


# The Delirium Interview as a new reference standard in studies on delirium assessment tools

Fienke L. Ditzel MD<sup>1</sup>  | Arjen J. C. Slooter MD, PhD<sup>1,2,3</sup> |  
Mark van den Boogaard RN, PhD<sup>4</sup> | Michel Boonstra MSc<sup>1</sup> |  
Timotheus A. van Nesselrooij Msc<sup>2</sup> | Marjan Kromkamp MD, PhD<sup>2</sup> |  
Monica Pop-Purceleanu MD, PhD<sup>5</sup> | Paul J. T. Rood RN, PhD<sup>4,6</sup> |  
Robert Jan Osse MD, PhD<sup>7</sup> | Carol K. Chan MBBCh, MSc<sup>8</sup> |  
Alasdair M. J. MacLulich MRCP, PhD<sup>9</sup> | Zoë Tieges PhD<sup>9,10</sup> |  
Karin J. Neufeld MD, MPH<sup>8</sup> | Suzanne C. A. Hut PhD<sup>1</sup>

<sup>1</sup>Department of Intensive Care Medicine and UMC Utrecht Brain Center, University Medical Center Utrecht, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

<sup>2</sup>Department of Psychiatry and UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands

<sup>3</sup>Department of Neurology, UZ Brussel and Vrije Universiteit Brussel, Brussels, Belgium

<sup>4</sup>Department of Intensive Care Medicine, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, the Netherlands

<sup>5</sup>Department of Psychiatry, Radboud University Medical Center, Nijmegen, the Netherlands

<sup>6</sup>School of Health Studies, Research Department of Emergency and Critical Care, HAN University of Applied Sciences, Nijmegen, the Netherlands

<sup>7</sup>Department of Psychiatry, Erasmus Medical Centre Rotterdam, Rotterdam, the Netherlands

<sup>8</sup>Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

## Abstract

**Background:** The reference standard in studies on delirium assessment tools is usually based on the clinical judgment of only one delirium expert and may be concise, unstandardized, or not specified at all. This multicenter study investigated the performance of the Delirium Interview, a new reference standard for studies on delirium assessment tools allowing classification of delirium based on written reports.

**Methods:** We tested the diagnostic accuracy of our standardized Delirium Interview by comparing delirium assessments of the reported results with live assessments. Our reference, the live assessment, was performed by two delirium experts and one well-trained researcher who registered the results. Their delirium assessment was compared to the majority vote of three other independent delirium experts who judged the rapportage of the Delirium Interview. Our total pool consisted of 13 delirium experts with an average of  $13 \pm 8$  years of experience.

**Results:** We included 98 patients (62% male, mean age  $69 \pm 12$  years), of whom 56 (57%) intensive care units (ICUs) patients, 22 (39%) patients with a Richmond Agitation Sedation Scale (RASS)  $< 0$  and 26 (27%) non-verbal assessments. The overall prevalence of delirium was 28%. The Delirium Interview had a sensitivity of 89% (95% confidence interval [CI]: 71%–98%) and specificity of 82% (95% CI: 71%–90%), compared to the diagnosis of an independent panel of two delirium experts and one researcher who examined the patients themselves. Negative and positive predictive values were 95% (95% CI:

<sup>9</sup>Edinburgh Delirium Research Group, Ageing and Health, Usher Institute, University of Edinburgh, Edinburgh, UK

<sup>10</sup>SMART Technology Centre, School of Computing, Engineering and Built Environment, Glasgow Caledonian University, Glasgow, UK

#### Correspondence

Fienke L. Ditzel, Department of Intensive Care Medicine, University Medical Center Utrecht, PO Box 85500 3508 GA, Utrecht, the Netherlands.

Email: [f.l.ditzel-2@umcutrecht.nl](mailto:f.l.ditzel-2@umcutrecht.nl)

#### Funding information

Horizon 2020 Framework Programme, Grant/Award Number: 820555

86%–0.99%), respectively, 66% (95% CI: 49%–80%). Stratification into ICU and non-ICU patients yielded similar results.

**Conclusion:** The Delirium Interview is a feasible reference method for large study cohorts evaluating delirium assessment tools since experts could assess delirium with high accuracy without seeing the patient at the bedside.

#### KEYWORDS

cognitive assessment, delirium, neuropsychological tests, reference standard

## INTRODUCTION

Delirium is a neuropsychiatric syndrome characterized by disturbances of awareness, attention, cognitive functions, and behavior.<sup>1</sup> It occurs in the severely ill, often older patients<sup>2</sup> and is associated with prolonged hospitalization,<sup>3</sup> institutionalization, increased costs, long-term cognitive decline, and mortality.<sup>2–4</sup> Recognition of delirium is essential for the treatment of underlying conditions, optimal communication with affected patients and informing family members.<sup>5</sup> In the previous decade, several delirium assessment tools have been developed.<sup>6</sup> However, validating these instruments can be subject to various methodological issues.

The reference standard in studies on delirium assessment tools is usually based on the clinical judgment of only one delirium expert and may be concise, unstandardized, or not specified.<sup>6</sup> The expert relies on diagnostic criteria, most often the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) criteria,<sup>7</sup> but it is unclear which tests should be used to evaluate these criteria and which cut-off values should be used to interpret these tests.<sup>8</sup> In the absence of standardized testing, experts often base their diagnosis on unspecified clinical judgment.<sup>9</sup> Importantly, even if experts had access to exactly the same clinical information, they were found to disagree on delirium classification in 21% of the cases.<sup>10</sup> To minimize subjectivity and correct for differences in the level of experience, a panel of experts is needed to classify delirium by majority vote.<sup>9</sup> These experts must examine the patient simultaneously due to the fluctuating character of delirium and rule out learning effects associated with repeated measurements. However, it appears to be logistically challenging to arrange for two (or more) experienced delirium experts to be present simultaneously, especially when many patients need to be enrolled. Therefore, studies that

### Key points

- This paper proposes the Delirium Interview as a new reference for delirium assessment tools used in intensive care unit (ICU) and non-ICU departments.
- The diagnosis was based on the majority vote of an expert panel of three, accounting for the known disagreement on the diagnosis of delirium.
- Our panel could assess delirium with a high sensitivity of 89% (95% confidence interval [CI]: 71%–98%) and specificity of 82% (95% CI: 71%–90%) without being present at the bedside.

### Why does this paper matter?

Delirium is a major healthcare problem associated with longer hospital duration and increased healthcare costs. Recognition of delirium is essential for the treatment of underlying conditions and optimal communication with affected patients. The Delirium Interview is a reference method standardized for a heterogeneous population, including patients with low consciousness and intubated patients. The methodology is logistically efficient for studies with large cohorts, as we found that delirium experts could diagnose delirium with high accuracy without being present at the bedside.

validate delirium assessment tools are mostly based on small numbers and use a self-selected reference method that is often poorly described.<sup>6,9</sup>

**TABLE 1** The DSM-5 criteria and the items of the Delirium Interview.

DSM-5 Criteria	Delirium Interview item
A. Disturbance in attention (i.e., reduced ability to direct, focus, sustain, and shift attention) and awareness	RASS, <sup>12</sup> OSLA, <sup>13</sup> ASE (item 2 of the CAM-ICU), <sup>14</sup> MOTYB (item 3 of the 4AT) <sup>15</sup>
B. The disturbance develops over a short period of time (usually hours to a few days), represents a change from baseline attention and awareness, and tends to fluctuate in severity during the course of a day.	Review of EHR 24 h before- and 12 h after the delirium assessment.
C. An additional disturbance in cognition (e.g., memory deficit, disorientation, language, visuospatial ability, or perception).	Orientation (person, place and time), memory (item 4 and 5 of the CTD), <sup>16</sup> conceptual reasoning (item 7 of the CTD), <sup>16</sup> visuospatial ability (inspired by item 13 of the DRS-R-98), <sup>17</sup> comprehension (item 8 of the MMSE) <sup>18</sup>
D. The disturbances in Criteria A and C are not explained by another pre-existing, established, or evolving neurocognitive disorder and do not occur in the context of a severely reduced level of arousal, such as coma	To fulfill this criterion, we adhered to the following exclusion criteria: Acute macro-brain injury within 6 weeks prior to the measurement (including post-anoxic encephalopathy and traumatic brain injury), admission because of a primary neurological or neurosurgical disease or pre-existing dementia as documented in the EHR.
E. There is evidence from the history, physical examination, or laboratory findings that the disturbance is a direct physiological consequence of another medical condition, substance intoxication or withdrawal (i.e., due to a drug of abuse or to a medication), or exposure to a toxin, or is due to multiple etiologies.	To fulfill this criterion, we adhered to the following inclusion criteria: admission to the ICU or non-ICU department
Additional information	Self-report experiences of sleep assessment, hallucinations and paranoia

Abbreviations: 4AT, 4 “A”s Test; ASE, Attention Screening Examination; CAM-ICU, Confusion assessment method for the Intensive Care Unit; CTD, Cognitive Test for Delirium; DRS-R98, Delirium Rating Scale Revised-98; EHR, Electronic Health Record; MMSE, Minimal Mental State Examination; MOTYB, Month of the Year Backwards; OSLA, Observational Scale of Level of Arousal; RASS: Richmond Agitation Sedation Scale.

One way to solve the problem of expert availability is by making video recordings of patients for later classification.<sup>10</sup> However, video recordings could be burdensome and intimidating for a very ill, vulnerable patient who cannot provide informed consent at the moment of assessment. To overcome challenges pertaining to video-recorded interviews while recognizing the need for an efficient method, we composed the Delirium Interview.

The Delirium Interview is a standardized cognitive test battery capturing all aspects of delirium described by the DSM-5 criteria. It was composed of the input of specialist experts (e.g., geriatricians, psychiatrists) and pre-existing cognitive tests and can be applied to either mechanically ventilated intensive care unit (ICU) or non-ICU patients. This study aimed to investigate to what extent delirium experts can diagnose delirium based on written reports of the Delirium Interview without the need to be present at the bedside.

## METHODS

### Setting and study population

This study is part of the DELTA study, a multicenter clinical validation study of DeltaScan for assessing delirium

in the ICU and non-ICU wards. The study design was approved by the medical ethical committee of the University Medical Center Utrecht, the Netherlands (protocol 17–857), which waived the need for informed consent, and registered at [clinicaltrials.gov](https://clinicaltrials.gov) (NCT03966274). This manuscript adheres to the Standards for Reporting of Diagnostic Accuracy Studies (STARD) guidelines.<sup>11</sup> This study recruited patients between February 2019 and December 2021 from the University Medical Center Utrecht or the Radboud University Medical Center Nijmegen in the Netherlands. Patients were included when they were admitted to either the ICU (aged 18 years or older) or admitted to the non-ICU ward (aged 60 years or older), and at a consciousness level of a Richmond Agitation Sedation Scale (RASS)<sup>12</sup> of  $-2$  or higher. Exclusion criteria were: acute brain injury within 6 weeks prior to the measurement, including post-anoxic encephalopathy and traumatic brain injury; admission because of a primary neurological or neurosurgical disease; a condition hampering delirium assessment (e.g., language barrier or deafness); and pre-existing dementia as documented in the Electronic Health Record (EHR). Patients gave permission before study inclusion and were provided with an opt-out form included in the information letter that was given to them and their family or legal representatives.

TABLE 2 Patient characteristics.

	Total (N = 98)	ICU (N = 56)	Non-ICU (N = 42)
Age in years, mean (SD)	69 (12)	63 (11)	76 (7)
Male sex, n (%)	60 (61%)	34 (61%)	26 (62%)
Female sex, n (%)	38 (39%)	22 (39%)	16 (38%)
UMCU, n (%)	73 (75%)	31 (55%)	42 (100%)
Radboud, UMC n (%)	25 (26%)	25 (45%)	0 (0%)
Surgical patients, n (%)	77 (79%)	44 (79%)	33 (79%)
Delirium, n (%)	27 (28%)	15 (27%)	12 (29%)
Hyperactive delirium	1 (4%)	1 (7%)	0 (0%)
Hypoactive delirium	12 (44%)	7 (47%)	5 (42%)
Mixed delirium	14 (52%)	7 (47%)	7 (58%)
Relevant psychiatric or neurological medical history			
Depression, n (%)	8 (8%)	6 (11%)	2 (5%)
Alcohol abuse, n (%)	5 (5%)	3 (5%)	3 (5%)
CVA, n (%)	14 (14%)	6 (11%)	8 (19%)
Parkinson, n (%)	3 (3%)	1 (2%)	2 (5%)
Other, n (%)	8 (8%)	2 (4%)	6 (14%)
Relevant medication 2 h prior to measurement			
Antipsychotics, n (%)	4 (4%)	2 (36%)	2 (5%)
Alpha-2-antagonists, n (%)	11 (11%)	11 (20%)	0 (0%)
Benzodiazepines, n (%)	2 (2%)	2 (4%)	0 (0%)
Opioids, n (%)	19 (19%)	15 (27%)	4 (10%)
Esketamine, n (%)	2 (2%)	2 (4%)	0 (0%)
Propofol, n (%)	4 (4%)	4 (7%)	0 (0%)
Relevant medication 24 h prior to measurement			
Antipsychotics, n (%)	23 (23%)	17 (30%)	6 (14%)
Alpha-2-antagonists, n (%)	17 (17%)	16 (29%)	1 (2%)
Benzodiazepines, n (%)	19 (19%)	12 (21%)	7 (17%)
Opioids, n (%)	55 (56%)	33 (59%)	22 (52%)
Esketamine, n (%)	7 (7%)	5 (9%)	2 (5%)
Propofol, n (%)	8 (8%)	7 (13%)	1 (2%)

Note: Patients with another psychiatric or neurological history suffered from: epilepsy ( $n = 1$ ), meningitis ( $n = 2$ ), subdural hematoma 2 months earlier ( $n = 1$ ), intraventricular bleeding ( $n = 1$ ), mild cognitive impairment ( $n = 1$ ), amyotrophic lateral sclerosis ( $n = 1$ ), anxiety disorder ( $n = 1$ ). Registered antipsychotics are Haloperidol, Olanzapine, Quetiapine and Clozapine. Registered alpha-2 antagonists are Clonidine and Dexmedetomidine. Registered benzodiazepines are: Midazolam, Lorazepam, Temazepam, Oxazepam, Diazepam. Registered Opioids are: Morphine, Fentanyl, Remifentanyl, Sufentanil, Tramadol, Piritramide, Oxycodon.

## Data collection and study design

The 10 min Delirium Interview was conducted by a trained researcher in the presence of two delirium experts: the Live Interview Assessor Panel (LIAP) that we used as a reference. More details on the Delirium Interview are described in Table 1 and Supplement 1–5. After the Delirium Interview, the delirium experts were

allowed to ask additional questions to the patient and inspect their EHR. The results of the Delirium Interview were sent out to three other independent delirium experts: the Written Interview Assessor Panel (WIAP). The total pool consisted of 13 delirium experts with an average of 13 years (standard deviation [SD]: 8) of clinical experience. The pool of researchers consisted of six members, trained to conduct the Delirium Interview correctly.

**TABLE 3** Performance and interrater reliability of the Delirium Interview.

	Total (N = 98)	ICU (N = 56)	Non-ICU (N = 42)
Overall accuracy	84% (75%–90%)	79% (66%–89%)	90% (77%–97%)
Sensitivity	89% (72%–98%)	80% (52%–96%)	100% (74%–100%)
Specificity	82% (71%–90%)	78% (62%–89%)	87% (69%–96%)
NPV	95% (86%–99%)	91% (77%–98%)	100% (87%–100%)
PPV	66% (49%–80%)	57% (34%–78%)	75% (48%–93%)
Kappa LIAP	0.75 (0.63–0.87)	0.71 (0.54–0.88)	0.79 (0.62–0.97)
Kappa WIAP	0.46 (0.35–0.58)	0.44 (0.29–0.59)	0.49 (0.32–0.67)
ICC LIAP	0.90 (0.86–0.93)	0.91(0.86–0.94)	0.88 (0.81–0.93)
ICC WIAP	0.67 (0.57–0.75)	0.67 (0.55–0.78)	0.66 (0.51–0.79)

*Note:* We tested the diagnostic accuracy of paper-based assessments of the Delirium Interview by comparing assessments of the Live Interview Assessor Panel (LIAP, 2 delirium experts and a researcher assessing the patient at bedside) with the Written Interview Assessor Panel (WIAP, 3 different independent delirium experts assessing the patient by reading the Delirium Interview). Delirium experts were selected out of a pool of 13 and researchers out of a pool of 6. Data are shown as value with 95% confidence interval (CI). The delirium probability score included a few missing's: 18 (6% of 3 × 98) within the Live Interview Assessor Panel and 1 (0%) within the Written Interview Assessor Panel. Complete case analysis of the Live Interview Assessor Panel yielded similar results (ICC = 0.89 95% CI: 0.84–0.92, N = 80).

## Statistical analysis

Statistical analyses were performed in Ri386 version 4.0.3. For our main analyses, we tested the performance of the Delirium Interview by comparing the majority vote of the WIAP with the reference: the majority vote of the LIAP (Figure S1). Predictive values, sensitivity, specificity, and overall accuracy were calculated using a 2 × 2 contingency table and 95% confidence intervals based on 2000 bootstrap samples. Overall accuracy was calculated by dividing the number of correct classifications by the total number of subjects. We performed stratified analyses according to admittance to the ICU, the ability to verbally communicate, and a decreased level of consciousness indicated by a RASS < 0 as compared to RASS = 0. Next, interrater reliability for the classification of delirium was calculated with Fleiss' Kappa within the three raters of a panel and Cohen's kappa between two panels. Comparable analyses were performed for the interrater reliability of the delirium probability score (1–10) with the interclass correlation coefficient (ICC, The Oneway Model for randomly unbalanced datasets, including a complete cases analysis since we had more than 5% of missing cases).<sup>19</sup> More details of the statistical analyses are available in Supplement 6.

## RESULTS

### Demographics

In total, 109 patients were assessed by the LIAP, of whom 101 interviews were sent out to the WIAP (for participant

flowchart, see Supplement 7). The final cohort consisted of 98 patients, of whom 56 (57%) were ICU patients and 42 (43%) were non-ICU patients. The cohort included 26 (27%) non-verbal assessments (e.g., due to mechanical ventilation), and 22 (39%) patients had a reduced level of consciousness indicated by a RASS < 0. The mean age was 77 (SD: 12) years, 61% (N = 60) of the population was male, and the overall prevalence of delirium was 28% (N = 27). Patient characteristics are shown in Table 2.

### Performance

The overall accuracy of the WIAP was 84% (95% CI: 75%–90%) compared to the reference: the LIAP. The Delirium Interview had a sensitivity of 89% (95% CI: 72%–98%) and specificity of 82% (95% CI: 71%–90%). The negative predictive value (NPV) and positive predictive value (PPV) were 95% (95% CI: 86%–99%) and 66% (95% CI: 49%–80%), respectively. Overall accuracy was especially high in non-ICU patients (90%, 95% CI: 77%–97%), since all delirious patients were recognized by the WIAP (sensitivity of 100%, 95% CI: 74%–100%), and there were no false negatives (NPV of 100%, 95% CI: 87%–100%). Remarkably, sensitivity remained high in patients with non-verbal assessments (90%, 95% CI: 55%–100%) and those with a decreased level of consciousness indicated by RASS < 0 (83%, 95% CI: 52%–98%), whereas specificity was 64% (95% CI: 35%–87%) and 67% (95% CI: 30%–93%), respectively. However, we should note that the study was not powered for this stratified analysis. Table 3 shows an overview of the performance, Supplement 8 stratifies

results for non-verbal assessments and sedated patients, and Supplement 9 provides details on the certainty of the Assessor Panels.

## Interrater reliability

Interrater reliability was substantial within the LIAP with a kappa coefficient of 0.75 (95% CI: 0.63–0.87) and moderate in the WIAP (0.46, 95% CI: 0.35–0.58).<sup>20</sup> A higher agreement was measured within the delirium probability score with an excellent Intraclass Correlation (ICC) within LIAP (0.90 95% CI: 0.86–0.93) and moderate ICC in the WIAP (0.67, 95% CI: 0.57–0.75).<sup>21</sup> The delirium probability score included a few missings: 18 (6%) within the LIAP and 1 (0%) within the WIAP. A complete case analysis of the LIAP yielded similar results (ICC = 0.89 95% CI: 0.84–0.92,  $N = 80$ ).

## DISCUSSION

In summary, this prospective multicenter study assessed the diagnostic accuracy of the Delirium Interview, a proposed reference standard for studies on delirium assessment tools for ICU and non-ICU settings. We compared paper-based assessments of the Delirium Interview with live assessments of patients and observed a sensitivity of 88% (95% CI: 72%–98%) and specificity of 82% (95% CI: 71%–90%).

Our results show that this method is accurate and provides important logistical benefits, as it allows the experts to classify delirium at a suitable moment instead of performing a time-consuming bedside assessment. Therefore, this method seems feasible in studies on delirium assessment tools, especially when a large number of patients need to be enrolled.

Our moderate<sup>20</sup> interrater reliability of the WIAP was similar to video assessments of patients (kappa = 0.61).<sup>10</sup> As expected, the LIAP showed a higher kappa of 0.78, which was comparable with studies investigating the interrater reliability of brief delirium monitoring tools in live assessments.<sup>22</sup> These results confirm that although experts may disagree on the classification of delirium, relying on the majority vote of delirium experts yields high diagnostic accuracy. The interrater reliability of our delirium probability scale showed less disagreement than the final diagnosis (delirium/no delirium). This is in accordance with the concept of regarding delirium rather as a continuous spectrum instead of a dichotomous entity.<sup>23–25</sup> We recommend that future studies on delirium assessment tools always use the classification of at least three delirium experts and include a delirium

probability scale to overcome this problem. To minimize subjectivity, it would be best to randomly pair three experts from a considerably large panel with experience in various clinical settings.

The following issues were encountered based on feedback from our Interview Assessor Panels on the Delirium Interview. First of all, we received feedback that the Month of the Year Backwards (MOTYB) was one of the most informative and reliable components of the Delirium Interview. A study focusing on the accuracy of attention tests in delirium concluded that no test could become a single-item screening tool, but that the MOTYB showed the best balance between sensitivity and specificity.<sup>26</sup> Second, it appears crucial to report the reason for a missing item in the Delirium Interview. One study showed that half of the patients who refused to answer questions on an attention test were assessed as delirious by a second examiner.<sup>27</sup> Third, assessors acknowledged having difficulties in diagnosing subsyndromal delirium. In the medical setting, resolving delirium in an improving patient has less diagnostic value, but in research, there is a definite need to define this group of patients. Some experts mentioned that the Delirium Interview could be improved by adding the educational level of the patient, although others disagreed and argued that anyone without an altered mental state should have been able to perform all items of the test, independent of their prior education. Finally, the “narrative” in which the researcher could include subjective elements in the patient report appeared to be a section of great value to the Delirium Interview. To pursue a good narrative, it is crucial that independent researchers are appropriately trained, have intervision sessions and are supervised in the beginning as well as regularly thereafter.

The strengths of this study are the multicenter setting on ICU and non-ICU departments, the various groups of delirium experts, and the expertise used to compose the Delirium Interview. A limitation of the Delirium Interview is that we did not include an objective measurement of pre-existent cognitive status. Since dementia affects all the selected attentional and cognitive tests,<sup>15</sup> we chose to exclude patients with cognitive impairment based on the medical history reported in the EHR. Further research could explore a reference standard for delirium superimposed on dementia.<sup>28</sup> Performance was not majorly influenced by difficult subgroups since sensitivity remained high and specificity was sufficient in patients unable to communicate verbally (e.g., mechanically ventilated patients) and in those with a low level of consciousness. However, our sample was unpowered for such stratified analysis, and thus we recommend future studies focus on the isolated performance of these patients. Furthermore, future studies may benefit from exploring delirium

severity, for example, by ranking the various items of the Delirium Interview.

## CONCLUSION

Delirium experts could assess delirium with high accuracy based on written reports and responses to the Delirium Interview without being present at the bedside, which makes this method feasible to study large cohorts. We believe that the Delirium Interview may serve as an appropriate reference standard for the diagnosis of delirium for both ICU and non-ICU patients when used in studies on the evaluation of delirium assessment tools.

## AUTHOR CONTRIBUTIONS

All authors confirmed approval for the manuscript to be published. Fienke L. Ditzel, MD: This author contributed to data collection, data preparation, analysis and preparation of the manuscript. Arjen J. C. Slooter, MD, PhD: This author contributed to the study concept and design, composing the Delirium Interview, data collection and revision of the manuscript. Mark van den Boogaard, RN, PhD: This author contributed to the study concept, composing the Delirium Interview, data collection and revision of the manuscript. Michel Boonstra, MSc: This author contributed to the study concept and design, composing the Delirium Interview, data collection, data preparation and revision of the manuscript. Tim van Nesselrooij: This author contributed to data collection. Marjan Kromkamp, MD, PhD: This author contributed to data collection. Monica Pop-Purceanu, MD, PhD: This author contributed to data collection and revision of the manuscript. Paul J. T. Rood, RN, PhD: This author contributed to data collection. Robert Jan Osse, MD, PhD: This author contributed to data collection. Carol K. Chan, MBBCh, MSc: This author contributed to revision of the manuscript. Alasdair M. J. MacLulich, MRCP, PhD: This author contributed to the study concept, composing the Delirium Interview and revision of the manuscript. Zoë Tieges, PhD: This author contributed to the study concept, composing the Delirium Interview and revision of the manuscript. Karin J. Neufeld, MPH, MD: This author contributed to the study concept, composing the Delirium Interview and revision of the manuscript. Suzanne C. A. Hut, PhD: This author contributed to the study concept and design, composing the Delirium Interview, data collection, data preparation, analysis and revision of the manuscript.

## ACKNOWLEDGMENTS

We would like to thank A. Vondeling, MD, C. H. Röder, MD, PhD, J.W.M. Krulder, MD, R. Faaij, MD, PhD,

A. Kamper, MD, A. de Jonghe, MD, PhD, P. Bouvy, MD, PhD, E. van Dellen MD, PhD, K. Milisen, RN, PhD, B. van Munster, MD, PhD and especially M. Coesmans for their share in the paper-based assessments of the Delirium Interview and K. Luijken, PhD for help with the analysis.

## CONFLICT OF INTEREST

None of the other authors reports any conflicts of interest.

## SPONSOR'S ROLE

This work was supported by European Union Horizon 2020 (grant number 820555). The sponsor had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit for publication.

## ORCID

Fienke L. Ditzel  <https://orcid.org/0000-0003-2011-9917>

## REFERENCES

1. American Psychiatric Association. DSM-5 Diagnostic Classification. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association Publishing; 2013. doi:10.1176/appi.books.9780890425596.x00diagnosticclassification
2. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. *Lancet*. 2014;383(9920):911-922. doi:10.1016/S0140-6736(13)60688-1.Delirium
3. Salluh JIF, Wang H, Schneider EB, et al. Outcome of delirium in critically ill patients: systematic review and meta-analysis. *BMJ*. 2015;350:1-10. doi:10.1136/bmj.h2538
4. Tsui A, Searle SD, Bowden H, et al. The effect of baseline cognition and delirium on long-term cognitive impairment and mortality: a prospective population-based study. *Lancet Health Longev*. 2022;3:232-241. doi:10.1016/s2666-7568(22)00013-7
5. Marra A, Kotfis K, Hosie A, et al. Delirium monitoring: yes or No? That is the question. *Am J Crit Care*. 2019;28(2):127-135. doi:10.4037/ajcc2019874
6. 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(6):1946-1951. doi:10.1097/CCM.0b013e31824e16c9
7. Gibb K, Seeley A, Quinn T, et al. The consistent burden in published estimates of delirium occurrence in medical inpatients over four decades: a systematic review and meta-analysis study. *Age Ageing*. 2020;49(3):352-360. doi:10.1093/ageing/afaa040
8. Tieges Z, Evans JJ, Neufeld KJ, MacLulich AMJ. The neuropsychology of delirium: advancing the science of delirium assessment. *Int J Geriatr Psychiatry*. 2018;33(11):1501-1511. doi:10.1002/gps.4711
9. Neufeld K, Nelliott A, Inouye SK, et al. Delirium diagnosis methodology used in research: a survey-based study. *Bone*. 2012;23(1):1-7. doi:10.1016/j.jagp.2014.03.003.Delirium
10. Numan T, van den Boogaard M, Kamper AM, Rood PJT, Peelen LM, Slooter AJC. Recognition of delirium in postoperative elderly patients: a multicenter study. *J Am Geriatr Soc*. 2017;65(9):1932-1938. doi:10.1111/jgs.14933

11. Cohen JF, Korevaar DA, Altman DG, et al. Guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2015;6(11). doi:10.1136/bmjopen-2016-012799
12. Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond agitation-sedation scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med*. 2002;166(10):1338-1344. doi:10.1164/rccm.2107138
13. Hall R, Stíobhairt A, Allerhand M, MacLulich AMJ, Tiegies Z. The observational scale of level of arousal: a brief tool for assessing and monitoring level of arousal in patients with delirium outside the ICU. *Int J Geriatr Psychiatry*. 2020;35(9):1021-1027. doi:10.1002/gps.5324
14. 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(7):1370-1379. doi:10.1097/00003246-200107000-00012
15. Bellelli G, Morandi A, Davis DHJ, et al. Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. *Age Ageing*. 2014;43(4):496-502. doi:10.1093/ageing/afu021
16. Hart RP, Levenson JL, Sessler CN, Best AM, Schwartz SM, Rutherford LE. Validation of a cognitive test for delirium in medical ICU patients. *Psychosomatics*. 1996;37(6):533-546. doi:10.1016/S0033-3182(96)71517-7
17. Trzepacz PT, Mittal D, Torres R, Canary K, Norton J, Jimerson N. 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(2):229-242. doi:10.1176/jnp.13.2.229
18. Folstein M, Anthony JC, Parhad I, Duffy B, Gruenberg EM. The meaning of cognitive impairment in the elderly. *J Am Geriatr Soc*. 1985;33(4):228-235.
19. Ebel RL. Estimation of the reliability of ratings. *Psychometrika*. 1951;16(4):407-424.
20. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica*. 2012;22(3):276-282. doi:10.11613/bm.2012.031
21. Koo TK, Li MY. A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
22. Bhat R, Rockwood K. Inter-rater reliability of delirium rating scales. *Neuroepidemiology*. 2005;25(1):48-52. doi:10.1159/000085440
23. Numan T. *EEG in Delirium*. Vol 5. Ridderprint BV; 2017.
24. Ouimet S, Riker R, Bergeon N, Cossette M, Kavanagh B, Skrobik Y. Subsyndromal delirium in the ICU: evidence for a disease spectrum. *Intensive Care Med*. 2007;33(6):1007-1013. doi:10.1007/s00134-007-0618-y
25. Stevens RD, Pronovost PJ. The spectrum of encephalopathy in critical illness. *Semin Neurol*. 2006;26(4):440-451. doi:10.1055/s-2006-948325
26. Voyer P, Champoux N, Desrosiers J, et al. Assessment of inattention in the context of delirium screening: one size does not fit all! *Int Psychogeriatrics*. 2016;28(8):1293-1301. doi:10.1017/S1041610216000533
27. Neufeld KJ, Leoutsakos JS, Sieber FE, et al. Evaluation of two delirium screening tools for detecting post-operative delirium in the elderly. *Br J Anaesth*. 2013;111(4):612-618. doi:10.1093/bja/aet167
28. Leonard M, O'Connell H, Williams O, et al. Attention, vigilance and visuospatial function in hospitalized elderly medical patients: relationship to neurocognitive diagnosis. *J Psychosom Res*. 2016;90:84-90. doi:10.1016/j.jpsychores.2016.09.011

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Supplement 1. Figure S1.** Study overview

**Supplement 2.** The development of the Delirium Interview

**Supplement 3A.** The Delirium Interview

**Supplement 3B.** Dutch translation of The Delirium Interview

**Supplement 4.** Illustrative case of the Delirium Score Card

**Supplement 5.** Overview of known Delirium monitoring scales and the incorporation in the Delirium Interview

**Supplement 6.** Statistical analysis

**Supplement 7. Figure S2.** Participant flowchart

**Supplement 8.** Stratified performance of the Delirium Interview

**Supplement 9.** Uncertainty of delirium

**How to cite this article:** Ditzel FL, Slooter AJC, van den Boogaard M, et al. The Delirium Interview as a new reference standard in studies on delirium assessment tools. *J Am Geriatr Soc*. 2023;71(6):1923-1930. doi:10.1111/jgs.18263