care unit (ICU) and hospital stays,[5–7] higher healthcare costs,[2,8,9] and deterioration of long-term cognitive function.[4,7,8,10] Therefore, delirium is a serious healthcare problem.

The reference standard for diagnosing delirium is assessment by a delirium expert (geriatrician, psychiatrist, neurologist), usually using the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) criteria.[1] Delirium diagnosis may be based on one or a combination of clinical impression, cognitive testing, clinical notes, physical examination, and laboratory results, which has been found to be differ greatly for the reference standard.[11] The consistency of the diagnosis of delirium is unclear, because to what extent delirium experts agree on a delirium diagnosis when independently assessing the same information has never been studied.

In clinical practice, it is infeasible for delirium experts to examine all patients. Therefore, delirium screening tools have been developed, and several organizations, including the American Geriatrics Society[12] and the Society of Critical

Care Medicine,[13] recommend routine delirium screening. The feasibility and interrater reliability of delirium instruments have been investigated in several studies,[14–18] but almost all of these studies were performed in a research setting, where dedicated researchers performed the assessments, and almost all were performed in the centers where the screening tools were developed. However, the sensitivity of delirium screening in routine, clinical practice by clinical nurses in ICU patients has been found to be low (overall 47% using the Confusion Assessment Method for the ICU (CAM-ICU)),[19] as well as in patients not in the ICU when using the CAM (25%).[20]

The aim of this study was to determine agreement in diagnosing delirium of delirium experts when independently confronted with the same clinical information and to determine the sensitivity of delirium screening tools in routine daily practice by clinical nurses.

METHODS

Design, Setting, and Study Population

This prospective, multicenter observational study was performed in three centers: University Medical Center Utrecht (UMCU); Radboud University Medical Center Nijmegen; and Isala Klinieken Zwolle, a nonacademic teaching hospital (clinicaltrials.gov identifier NCT02404181). The current study was performed in the context of the validation of an electroencephalogram (EEG)-based delirium monitor that the medical ethical board of UMCU approved (protocol number 13–643).

Individuals scheduled for major surgery (expected hospital stay of ≥2 days) aged 60 and older and considered to be at risk of delirium were informed about the study and provided written informed consent. Patients were considered to be at risk of delirium if they were aged 75 and older or had a history of transient ischemic attack (TIA) stroke, alcohol abuse, depression, dementia, delirium, or cognitive problems, severe cardiac or respiratory disease.[2] Exclusion criteria were inability to speak Dutch or English, deafness, neurological surgery, and inability to perform EEG-based delirium monitor recording.

Reference Standard Delirium Diagnosis

Trained researchers examined participants preoperatively (T-1) and during the first three consecutive days after surgery (T1, T2, T3). Participants were interviewed using the 13 severity items of the Delirium Rating Scale Revised 1998 Edition (DRS-R-98)[21,22] and the CAM-ICU.[15] The DRS-R-98 is designed for psychiatrists or other professionals with substantial training in mental health assessment (Appendix S2). Items were rated between 0 and 3, yielding a total DRS-R-98 score ranging from 0 to 39, with higher scores indicating more-severe delirium symptoms. No predefined cut-off values were used to ensure unbiased delirium diagnosis by the experts. Interviews took approximately 10 to 15 minutes, and all interviews including a cognitive assessment were recorded on video. When participants refused to participate on a specific recording day, no video assessment was performed, and no formal expert diagnosis was available for that day.

Delirium experts from academic and nonacademic hospitals in the Netherlands evaluated the videos. The experts were 17 psychiatrists, 15 geriatricians, four neurologists, one neuropsychologist, and one nurse–scientist; had a median of 11.5 years (interquartile range (IQR) 6.3–19.0 years) of experience; and saw an estimated median of 15 individuals with delirium monthly (IQR 10–25). All experts received written instruction about the evaluation procedure. The experts evaluated all video recordings, read participants' medical and nursing record information for the last 24 hours, and filled in the DRS-R-98. In different combinations, two delirium experts evaluated the patients. The delirium experts were blinded to each others evaluations.

Two classifications had to be made: one based solely on the cognitive testing recorded on video and one based on all information from the last 24 hours, including the description in the medical and nursing files and the video. Both classifications were based on DSM-5 criteria. For each classification, a likelihood of the participant having delirium had to be reported on a numeric rating scale (NRS: 0 = definite no delirium, 10 = definite delirium), as well as a final diagnosis: no delirium, possible delirium, or delirium. When the participant was classified as having (possible) delirium, the motor subtype (hypoactive, hyperactive, mixed) was reported.

Classifications of the two delirium experts were compared. If there was no consensus on the video recording or all information from the last 24 hours including the video recording, a third expert was consulted. The third expert was blinded to the classification of the first two experts but knew that he or she was consulted because of disagreement. Final classification and motor subtype were based on the median score of the three experts. The final NRS and DRS-R-98 scores of the two or three (when a third expert was consulted) delirium experts were averaged.

Delirium Screening by the Clinical Nurse

In the participating hospitals, the CAM-ICU and Delirium Observation Screening (DOS) scale were used in daily practice for delirium assessment two or three times daily, in the ICU and general wards, respectively. The DOS scale consists of 13 items (Appendix S2).[17,23] Participants' records were reviewed for available documentation of CAM-ICU or DOS score for the 24 hours before the video recording. In participants transferred from the ICU to the general ward, both scores could have been administered in the last 24 hours. The highest DOS score per day was used; when positive and negative CAM-ICU scores were documented, the positive score was used. A combined clinical nurse score was calculated and defined positive when one or both CAM-ICU and DOS scores were positive. Clinical nurses and physicians were blinded to the classification of the delirium experts.

Other Data Collection

The following baseline characteristics were recorded: age, sex, alcohol consumption (0, 1–14, >14 alcohol units per

week), transient ischemic attack (TIA) or stroke in the medical history, preexisting psychiatric disease, and previous cognitive problems (according to preoperative Mini-Mental State Examination (MMSE) score). Furthermore, characteristics of the surgical procedure were documented, including type and duration of surgery.

Statistical Analysis

Linear weighted Cohen's kappa statistic was used separately for the classification based on cognitive testing on video only and for the classification based on all information over the last 24 hours including cognitive testing on video to assess interrater agreement between the two experts for all postoperative assessments. To assess reliability, the intraclass correlation coefficient was calculated, based on a two-way random single measures model, between the ratings of the two delirium experts for the likelihood (NRS) and severity (DRS-R-98) scores.

Furthermore, variability in NRS and DRS-R-98 scores between the two delirium experts was evaluated for the motor subtypes by calculating the average and absolute difference of the NRS and DRS-R-98 scores between the initial two experts. These scores were compared between the different motor subtypes using the Mann-Whitney test.

Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the CAM-ICU, DOS, and the combined clinical nurse score were calculated using the classification of delirium experts based on all information over the last 24 hours including cognitive testing as reference. For these calculations, possible delirium and delirium were combined and compared with no delirium. To assess the effect of including possible delirium in the delirium group, a sensitivity analysis was performed in which possible delirium was included in the group without delirium.

All postoperative assessments were included as independent observations. To assess the influence of repeated measures, the diagnostic values of the CAM-ICU scored by clinical nurses based on T1 only were recalculated. $P < .05$ was assumed to be statistically significant. Statistical analyses were performed in SPSS version 21 (IBM Corp., Armonk, NY).

RESULTS

Of 196 individuals from whom informed consent was obtained, 29 (14.8%) were excluded for logistical reasons. The remaining study population consisted of 167 participants, in whom 424 postoperative delirium assessments were performed. Of the 501 potential postoperative assessments, 31 (6.2%) were excluded because the individual refused participation on that day and 46 (9.2%) for logistical reasons, for example discharge. No participant was intubated at the time of cognitive testing.

Population Characteristics

Baseline characteristics of the study population are presented in Table 1. One hundred forty-seven participants (88.6%) underwent cardiothoracic surgery. The median preoperative MMSE score was 28 (range 14–30); 10

participants had an MMSE score of less than 24. Assessments were performed in the ICU (n = 71, 16.7%), medium care or coronary care-unit (n = 93, 21.9%), and general ward (n = 260, 61.3%).

Table 1. Baseline Characteristics of Study Population (N = 167)

Characteristic	Value
Age, mean ± standard deviation	76.8 (6.4)
Male, n (%)	113 (67.7)
Alcohol consumption, drinks/wk, n (%)	
0	64 (38.3)
1–14	73 (43.7)
>14	26 (15.6)
Medical history, n (%)	
Transient ischemic attack or stroke	47 (28.1)
Psychiatric illness	11 (6.6)
MMSE score, median (range)	28 (14–30)
Surgery type, n (%)	
Cardiothoracic or vascular	147 (88.6)
Orthopedic	12 (7.2)
Other[a]	7 (4.2)
Duration of surgery, minutes, median (range)	169 (20–401)

Data were missing for alcohol consumption (n = 4, 2.4%), Mini-Mental State Examination (MMSE) score (n = 43, 25.7%), duration of surgery (n = 4, 4.2%).
[a]Otorhinolaryngological (n = 3), gastrointestinal or general (n = 3), urological (n = 1).

Table 2. Interrater Variability of Classification of Delirium According to Delirium Experts

A	Classification: *cognitive tests only* (422 assessments*)	Expert X		
		No delirium	Possible delirium	Delirium
Expert Y	No delirium	313 (74.2%)	–	–
	Possible delirium	45 (10.7%)	19 (4.5%)	–
	Delirium	9 (2.1%)	24 (5.7%)	12 (2.8%)

B	Classification: *all information in last 24 hours* (424 assessments)	Expert X		
		No delirium	Possible delirium	Delirium
Expert Y	No delirium	286 (67.5%)	–	–
	Possible delirium	49 (11.6%)	16 (3.8%)	–
	Delirium	11 (2.6%)	29 (6.8%)	33 (7.8%)

In each patient the two initial experts classified each participant as having no delirium, possible delirium, or delirium based on cognitive tests recorded on video (weighted kappa = 0.53, 95% confidence interval (CI) = 0.44–0.62) and on all information in the last 24 hours, including the cognitive testing recorded on video (weighted kappa = 0.61, 95% CI = 0.53–0.68). The videos of two postoperative assessments were incomplete and therefore not included in this comparison.

Table 3. Absolute Difference in Likelihood and Severity of Delirium Between Two Initial Experts

Absolute Difference	Hypoactive (29 Assessments)	Mixed (58 Assessments)	Hyperactive (13 Assessments)	P-Value
Likelihood of delirium (Numeric Rating Scale)[a]	3.0 (1.5–5.0)	2.0 (1.0–3.3)	1.0 (0–3.5)	.01[c,d]
Severity of delirium (Delirium Rating Scale Revised)[b]	3.0 (1.0–8.0)	3.0 (1.8–5.3)	2.0 (0.5–5.0)	.45

Motor subtype classification based on two (or majority vote of three) delirium experts based on all available information over last 24 hours, including cognitive testing recorded on video. The likelihood and severity of delirium were assessed using the Kruskal Wallis test.
[a]Average score between 0 and 10 for likelihood of an individual having delirium.
[b]Average score between 0 and 39 for severity of delirium.
[c]Significant difference between hypoactive and mixed type delirium (post-hoc Mann-Whitney U test).
[d]Significant difference between hypoactive and hyperactive delirium (post-hoc Mann-Whitney U test).

Interrater Reliability of Diagnosis of Delirium

Of the 424 assessments, both experts classified 344 (81.5%) similar based on cognitive testing only (Table 2); two observations were excluded because of an incomplete video. In 78 (18.5%) assessments, there was disagreement between the experts on the classification of video-recorded cognitive testing ($\kappa = 0.53$, 95% confidence interval (CI) = 0.44–0.62).

The extent of agreement between experts classifying delirium based on all information over the last 24 hours including cognitive testing resulted in an inter-rater agreement of $\kappa = 0.61$, (95% CI = 0.53–0.68) (Table 2). In the 89 (21.0%) assessments with disagreement based on all information, a third expert was consulted to obtain final classification.

Based on all information including cognitive testing recorded on video, 50 (11.8%) postoperative assessments were classified as delirious and 50 (11.8%) as possibly delirious, which were combined as 100 (23.6%) positive delirious assessments. The other 324 (76.4%) assessments were classified as not delirious. This corresponded with 31 (18.6%) participants classified as delirious and 30 (18.0%) as possibly delirious in the first three postoperative days. Median DRS-R-98 score was 3.0 (IQR 2.0–4.0) participants without delirium, 7.7 (IQR 6.3–12.7) for those with possible delirium, and 12.7 (IQR 9.6–17.5) for those with delirium. Futher analyses were performed using the final delirium expert classification based on all information over the last 24 hours.

Variability of Likelihood and Severity of Delirium

The intraclass correlation coefficient between the NRS and DRS-R-98 scores of the first two delirium experts was 0.73 (95% CI = 0.68–0.77) for NRS and 0.76 (95% CI = 0.72–0.80) for DRS-R-98 (Figure 1). The median absolute difference between both experts was 1 for NRS (IQR 0–2) and 1 for DRS-R-98 (IQR 1–3) scores. Agreement between the first two experts on the individual DRS-R-98 items is shown in Table S1.

The NRS scores that the experts assigned for the likelihood of the individuals having delirium did not differ between the hypoactive (median 6.3, IQR 4.4–7.2), mixed (median 6.2, IQR 4.6–9.0), and hyperactive (median 7.0, IQR 5.5–9.0) ($P = .18$) subtypes of delirium. The DRS-R-

98 that the experts assigned for severity of delirium did not differ between the three motor subtypes either (hypoactive: median 10.0, IQR 7.8–13.3; mixed: median 9.7, IQR 6.9–12.2; hyperactive: median 13.0, IQR 6.5–20.3) ($P = .33$).

There was more disagreement between the experts on the likelihood of the individuals having delirium in individuals with hypoactive delirium than in those with the other types of delirium; the absolute difference in NRS score between the initial two experts was significantly higher with the hypoactive subtype than with mixed ($P = .01$) or hyperactive ($P = .03$) subtype. No differences in assigned DRS-R-98 scores were found between individuals with different delirium subtypes ($P = .45$).

Clinical Use of Delirium Screening Tools

Of the 424 included assessments, only the CAM-ICU was performed in 117 (27.6%), only the DOS was performed in 87 (20.5%), both the CAM-ICU and DOS were performed in 14 (3.3%), and no clinical screening was documented in 206 (48.6%). The combined score (CAM-ICU and/or DOS) was performed in 218 (51.4%) delirium assessments). In 32 of the 100 delirium assessments that the experts classified as (possible) delirium, the CAM-ICU or DOS was positive. Screening by clinical nurses was performed in 44.4% of those without delirium, 60.0% of those with possible delirium, and 82.0% of those with delirium.

The CAM-ICU was performed in 44 assessments with (possible) delirium classification by delirium experts, and was positive in 11 (25.0%) of these assessments. The DOS was documented in 37 (possible) delirium assessments and was positive in 23 (62.2%) of the assessments (Table 4). Combining the two delirium screening tools resulted in sensitivity of 43.2% (95% CI = 31.6–55.2%) and specificity of 97.2% (95% CI = 92.5–99.1%). The diagnostic performance of the clinical nurses with screening improved slightly when possible delirium was grouped with no delirium (sensitivity 55.8%, specificity 93.1%, PPV 66.7%, NPV 89.6%).

Finally, calculating the diagnostic performance of the CAM-ICU for T1 only, to exclude effects of repeated measurements, resulted in sensitivity of 21.9% (95% CI = 9.9–40.4%), specificity of 95.7% (95% CI = 87.2–98.9%), PPV of 70.0% (95% CI=35.4–91.9%), and NPV of 72.8% (95% CI = 62.4–81.3%), which indicated no major differences from the main analysis over all postoperative days.
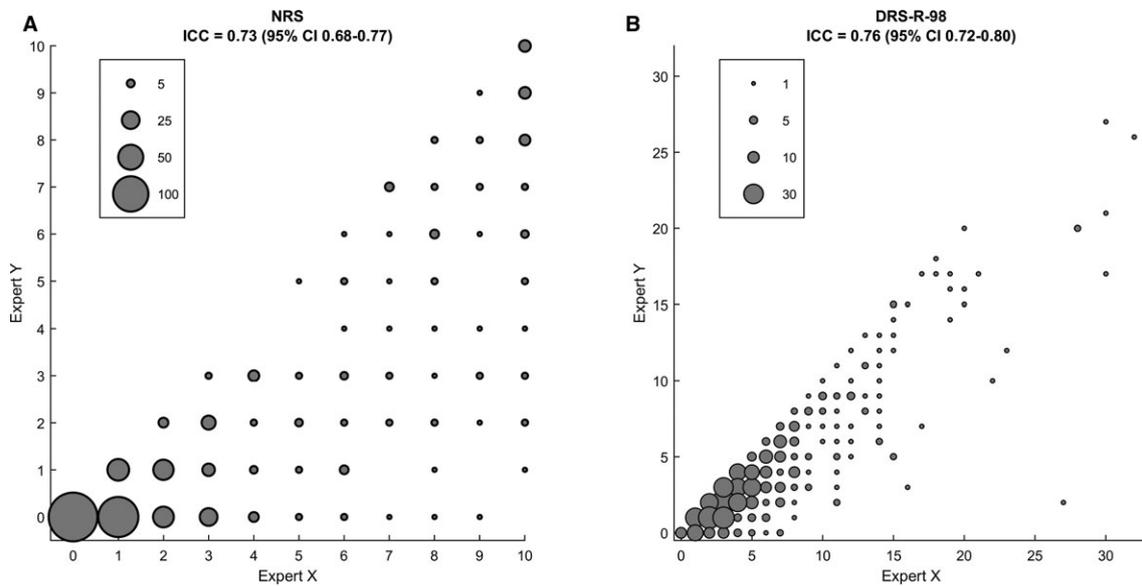
Figure 1. Scores of two delirium experts of the likelihood of delirium (numeric rating scale (NRS, left) and the severity of delirium (Delirium Rating Scale Revised 1998 Edition (DRS-R-98), right) and likelihood of delirium (numeric rating scale (NRS, right) based on all information within the last 24 hours including cognitive testing as recorded on video. Expert X had the highest NRS or DRS-R-98 score of the two experts, Expert Y had the lowest NRS or DRS-R-98 score of the two experts. The size of the circles indicates the number of postoperative assessments with these scores.

Table 4. Evaluation of Clinical Nurse Delirium Screening in Clinical Practice

Screen	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
	% (95% Confidence Interval)			
CAM-ICU (131 assessments, 33.6% (possibly) delirious[a])	25.0 (13.7–40.6)	96.6 (89.5–99.1)	78.6 (48.8–99.1)	71.8 (62.6–79.5)
DOS (101 assessments, 36.6% (possibly) delirious[a])	62.2 (44.8–77.1)	98.4 (90.4–99.9)	95.8 (90.4–99.9)	81.8 (71.0–89.4)
Combined CAM-ICU and DOS (218 assessments, 34.4% (possibly) delirious[a])	43.2 (31.6–55.2)	97.2 (92.5–99.1)	88.9 (73.0–96.4)	76.9 (70.0–82.7)

[a]Percentage of assessments classified as possibly delirious or delirious according to the final classification based on all information in the last 24 hours, including cognitive testing stored on video. In 14 assessments, Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) and Delirium Observation Screening (DOS) were documented, which are included in all three groups. DOS score ≥3 was used as the threshold for delirium detection. For combined CAM-ICU and DOS, the highest score (positive for CAM-ICU or DOS ≥ 3) documented in the last 24 hours.

DISCUSSION

This study found a considerable amount of disagreement in the classification of delirium by experts who independently assessed the same information. Variability between experts in the likelihood of patients having delirium was highest in those with the hypoactive motor subtype. Recognition of delirium by clinical nurses using screening tools appeared to be insufficient, with not more than 32% of the postoperative cases identified.

Recognition of Delirium in Clinical Practice

The CAM-ICU and DOS have been validated in research settings with good diagnostical properties,[15,17,24,25] but not all subsequent studies were able to reproduce these high diagnostic values.[14,19,26] The current study had a sensitivity of 25.0% for CAM-ICU and 62.2% for DOS, indicating underdetection of delirium in clinical practice. Cautious interpretation is needed, because cases were included only when the results of delirium screening by clinical nurses was documented in the medical record, which were present in half of the study population. Underlying reasons of the relative low number of documented screening results were not investigated. It is not clear how this might have influenced the diagnostic values. Another explanation for the poor diagnostic performance is the included 'possible delirium' assessments, which might not fullfil all criteria and therefore result in false-negative scores. Repeating the analysis with possible delirium as a negative assessment resulted in a slightly higher sensitivity (55.8%), although that is still not sufficient for clinical practice.

Strengths and Limitations

Strengths of this study are the prospective, multicenter setting in academic and nonacademic hospitals and the large sample size. The first assessment was performed before surgery, facilitating the delirium expert in the classification of delirium. Cognitive testing was extensive and

standardized. Because these assessments were recorded on video, there was a unique opportunity to investigate to what extent there is agreement on a diagnosis of delirium between different delirium experts truly independent of each other. The reference standard was robust, because it was based on all available information compatible with usual clinical practice and not on a DRS-R-98 cut-off score. Moreover, consensus of two delirium experts was required and, in the case of disagreement, a third expert was consulted. To account for the fluctuating nature of delirium, experts had access to the medical and nursing files in addition to the videos with the interview including cognitive assessment.

Some potential limitations need to be addressed. First, the possible delirium may have been difficult to interpret and may have influenced interrater agreement. Second, each postoperative assessment was analyzed as an individual case, and therefore the effect of repeated measures within participants was not taken into account, but sensitivity analyses evaluating the diagnostic values of the CAM-ICU based on one measure of each participant (T1) showed similar results. Third, 31 (6.2%) formal assessments could not be performed because the patient refused to participate, which may have led to underdetection of delirium, although the authors do not recall cases that showed signs of delirium based on clinical impression of the visiting researcher or the medical record. Finally, agreement between the experts and detection rates by clinical nurses might have been higher when the included individuals had more-severe delirium.

Recommendations for Future Studies

There is a clear need for an easy-to-use, reliable method of detecting delirium in clinical practice. Most of the current tools are based on observation, simple tests,[27] or a combination of both approaches.[15] Several cognitive tests have been described and validated in the literature, such as the abbreviated cognitive test for delirium,[28] Edinburgh Delirium test box,[29,30] or smartphone application.[31] These tools showed good performance, but interpretation by the performer is needed, and interaction with patients is essential but is not possible in all individuals, for example because of low levels of consciousness. Therefore, an objective tool is needed that is applicable in every individual and unambiguous to interpret. Bipolar EEG may be a candidate for objective delirium detection.[32] One-minute recording without artifacts was shown to be sufficient to distinguish individuals with delirium from those without,[32] although that study was performed in individuals who definitely had delirium or did not, so findings should be confirmed in an independent study population. It is not clear whether this approach could replace assessment of all aspects of delirium or whether some features (e.g., hallucinations or distress) would still require interaction.

CONCLUSION

This study showed a considerable amount of disagreement in the classification of delirium by experts who independently assessed the same information, indicating the difficulty of delirium diagnosis. Moreover, delirium recognition in clinical practice by clinical nurses was poor, and many cases of delirium were unrecognized despite the use of delirium screening tools. An objective and reliable screening tool may improve detection of delirium.

REFERENCES

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Ed. Arlington, VA: American Psychiatric Association, 2013.
2. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. Lancet 2014;383:911–922.
3. Koster S, Hensens AG, Schuurmans MJ, et al. Risk factors of delirium after cardiac surgery. Eur J Cardiovasc Nurs 2011;10:197–204.
4. Wolters AE, van Dijk D, Pasma W, et al. Long-term outcome of delirium during intensive care unit stay in survivors of critical illness: A prospective cohort study. Crit Care 2015;18:R125.
5. Gleason LJ, Schmitt EM, Kosar CM, et al. Effect of delirium and other major complications on outcomes after elective surgery in older adults. JAMA Surg 2015;150:1134–1140.
6. Salluh JIF, Wang H, Schneider EB, et al. Outcome of delirium in critically ill patients: Systematic review and meta-analysis. BMJ 2015;350:h2538.
7. Klein Klouwenberg PMC, Zaal IJ, Spitoni C, et al. The attributable mortality of delirium in critically ill patients: Prospective cohort study. BMJ 2014;349:1–10.
8. Rizk P, Morris W, Oladeji P, et al. Review of postoperative delirium in geriatric patients undergoing hip surgery. Geriatr Orthop Surg Rehabil 2016;7:100–105.
9. Leslie DL, Marcantonio ER, Zhang Y, et al. One-year health care costs associated with delirium in the elderly population. Arch Intern Med 2008;168:27–32.
10. Fong TG, Davis D, Growdon ME, et al. The interface between delirium and dementia in elderly adults. Lancet Neurol 2015;14:823–832.
11. Neufeld KJ, Nelliot A, Inouye SK, et al. Delirium diagnosis methodology used in research: A survey-based study. Am J Geriatr Psychiatry 2014;22:1513–1521.
12. American Geriatrics Society Expert Panel on Postoperative Delirium in Older Adults. Postoperative delirium in older adults: Best practice statement from the American Geriatrics Society. J Am Geriatr Soc 2015;220:142–150.
13. Barr J, Fraser GL, Puntillo K, et al. Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit. Crit Care Med 2013;41:263–306.
14. Van Velthuijsen EL, Zwakhalen SMG, Warnier RMJ, et al. Psychometric properties and feasibility of instruments for the detection of delirium in older hospitalized patients: A systematic review. Int J Geriatr Psychiatry 2016;31:974–989.
15. Ely EW, Margolin R, Francis J, 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;29:1370–1379.

16. Scheffer AC, van Munster BC, Schuurmans MJ, et al. Assessing severity of delirium by the Delirium Observation Screening Scale. Int J Geriatr Psychiatry 2011;26:284–291.
17. Van Gemert LA, Schuurmans MJ. The Neecham Confusion Scale and the Delirium Observation Screening Scale: Capacity to discriminate and ease of use in clinical practice. BMC Nurs 2007;6:3.
18. De J, Wand APF. Delirium screening: A systematic review of delirium screening tools in hospitalized patients. Gerontologist 2015;55:1079–1099.
19. Van Eijk MM, van den Boogaard M, van Marum RJ, et al. Routine use of the confusion assessment method for the intensive care unit: A multicenter study. Am J Respir Crit Care Med 2011;184:340–344.
20. Rice KL, Bennett M, Gomez M, et al. Nurses' recognition of delirium in the hospitalized older adult. Clin Nurse Spec 2011;25:299–311.
21. Trzepacz PT, Mittal D, Torres R, et al. Validation of the delirium rating scale-revised-98: Comparison with the delirium rating scale and the cognitive test for delirium. J Neuropsychiatry Clin Neurosci 2001;13:229–242.
22. De Rooij SE, van Munster BC, Korevaar JC, et al. Delirium subtype identification and the validation of the Delirium Rating Scale—Revised-98 (Dutch version) in hospitalized elderly patients. Int J Geriatr Psychiatry 2006;21:876–882.
23. Neto AS, Nassar AP, Cardoso SO, et al. Delirium screening in critically ill patients: A systematic review and meta-analysis. Crit Care Med 2012;40:1946–1951.
24. Koster S, Hensens AG, Oosterveld FGJ, et al. The delirium observation screening scale recognizes delirium early after cardiac surgery. Eur J Cardiovasc Nurs 2009;8:309–314.
25. Ely EW, Inouye SK, Bernard GR, et al. Delirium in mechanically ventilated patients: Validity and reliability of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). JAMA 2001;286:2703–2710.
26. van Eijk MMJ, van Marum JR, Klijn IAM, et al. Comparison of delirium assessment tools in a mixed intensive care unit. Crit Care Med 2009;37:1881–1885.
27. Adamis D, Meagher D, Murray O, et al. Evaluating attention in delirium: A comparison of bedside tests of attention. Geriatr Gerontol Int 2016;9:1028–1035.
28. Hart RP, Best AM, Sessler CN, et al. Abbreviated cognitive test for delirium. J Psychosom Res 1997;43:417–423.
29. Brown LJE, Fordyce C, Zaghdani C, et al. Detecting deficits of sustained visual attention in delirium. J Neurol Neurosurg Psychiatry 2011;82:1334–1340.
30. Tieges Z, McGrath A, Hall RJ, et al. Abnormal level of arousal as a predictor of delirium and inattention: An exploratory study. Am J Geriatr Psychiatry 2013;21:1244–1253.
31. Tieges Z, Stiobhairt A, Scott K, et al. Development of a smartphone application for the objective detection of attentional deficits in delirium. Int Psychogeriatrics 2015;27:1251–1262.
32. Van der Kooi AW, Zaal IJ, Klijn FA, et al. Delirium detection using EEG: What and how to measure. Chest 2015;147:94–101.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Dutch Delirium Detection Study Group
Appendix S2. Screening Tools – Delirium Rating Scale Revised 1998 Edition (DRS-R-98) and Delirium Observation Screening (DOS) Scale
Table S1. Agreement Between Two Experts on Individual Delirium Rating Scale Revised 98 Edition Items (n=424 Assessments)

Please note: Wiley-Blackwell is not responsible for the content, accuracy, errors, or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.