

**REVIEW****DELIRIUM SCREENING INSTRUMENTS ADMINISTERED BY NURSES FOR HOSPITALIZED PATIENTS – LITERATURE REVIEW****Blažena Ševčíková<sup>1</sup>, Hana Matějovská Kubešová<sup>2</sup>, Lenka Šáteková<sup>3</sup>, Elena Gurková<sup>3</sup>**<sup>1</sup>Department of Nursing and Midwifery, Faculty of Medicine, University of Ostrava, Czech Republic (student)<sup>2</sup>Department of Nursing and Midwifery, Faculty of Medicine, University of Ostrava, Czech Republic<sup>3</sup>Department of Nursing, Faculty of Health Sciences, Palacký University Olomouc, Czech Republic

Received January 28, 2019; Accepted September 30, 2019. Copyright: This is an open access article distributed under the terms of the Creative Commons Attribution International License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

**Abstract**

*Aim:* The aim of the paper was to create a literature review of valid and reliable delirium screening instruments administered by nurses for hospitalized patients. *Design:* Literature review. *Methods:* An advanced search of three licensed electronic databases (EBSCO, MEDLINE, PROQUEST) was selected. Twenty-one research studies complied with the inclusion criteria. *Results:* The predictive validity of ten delirium screening measuring tools were studied. The best predictive validity was found in the Nursing Delirium Screening Scale (Nu-DESC), with sensitivity values ranging from 32% to 100%, and specificity from 83% to 100%, with highest value for reliability of 0.94. The Delirium Observation Screening Scale (DOS) came second, with sensitivity from 25% to 97%, and specificity from 89% to 98.4%, with highest value for reliability of 0.77. The 4AT tool had sensitivity values from 83.3% to 90%, and specificity from 84% to 86.3%, with highest value for reliability of 0.99. *Conclusion:* The best screening tool for evaluating delirium by nurses was the Nu-DESC, followed by the DOS, and the 4AT. We recommend testing the predictive validity and reliability of selected screening tools administered by nurses in conditions of Czech clinical practice.

**Keywords:** adult, delirium, measuring instrument, nurse, psychometric properties, screening.

**Introduction***Delirium and the importance of diagnostic criteria*

Delirium is characterized by acute disturbance of attention and consciousness, impaired cognition, with a tendency to fluctuate. It is defined as a non-specific, pathological reaction of the brain to various adverse agents (Pečeňák, 2011; American Psychiatric Association, 2013). A standardization of the diagnostic criteria was first included in the 3<sup>rd</sup> edition of the Diagnostic and statistical manual of the APA in 1980 (Grover, Kate, 2012; De, Wand, 2015). Since this publication, terminological confusion has been reduced, the diagnostic criteria revised, and clinicians have a greater understanding of delirium. The Diagnostic and statistical manual, currently in its revised 4<sup>th</sup> edition, (hereinafter referred to as DSM-IV-TR) describes delirium as a disturbance of consciousness and change in cognition that develops over a short period of time. It includes changes in state of consciousness and a reduced ability to focus and retain attention during the course of a day, with evidence that the disturbance is caused by direct physiological consequences of a general medical

condition (American Psychiatric Association, 2000). In the International Classification of Diseases (ICD) used in the Czech Republic, the term delirium is listed as a qualitative disturbance of consciousness without specifying the severity and related symptomatology. Several authors (e.g., Grover, Kate, 2012; De, Wand, 2015) point out that despite developments in diagnostic criteria, differences in terminology relating to delirium persist. Using valid and reliable scales may help clinical and research staff, not only to detect delirium, but also to assess its severity and the efficacy of treatment.

*Prevalence and impact of delirium*

Delirium is a severe and common complication, particularly among older patients hospitalised in standard wards (van Velthuisen et al., 2016). It is also a frequent complication in patients in surgical wards after an operation, and in Intensive Care Units (ICU) (Godfrey et al., 2013; van de Steeg et al., 2014; Balková, Tomagová, 2018). The incidence of delirium among the adult population is estimated at between 3% and 29%. There is a particularly high incidence of delirium among older hospitalized patients with dementia (22–89%). Older adults from social care institutions with deteriorating cognitive functions present a particular risk group (De, Wand, 2015). Delirium may cause falls from bed, attempts to run

Corresponding author: Blažena Ševčíková, I. P. Pavlova 1000/36, Olomouc, Czech Republic; email: [blazena.sevcikova@upol.cz](mailto:blazena.sevcikova@upol.cz)

away from fictitious pursuers, and self-harm. Although screening for delirium is a standard procedure in many healthcare facilities, delirium related events are often not identified, with up to 72% not recognised as such or otherwise misevaluated (van de Steeg et al., 2014). Inadequate detection of delirium (staff correctly identified only 23% of cases) was observed despite an earlier targeted educational intervention (Wand et al., 2014).

#### *Screening tools administered by nurses*

The assessment of delirium and the prescription of appropriate medication help prevent serious complications from developing. For this reason, it is necessary to use valid and reliable screening tools. In the past two decades, several tools have been developed to detect delirium, designed for various clinical settings, such as ICUs or community care. The tools developed differ in the age group targeted, e.g., children and young adults; and degree of validity and reliability, with some being too time-consuming for routine assessments. These tools have been evaluated in several traditional and systematic reviews (Adamis et al., 2010; Wong et al., 2010; Grover, Kate, 2012; De, Wand, 2015; van Velthuisen et al., 2016; Balková, Tomagová, 2018).

Grover and Kate (2012) created a narrative overview of the tools available for delirium assessment, divided into nine separate groups according to their purpose: tools for assessing patient disturbance; screening tools for assessment of individual cognitive functions of patients; screening tools for assessment of delirium; measuring tools for the diagnosis of delirium; measuring tools for assessment of delirium severity; instruments for assessment of cognitive symptoms of delirium; instruments for assessment of motor symptoms of delirium; instruments for assessment of aetiology; instruments for assessment of risk factors; and scales used to assess distress due to delirium experience in patients. In their overview, the authors identified six screening tools feasible for use in nursing practice, designed for various clinical settings and age groups (the NEECHAM Confusion Scale, DOSS/DOS, Nu-DESC, ICDSC and PAED scale).

With regard to the overview above, in their systematic review of delirium screening tools for hospitalized patients, De and Wand (2015) focused on an analysis and comparison of 21 tools published in 31 studies. The strength of the review was in its rigorous assessment of quality (the *Standards for the Reporting of Diagnostic Accuracy* – STARD Score was used) and breakdown of tools in terms of the overall population (mixed population of hospitalized patients,

surgical and post-operative patients exclusively, cancer patients, palliative care, and emergency). The *Confusion Assessment Method* (CAM) was listed as the most frequently used tool, although it is designed for use by doctors. From 11 validation studies of CAM, only two studies included nurses as validators (van Velthuisen et al., 2016). Additionally, the *Nurses Delirium Screening Checklist* (Nu-DESC) was listed as the best screening tool for surgical setting. The *Memorial Delirium Assessment Scale* (MDAS) was recommended for palliative and postoperative settings.

A systematic review by van Velthuisen et al. (2016) focused on the analysis and comparison of psychometric properties and performance of delirium detection tools among older adults. The quality of the studies included in the review was assessed using the *QUADAS 2*. The authors identified 28 tools, classified into several groups – observational, interactive, diagnostic, screening, and tools for severity assessment and delirium typology. The authors identified 14 studies and seven tools in which nurses were listed as eligible raters. Balková and Tomagová (2018) analysed tools for postsurgical delirium screening that could be administered by nurses. However, their review includes only a specific group of hospitalised patients.

The above-mentioned reviews do not analyse the performance and psychometric properties in great detail – i.e., validity (specificity, sensitivity, positive and negative predictive values of the tools); and reliability (inter-rater reliability and internal consistency) of the screening tools for which nurses can be raters. The latest overview of tools for delirium screening in hospitalized patients was published in 2015 (De, Wand, 2015), but does not focus exclusively on validation studies which include nurses as raters. Since then, new or modified tools and new validation studies have appeared. However, none of the reviews focuses on delirium screening tools for hospitalized patients administered exclusively by nurses. In the Czech Republic, delirium assessment administered by nurses is not a standard procedure. Currently, only the *Confusion Assessment Method for the Intensive Care Unit* (CAM-ICU) has been validated in a Czech context (Mitášová et al., 2010). Since the number of studies dealing with the issue of tools for screening delirium is rising, it is necessary to offer an up-to-date overview of findings.

#### **Aim**

The aim of the article was to provide an overview of valid and reliable delirium screening tools for

assessment performed by nurses for hospitalized patients. The overview focuses on evaluation of sensitivity, specificity and inter-rater reliability.

## Methods

### Design

Design of this paper is literature review.

### Eligibility criteria

These results were subsequently analysed using inclusion and exclusion criteria. Inclusion criteria were: period 2008–2018, and English full-texts. The literature review included studies, which:

- published psychometric properties of the tools (validity – specificity, sensitivity, positive and negative predictive values; reliability – inter-rater reliability or internal consistency);
- focused exclusively on screening delirium in adult hospitalized patients (included were: observational tools – based solely on observation data, without the need for direct interaction with the patient; interactive tools – based on data obtained from the patient, e.g., from an interview or cognitive test; and mixed tools – combining the above-mentioned types). A similar distinction was used in a review by van Velthuisen et al. (2016);
- included delirium diagnosed according to DSM or ICD by a doctor as the reference standard;
- had a nurse as the rater;
- included tools for diagnosis of delirium in ICUs, palliative care units, and Emergency departments (ED);
- had the tool validated in at least one other language in order to be culturally sensitive.

Exclusion criteria were:

- sources that included information about the tool being used for other purposes than solely screening delirium (tools for assessment of disturbance, tools for delirium diagnosis, tools for assessment of delirium severity, tools for assessment of cognitive symptoms, tools for assessment of motor symptoms, tools for assessment of aetiology, tools for assessment of risk factors, and tools for assessment of distress due to experience of delirium);
- studies assessing delirium screening in paediatric clinical settings;
- tools for which the raters were not nurses – doctors, psychologists, carers, or family members;
- overviews or descriptive studies.

### Search strategy

A systematic literature search was carried out between May 2018 and August 2018. The literature search was done in databases EBSCO, MEDLINE, and PROQUEST.

### Study selection

An advanced search was conducted in four stages. In all stages, the key words measuring tool, screening, delirium, nurse, adult, and psychometric properties were used, together with their synonyms using the Boolean operator “OR”. In the fourth stage, the results from the previous three stages were linked using the Boolean operator “AND”. In the fourth stage, there were 363 relevant results. 24 studies were included in the review (Figure 1).

### Data extraction

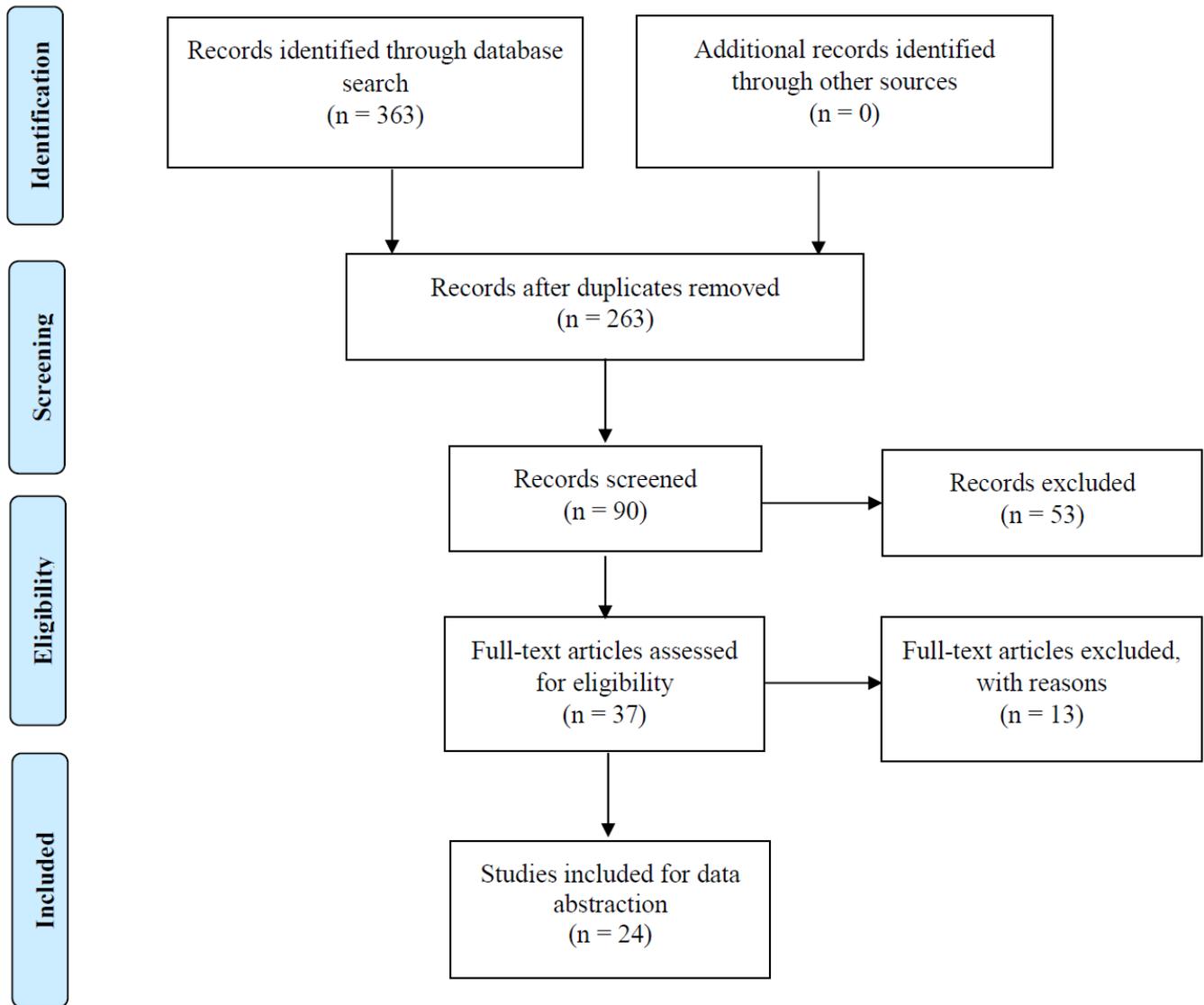
The data on the study design (reference standard, sample, clinical setting), psychometric properties, type (observational, interactive, mixed (van Velthuisen et al., 2016) and appropriateness (country, type of tool, rater, duration of the assessment, costs, previous training) were acquired from the included studies.

The findings of the studies were analysed based on validity criteria – sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic (ROC) curve (AUC) and reliability (inter-rater reliability and internal consistency). Predictive validity determines the likelihood of agreement between the results of measurement and the behaviour of the studied subjects within a certain period. Sensitivity is the ability of the tool to return positive results if risk is present. False negatives lower the value of sensitivity: the test fails to identify individuals that should have been diagnosed as positive for risk. Specificity is the ability of the tool to give negative results if risk is not present. This means giving negative results for healthy individuals (Dušek et al., 2011). Positive predictive value of the tool means the ability to identify patients who are “at risk” of developing delirium, and negative predictive value is the ability to identify patients who will not develop delirium. The ROC curve measures the performance of the tool. The higher the ROC, the more efficient the tool.

## Results

### Results of searching and evaluation

A systematic search returned 363 eligible studies. After an analysis and revision of literature, 24 studies met the inclusion criteria, featuring ten screening tools for assessment of delirium administered by nurses. (Figure 1).



**Figure 1** Flow diagram of the selection and classification of quantitative studies on predictive validity of delirium screening instruments

### **Description and analysis of the studies, results of extraction**

#### *The Nurses Delirium Screening Checklist (Nu-DESC)*

The most frequently tested tool was the Nu-DESC (Koster et al., 2009; Detroyer et al., 2014; Gavinski, Carnahan, Weckmann, 2016; Jorgensen, Carnahan, Weckmann, 2017; Numan et al., 2017). It is a five-item tool designed specifically for nurses. It can be administered in three minutes. Training is required before use (Table 1) (van Velthuisen et al., 2016). The Nu-DESC has been tested in Finland, Sweden, Germany, Italy, China, and the USA in departments of surgery, post-anaesthesia, geriatrics, and orthogeriatrics. The sample size ranged from 88 to 156 and the studies took between two and 20 months (Table 1). The predictive validity values are as

follows: sensitivity from 32% (Neufeld et al., 2013) to 100% (Leung et al., 2008), and specificity from 83% (Luetz et al., 2010) to 100% (Spedale et al., 2016). Positive predictive values and negative predictive values were only provided in one of the studies (Table 2). AUC and ROC ranged between 0.76 (Linghall et al., 2012) and 0.99 (Radtke et al., 2010). Reliability ranged from 0.47 (Poikajärvi et al., 2017) to 0.94 (Spedale et al., 2016). Cut off score ranged from 2 to 3 (Table 2).

#### *Delirium Observation Screening Scale (DOS)*

The second most frequently tested screening tools administered by nurses was the DOS (Koster et al., 2009; Detroyer et al., 2014; Gavinski, Carnahan, Weckmann, 2016; Jorgensen, Carnahan, Weckmann, 2017; Numan et al., 2017). The tool was created by

**Table 1** Characteristics of the screening tools

Screening tool	Type	Number of items	Rater	Administration	Training
DOS	O/S	13	nurse	< 5 min	yes
DEAR	O/S	5	nurse	< 5 min	no
DDS	M/S	8	nurse	< 5 min	yes
mRAAS	M/S	3	nurse	< 30 s	no
Nu-DESC	O/S	5	nurse	3 min	no
NEECHAM	M/S	9	nurse	8 min	yes
SQeeC	O/S	2	nurse	< 5 min	no
SQid	I/S	1	nurse	< 5 min	no
The Sour Seven Questionnaire	I/S	7	nurse	5 min	no
4AT	M/S	4	nurse	< 4 min	no

DOS – Delirium Observation Screening Scale; DEAR – The Delirium Elderly At Risk Instrument; DDS – Delirium Detection Scale; mRAAS – modified Richmond Agitation and Sedation Scale; Nu-DESC – Nursing Delirium Screening Scale; NEECHAM – Neecham Confusion Scale; SQEEC – Simple Query for Easy Evaluation of Consciousness; SQid – Single Question in Delirium; 4AT – The 4A's test; O – observational; M – mixed; I – interactive; S – screening

**Table 2** Psychometric properties of the delirium screening tools (tools sorted by sensitivity values)

Screening tool	Author (year)	Cut-off score	Delirium incidence (%)	Sensitivity (%)	Specificity (%)	PPH (%)	NPH (%)	ROC	Inter-rater reliability (κ, ICC)
Nu-DESC	Leung et al. (2008)	≥ 2	-	100	88	-	-	-	0.94
	Luetz et al. (2010)	2	17	82	83	-	-	-	0.68
	Radtke et al. (2010)	0.83	19	97.65	92.30	-	-	0.99	0.83
	Lingehall et al. (2012)	-	63	65.6	94.9	-	-	0.76	-
	Neufeld et al. (2013)	≥ 2	-	32–80	92–69	-	-	-	-
	Spedale et al. (2016)	3	6.1	76.1	100	-	-	0.94	0.87
	Poikajärvi et al. (2017)	-	86	85.7	86.8	5.4	78.4	-	0.47
	Koster et al. (2009)	-	21	25.0	95.5	60.0	82.4	0.85	-
	Detroyer et al. (2014)	3	22.9	88.1	96.1	-	-	0.93	0.77
DOS	Gavinski Carnahan, Weckmann (2016)	3	9.90	90	91	53	99	0.91	-
	Jorgensen Carnahan, Weckmann (2017)	-	38	97	89	-	-	-	-
	Numan et al. (2017)	≥ 3	32	62.2	98.4	95.8	81.8	-	0.73
NEECHAM	Sörensen Duppils, Johansson (2011)	25	-	100	91	-	-	-	-
	Poikajärvi et al. (2017)	-	-	-	-	26.3	73.7	-	0.87
4AT	Bellelli et al. (2014)	4	12	90	84	-	-	0.92	0.80
	De et al. (2016)	-	62	87	80	87	80	0.92	-
	Kuladee, Prachason (2016)	4	-	83.3	86.3	66.7	94.0	0.91	0.99
The Sour Seven Questionnaire	Shulman, Kalra, Jiang (2016)	4	85.7	89.5	90.0	89.5	90.0	0.92	-
	Chester et al. (2012)	< 0 <	-	85	92	-	-	-	0.48
DEAR	Freter et al. (2015)	2	58	93.2	41.8	53.4	89.6	-	-
DDS	Radtke et al. (2010)	0.77	19	71.18	87.11	-	-	0.88	0.77
	Luetz et al. (2010)	> 7	18	25	89	-	-	-	0.79
SQeeC	Lin et al. (2015)	-	-	83	81	39	97	-	-
SQid	Lin et al. (2015)	-	-	77	51	42	83	-	-

DOS – Delirium Observation Screening Scale; DEAR – The Delirium Elderly At Risk Instrument; DDS – Delirium Detection Scale; mRAAS – modified Richmond Agitation and Sedation Scale; Nu-DESC – Nursing Delirium Screening Scale; NEECHAM – Neecham Confusion Scale; SQEEC – Simple Query for Easy Evaluation of Consciousness; SQid – Single Qu

a team of Schuurmans, Shortridge-Baggett, Duursma (2003) and was first tested in a clinical geriatric setting after a hip fracture. The original 25-item scale was later reduced to a 13-item scale. It can be administered in five minutes. No previous training is necessary (Table 1) (Koster et al., 2009). This tool was also developed for general nurses and has been tested in the Netherlands, Belgium, and the USA in departments of surgery, internal medicine, and palliative care. The sample size ranged from 48 to 167. Only two studies listed duration (Table 1). The predictive validity values are listed in Table 3: sensitivity ranged from 25% (Koster et al., 2009) to 97% (Jorgensen, Carnahan, Weckmann, 2017), specificity values from 89% (Jorgensen, Carnahan, Weckmann, 2017) to 98.4% (Numan et al., 2017), positive predictive value from 53% (Gavinski, Carnahan, Weckmann, 2016) to 95.8% (Numan et al., 2017), negative predictive values from 81.8 % (Numan et al., 2017) to 99 % (Gavinski, Carnahan, Weckmann, 2016). AUC ranged from 0.85 (Koster et al., 2009) to 0.93 (Detroyer et al., 2014). Reliability values were mentioned in only two studies and ranged from 0.73 (Numan et al., 2017) to 0.77 (Detroyer et al., 2014). The cut-off score ranged from  $\geq 2$  to 3 (Table 2).

#### “4 A’s test” (4AT)

The “4 A’s test” (4AT) was included in three studies (Bellelli et al., 2014; De et al., 2016; Kuladee, Prachason, 2016). It is a four-item tool for nurses. It can be administered in less than four minutes. No training is required before use (van Velthuisen et al., 2016) (Table 3). The 4AT tool has been tested in Italy, Australia, and Thailand in geriatrics, orthogeriatrics and ED’s. The length ranged from five to 18 months (Table 1). Values for predictive validity ranged as follows: sensitivity from 83.3% (Kuladee, Prachason, 2016) to 90% (Bellelli et al., 2014), and specificity from 80% (De et al., 2017) to 86.3% (Kuladee, Prachason, 2016). Positive predictive value was between 66.7 (De et al., 2017) to 87 (Kuladee, Prachason, 2016). Negative predictive value ranged from 80 (De et al., 2017) to 94 (Kuladee, Prachason, 2016). AUC was 0.92 (Bellelli et al., 2014; De et al., 2017). Reliability ranged from 0.80 (Bellelli et al., 2014) to 0.99 (Kuladee, Prachason, 2016). The cut-off score 4 was found in three studies (Table 2).

#### Neerham Confusion Scale (NEECHAM)

Two of the included studies tested the NEECHAM *Delirium Detection Scale* (DDS) (Sörensen Duppils, Johansson, 2011; Poikajärvi et al., 2017). NEECHAM is a nine-item measuring tool intended for use by general nurses. It can be administered in eight minutes. Training is required beforehand (van Velthuisen et al., 2016) (Table 1). The studies were conducted

in Finland and Sweden at surgical and orthopaedic departments, with samples of 112 and 147 participants, respectively. Only Poikajärvi et al. (2017) included the duration (20 months) of the study (Table 3), while only Sörensen Duppils, Johansson (2011) provided values for sensitivity (100%) and specificity (91%). Positive and negative predictive values and reliability were included only in the study by Poikajärvi et al. (2017) (Table 2). No reliability values were provided in either study. AUC was provided only in Poikajärvi et al. (2017) at 0.87. The cut-off score was 25 points (Sörensen Duppils, Johansson, 2011).

#### *Delirium Detection Scale (DDS)*

The DDS is an eight-item tool designed for nurses. It can be administered in less than five minutes. Training is required before use (van Velthuisen et al., 2016) (Table 1). Testing was conducted in Germany in departments of surgery. The sample size ranged from 88 (Radtke et al., 2010) to 156 (Luetz et al., 2010). The testing took from three to 12 months (Table 1). Sensitivity values were between 25% (Luetz et al., 2010) and 71.18% (Radtke et al., 2010). Specificity ranged from 89% (Luetz et al., 2010) to 87.11% (Radtke et al., 2010). Positive and negative predictive values were not mentioned in either of the studies which tested the DDS. Only Radtke et al. (2010) included a value for AUC (0.88). Reliability ranged from 0.77 (Radtke et al., 2010) to 0.79 (Luetz et al., 2010) (Table 2). The cut-off score ranged from 0.77 (Radtke et al., 2010) to  $> 7$  (Luetz et al., 2010).

#### *Other screening tools*

Five other screening tools were tested in one study only: *The Delirium Elderly At Risk Instrument* (DEAR), the *modified Richmond Agitation and Sedation Scale* (mRAAS), *Simple Query for Easy Evaluation of Consciousness* (SQeeC), *Single Question in Delirium* (SQID) and *The Sour Seven Questionnaire*. All are screening tools developed for use by nurses. The number of items ranged from seven in the *Sour Seven Questionnaire* (Shulman, Kalra, Jiang, 2016) to one in the SQID (Lin et al., 2015). They can be administered in a period of  $< 30$  seconds to five minutes (van Velthuisen et al., 2016). Studies were conducted in England, Australia, and Canada in departments of internal medicine, orthopaedics, and geriatrics. The samples ranged from 80 (Shulman, Kalra, Jiang, 2016) to 283 (Freter et al., 2015) and took three months. In one case, the duration was not mentioned (Table 1). Values of predictive validity were as follows: sensitivity from 77% (Lin et al., 2015) to 93.2% (Freter et al., 2015), specificity from 41.8% (Freter et al., 2015) to 92% (Chester et al.,

**Table 3** Studies characteristics (Part 1)

Author (year), country	Aim of the study	Sample size	Clinical setting	Methodology	Research duration	Conclusion	Criteria
<b>Screening tool DOS</b>							
Koster et al. (2009), Netherlands	to investigate the predictive validity of the DOS tool	112	department of cardiosurgery	prospective observational study	N/A	the DOS has a good predictive validity	DSM-IV
Detroyer et al. (2014), Belgium	to investigate the predictive validity of the DOS tool	48	palliative care	prospective observational study	8 months	the DOS tool was deemed valid	DSM-IV
Gavinski, Carnahan, Weckmann (2016), USA	to investigate the predictive validity of the DOS tool	101	department of internal medicine	prospective observational study	N/A	the DOS is a valid and easy-to-use tool	DSM-IV
Jorgensen, Carnahan, Weckmann (2017), USA	to investigate the validity of DOS when identifying delirium in home hospice patients	75	palliative care	prospective observational study	N/A	the DOS is valid in hospice care	DSM-IV
Numan et al. (2017), Netherlands	to investigate the predictive validity and reliability of the DOS tool	167	department of surgery	prospective observational study	N/A	sensitivity was lower, reliability was satisfactory	DSM-V
<b>Screening tool DEAR</b>							
Freter et al. (2015), Canada	to ascertain the ability of the DEAR to identify patients with high risk of postoperative delirium	283	orthopaedics	prospective observational study	N/A	the DEAR may help identify patients at risk of postoperative delirium	N/A
<b>Screening tool DDS</b>							
Radtke et al. (2010), Germany	to investigate the predictive validity and reliability of the DDS tool	88	department of surgery	prospective observational study	12 months	the DDS has low sensitivity scores	DSM-IV
Luetz et al. (2010), Germany	to compare the predictive validity and reliability of three tools for assessment of delirium	156	department of surgery	prospective observational study	3 months	the DDS should not be used as a screening tool	DSM-IV
<b>Screening tool mRASS</b>							
Chester et al. (2012), UK	to investigate the predictive validity and reliability of the tool for delirium assessment	95	geriatrics	prospective study	N/A	the mRASS has good psychometric properties	DSM-IV

**Table 3** Studies characteristics (Part 2)

Author (year), country	Aim of the study	Sample size	Clinical setting	Methodology	Research duration	Conclusion	Criteria
<b>Screening tool – Nu-DESC</b>							
Leung et al. (2008), China	to ascertain the predictive validity and reliability of the screening tool	100	geriatrics	prospective observational study	3 months	the Nu-DESC had high psychometric property scores	DSM-IV
Luetz et al. (2010), Germany	to compare the predictive validity and reliability of three tools for assessment of delirium	156	department of surgery	prospective observational study	3 months	the Nu-DESC had the best psychometric properties	DSM-IV
Radtke et al. (2010), Germany	to investigate the predictive validity and reliability of the tools for postoperative delirium assessment	88	department of surgery	prospective observational study	12 months	the Nu-DESC has the best psychometric properties	DSM-IV
Lingehall et al. (2012), Sweden	to investigate the predictive validity of the Nu-DESC tool	142	department of cardiosurgery	prospective observational study	8 months	the Nu-DESC has low prediction validity	DSM-IV
Neufeld et al. (2013), USA	to investigate the predictive validity of the Nu-DESC tool	91	post-anaesthetic department	prospective observational study	2 months	the tool has high prediction validity	DSM-IV
Spedale et al. (2016), Italy	to investigate the predictive validity of the Nu-DESC tool	101	geriatrics and orthogeriatrics	prospective observational study	4 months	the tool has good psychometric properties	DSM-IV
Poikajarvi et al. (2017), Finland	psychometric testing of the Nu-DESC	112	department of surgery	prospective observational study with a randomized sample	20 months	the Nu-DESC had the best psychometric properties	DSM-IV
<b>Screening tool – NEECHAM</b>							
Duppils, Johansson et al. (2011), Sweden	to evaluate the predictive validity and reliability of the NEECHAM tool	149	orthopaedics	prospective observational study	N/A	the NEECHAM is a reliable screening tool	DSM-IV
Poikajarvi et al. (2017), Finland	psychometric testing of the NEECHAM	112	department of surgery	prospective observational study with a randomized sample	20 months	the NEECHAM tool had appropriate psychometric properties	DSM-IV

**Table 3** Studies characteristics (Part 3)

Author (year), country	Aim of the study	Sample size	Clinical setting	Methodology	Research duration	Conclusion	Criteria
<b>Screening tool – SQeeC</b>							
Lin et al. (2015), Australia	to evaluate the predictive validity and reliability of the SQeeC tool	100	department of internal medicine	prospective observational study	3 months	the SQeeC had appropriate psychometric properties	N/A
<b>Screening tool – SQiD</b>							
Lin et al. (2015), Australia	to evaluate the predictive validity and reliability of the SQiD tool	100	department of internal medicine	prospective observational study	3 months	the SQiD had satisfactory psychometric properties	N/A
<b>Screening tool – The Sour Seven Questionnaire</b>							
Shulman, Kalra, Jiang (2016), Canada	to evaluate the predictive validity and reliability of the The Sour Seven Questionnaire	80	N/A	prospective observational study	3 months	the Sour Seven Questionnaire shows good values for predictive validity	DSM-IV
<b>Screening tool – 4AT</b>							
Bellelli et al. (2014), Italy	to evaluate psychometric properties of the 4AT tool	234	geriatrics	prospective observational study	5 months	the 4AT has good psychometric properties	DSM-IV
De et al. (2016), Australia	to verify the predictive validity and reliability of the 4AT in geriatric and orthogeriatric patients	257	geriatrics and orthogeriatrics	prospective observational study	6 months	the 4AT demonstrates high predictive validity	DSM-V
Kuladee, Prachason (2016), Thailand	to evaluate predictive validity of the 4AT tool	97	emergency department	prospective observational study	18 months	the 4AT showed good predictive validity	DSM-IV

DOS – Delirium Observation Screening Scale; DEAR – The Delirium Elderly At Risk Instrument; DDS – Delirium Detection Scale; mRAAS – modified Richmond Agitation and Sedation Scale; Nu-DESC – Nursing Delirium Screening Scale; DDS – Delirium Detection Scale; NEECHAM – Neecham Confusion Scale; SQeeC – Simple Query for Easy Evaluation of Consciousness; SQiD – Single Question in Delirium; 4AT – The 4A's test

2012). Positive predictive values were between 39 (Lin et al., 2015) and 89.5 (Shulman, Kalra, Jiang, 2016), negative predictive values ranged from 83 to 97 (Lin et al., 2015). The AUC was provided only in the study dealing with *The Sour Seven Questionnaire*, with a value of 0.92 (Shulman, Kalra, Jiang, 2016). A value for reliability (0.48) was provided only in the study on the mRAAS by Chester et al. (2012).

## Discussion

Our literature review identified ten observational or interactive tools designed for delirium screening administered by nurses. Five tools were tested only

in single studies. The remaining five tools (Nu-DESC; DOS; 4AT; NEECHAM; DDS) were the subject of two or more studies. In accordance with previous systematic reviews (Grover, Kate, 2012; De, Wand, 2015; van Velthuisen et al., 2016), we can say that the Nu-DESC and DOS are the most suitable tools for clinical delirium screening administered by nurses (not only in geriatrics, but also in surgical and postoperative clinical environments).

We also found that the Nu-DESC is the tool most tested with nurses as the raters (seven studies). The CAM can be regarded as the “gold standard” in delirium assessment, which is in accordance with

the latest reviews published (De, Wand, 2015; van Velthuisen et al., 2016; Balková, Tomagová, 2018). However, van Velthuisen et al. (2016) point out that CAM is an interactive tool, which requires not only observation of the patient, but also cognitive testing, which makes it more time-consuming. For this reason, the CAM requires training before use, and is more suitable for use by doctors, rather than as a routine, indicative screening test administered by nurses. Due to its diagnostic focus, we did not include CAM in the review. The CAM has been modified for use in intensive care (CAM – ICU), and is considered a better tool for nurses to identify postoperative delirium, and delirium in a general intensive care setting (Balková, Tomagová, 2018). De and Wand (2015) consider delirium in ICUs to be a completely different entity to delirium in other departments, due to the severity of the disease, invasive treatment strategy, frequent IV sedation, etc. van Velthuisen et al. (2016) stress that its use in other departments should not be taken for granted and recommend that the tool's psychometric properties be tested outside intensive care units.

By analogy, what CAM is as a screening tool for diagnosis of delirium by doctors, Nu-DESC could be as a screening tool for diagnosis of delirium by nurses. Its feasibility lies in its small number of items requiring only patient observation, the short time needed for its administration, and the fact that nurses can administer it without previous training. However, despite its positive properties for practical application in daily nursing care, its psychometric values seem less beneficial, in particular, its sensitivity. The psychometric properties of the Nu-DESC differed from study to study, ranging between 32% to 100%, rising in relation to the cut-off score. The target population included in the testing of the tool may have caused the differences in results. In an American study focusing on screening of postoperative delirium, Neufeld et al. (2013) found that the tool's sensitivity was higher with a cut-off score of  $\geq 1$  (80%) in comparison to a cut-off score of  $\geq 2$  (32%). On the other hand, with a lower cut-off score, sensitivity increased at the cost of lower specificity. Chinese authors Leung et al. (2008) tested the Nu-DESC in a population of geriatric patients, in which sensitivity was high (100%).

In geriatrics, departments of surgery, and internal medicine, the second most frequently used tool tested in the included studies was the DOS. It has similar advantages to the Nu-DESC (it can be administered in a short time, observation only, no need for special training), and has comparable psychometric properties. In our review, the sensitivity values of the DOS varied considerably (25–97%). This means that

in clinical practice, there could be a problem with high false negatives (the higher the sensitivity, the fewer false negative results). The tool might not correctly detect patients presenting with delirium. The specificity values are good (89% and higher). The DOS tool has an excellent ability to identify patients without risk. Unlike the Nu-DESC, the values for sensitivity and specificity of the DOS are not balanced (there is a greater dispersion). An advantage of the DOS is the fact that it has been tested in several European countries. A disadvantage of both the DOS and the Nu-DESC is that they have not as yet been tested in a population of Czech patients.

Compared to the NEECHAM, both the Nu-DESC and DOS have better psychometric properties and are also easier to administer (focus of the items, time for administration, the need for training). According to van Velthuisen et al. (2016), as the NEECHAM is a tool combining interactive and observational aspects, it takes longer to administer and is therefore more time-consuming, and training is required for nurses. Finnish authors Poikajärvi et al. (2017) tested and compared the properties of the NEECHAM and the NU-DESC and confirmed that both tools are reliable and practical for use by nurses. In terms of validity of both tools, they suggest further testing. The Finnish study not only examined the psychometric properties in detail (internal consistency, inter-rater reliability and content validity), but also assessed the practical application of both tools. Nurses were mainly critical about the NEECHAM tool due to the length of administration, and the need to document certain parameters about the patient (e.g., to objectify physiological parameters), which may lead to duplication of work in patient documentation (when scaling and when completing nursing records). van Velthuisen et al. (2016) are critical about the methodology used in the studies that tested the NEECHAM. They point out the lack of evidence for the predictive validity of the tool. A positive is that NEECHAM has been tested in Slovakia as part of the validation of the nursing diagnosis Acute Confusion (Vörösová et al., 2007). Balková, Tomagová (2018) point to its time-consuming administration as a clear limitation to its use in clinical practice.

The 4AT is one of the five screening tools listed in three studies included in our review. Sensitivity of the 4AT in the studies by De et al. (2016) and Kuladee, Pracharson (2016) was lower (83.3%, 87%) than in the original study by Bellelli et al. (2014), who claim sensitivity of 90%. Yet, the 4AT had higher specificity values (80–86.3%) (De et al., 2016; Kuladee, Pracharson, 2016).

Nevertheless, these values are lower in comparison to the Nu-DESC, DOS, and NEECHEM. The values for specificity may be inaccurate if there is a high number of cognitive disorders (such as dementia of organic brain syndrome) in the sample. This fact may contribute to a higher number of false positives. According to the authors, another factor which could have caused the relatively low sensitivity and specificity values was staff shortage for shifts. This caused errors in tool administration due to time stress (Bellelli et al., 2014; Kuladee, Pracharson, 2016). However, Bellelli et al. (2014), De et al. (2016) and Kuladee and Pracharson (2016) agree that it is a concise screening tool for the assessment of delirium in older hospitalized patients, but that it is also suitable for patients with dementia and patients who are not English speakers. They also recommend further validation studies of the 4AT.

The DDS is a rarely used screening tool. Radtke et al. (2010) and Luetz et al. (2010) agree on its satisfactory to below-satisfactory sensitivity values. Luetz et al. (2010) reported lower sensitivity (25%) with a cut-off score > 7. Good sensitivity can be achieved if the cut-off score is 3, using the DSM-IV as a gold standard. A limitation of the study is the small sample size and the fact that delirium was only assessed once a day (Radtke et al., 2010). Luetz et al. (2010) and Radtke et al. (2010) agree that as a screening tool for assessment of delirium it is suitable for surgical settings.

Other screening tools were evaluated by single studies only, and for this reason are not discussed in detail here.

### Limitation of study

A limitation of this literature review is the fact that the methodological quality of the included studies about the tools was not evaluated. Another limitation might lie in the missing data about the psychometric properties of each tool in the studies (some included only predictive validity, some only reliability). The data on negative/positive predictive values were listed in only a few studies, with only 11 out of 24 studies providing this data. In further research into the screening tools, it is necessary to focus on false negative scores that may go on to affect negative/positive predictive values.

### Conclusion

Screening for delirium in hospitalized patients may be difficult for general nurses, yet it is essential in nursing care and treatment. Screening tools differ in the number of items, time required for administration, and levels of presumed knowledge and training required prior to administration. The tools with the highest

sensitivity, specificity, reliability, and brevity are the Nu-DESC or DOS, as described in five or more studies. These have the most satisfactory psychometric properties and feasibility, making the assessment of delirium possible during the regular nursing routine. The next best screening tools in terms of high sensitivity and specificity were the NEECHAM and 4AT. It is necessary to investigate the psychometric properties of delirium screening tools further in a Czech clinical environment, and to compare the results with studies from abroad. We can then decide which of the above-mentioned tools is the most valid in a Czech clinical setting. Prior to the research itself, a diligent and rigorous translation should be undertaken, according to recognised methodology.

### Ethical aspects and conflict of interest

All sources are duly cited. The authors declare no conflicts of interests.

### Acknowledgements

The publication was supported by a project at Ostrava University, Faculty of Medicine, Department of Nursing, No. SGS 07/LF/2018-2019.

### Author contributions

Concept and design (BŠ, HMK), analysis and data interpretation (BŠ, LŠ), preparation of the manuscript (BŠ), manuscript critical revision (HMK, LŠ, EG), final revision (BŠ, LŠ).

### References

- Adamis D, Sharma N, Whelan PJ, Macdonald AJ. Delirium scales: a review of current evidence. *Aging and Mental Health*. 2010;14(5):543–555.
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed., text revision. Washington: American Psychiatric Association; 2000.
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington: Arlington; 2013.
- Bellelli G, Morandi A, Davis DH, Mazzola P, Turco R, Gentile S, Ryan T, Cash H, Guerini F, Torpilliesi T, Del Santo F, Trabucchi M, Annoni G, MacLulich AM. Validation of the 4AT, a new instrument for rapid delirium screening: a study in 234 hospitalised older people. *Age and Ageing*. 2014;43(4):496–502.
- Balková M, Tomagová M. Use of measurement tools for screening of postoperative delirium in nursing practice. *Central European Journal of Nursing and Midwifery*. 2018;9(3):897–904.
- Chester JG, Beth Harrington M, Rudolph JL; VA Delirium Working Group. Serial administration of a modified Richmond Agitation and Sedation Scale for delirium screening. *Journal of Hospital Medicine*. 2012;7(5):450–453.

- De J, Wand AP. Delirium screening: a systematic review of delirium screening tools in hospitalized patients. *Gerontologist*. 2015;55(6):1079–1099.
- De J, Wand APF, Smerdely PI, Hunt GE. Validating the 4A's test in screening for delirium in a culturally diverse geriatric inpatient population. *International Journal of Geriatric Psychiatry*. 2017;32(12):1322–1329.
- Detroyer E, Clement PM, Baeten N, Decruyenaere M, Vandenberghe J, Menten J, Joosten E, Milisen K. Detection of delirium in palliative care unit patients: a prospective descriptive study of the Delirium Observation Screening Scale administered by bedside nurses. *Palliative Medicine*. 2014;28(1):79–86.
- Dušek L, Pavlík T, Jarkovský J, Koptíková J. Analýza dat v neurologii. XXV. Hodnocení diagnostických testů – senzitivita a specificita. *Česká a Slovenská Neurologie a Neurochirurgie*. 2011;74/107(1):97–103. (in Czech)
- Freter S, Dunbar M, Koller K, MacKnight C, Rockwood K. Risk of pre- and post-operative delirium and the Delirium Elderly At Risk (DEAR) tool in hip fracture patients. *Canadian Geriatrics Journal*. 2015;18(4):212–216.
- Gavinski K, Carnahan R, Weckmann M. Validation of the delirium observation screening scale in a hospitalized older population. *Journal of Hospital Medicine*. 2016;11(7):494–497.
- Godfrey M, Smith J, Green J, Cheater F, Inouye SK, Young JB. Developing and implementing an integrated delirium prevention system of care: a theory driven, participatory research study. *BMS Health Services Research*. 2013;13:341.
- Grover S, Kate N. Assessment scales for delirium: a review. *World Journal of Psychiatry*. 2012;2(4):58–70.
- Jorgensen SM, Carnahan RM, Weckmann MT. Validity of the Delirium Observation Screening Scale in identifying delirium in home hospice patients. *American Journals of Hospice and Palliative Medicine*. 2017;34(8):744–747.
- Koster S, Hensens AG, Oosterveld FG, Wijma A, van der Palen J. The delirium observation screening scale recognizes delirium early after cardiac surgery. *European Journal of Cardiovascular Nursing*. 2009;8(4):309–314.
- Kuladee S, Prachason T. Development and validation of the Thai version of the 4 'A's' test for delirium screening in hospitalized elderly patients with acute medical illnesses. *Neuropsychiatric Disease and Treatment*. 2016;12:437–443.
- Leung JI, Leung Vc, 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. *General Hospital Psychiatry*. 2008;30(2):171–176.
- Lin HS, Eeles E, Pandey S, Pinsky D, Brasch C, Yerkovich S. Screening in delirium: a pilot study of two screening tools, the Simple Query for Easy Evaluation of Consciousness and Simple Question in delirium. *Australasian Journal on Ageing*. 2015;34(4):259–264.
- Lingehall HC, Smulter N, Engström KG, Gustafson Y, Olofsson B. Validation of the Swedish version of the Nursing Delirium Screening Scale used in patients 70 years and older undergoing cardiac surgery. *Journal of Clinical Nursing*. 2012;22(19–20):2858–2866.
- Luetz A, Heymann A, Radtke FM, Chenitir C, Neuhaus U, Nachtigall I, von Dossow W, Marz S, Eggers V, Heinz A, Wernecke KD, Spies CD. Different assessment tools for intensive care unit delirium: which score to use? *Critical Care Medicine*. 2010;38(2):409–418.
- Mitášová A, Mitáš L, Urbánek I, Ryba L, Hanke I, Ruber M, Michalčáková R, Košťálová M, Bednařík J. Incidence a rizikové faktory pooperačního deliria. *Česká a Slovenská Neurologie a Neurochirurgie*. 2012;75/108(5):574–580. (in Czech)
- Neufeld KJ, Leoutsakos JS, Sieber FE, Joshi D, Wanamaker BL, Rios-Robles J, Needham DM. Evaluation of two delirium screening tools for detecting post-operative delirium in the elderly. *British Journal of Anaesthesia*. 2013;111(4):612–618.
- Numan T, van den Boogaard M, Kamper AM, Rood PJT, Peelen LM, Slooter AJC; Dutch Delirium Detection Study Group. Recognition of delirium in postoperative elderly patients: a multicenter study. *Journal of the American Geriatrics Society*. 2017;65(9):1932–1938.
- Pečeňák J. Liečba delíria. *Neurologie pro praxi*. 2011;12(5):307–310. (in Slovak)
- Poikajärvi S, Salanterä S, Katajisto J, Junttila K. Validation of Finnish Neecham Confusion Scale and Nursing Delirium Screening Scale using confusion assessment method algorithm as a comparison scale. *BMC Nursing*. 2017;16:7.
- Radtke FM, Franck M, Schust S, Boehme L, Pascher A, Bail HJ, Seeling M, Luetz A, Wernecke KD, Heinz A, Spies CD. A comparison of three scores to screen for delirium on the surgical ward. *World Journal of Surgery*. 2010;34(3):487–494.
- Shulman RW, Kalra S, Jiang JZ. Validation of the Sour Seven Questionnaire for screening delirium in hospitalized seniors by informal caregivers and untrained nurses. *BMC Geriatrics*. 2016;16:44
- Schuermans MJ, Shortridge-Baggett LM, Duursma SA. The Delirium Observation Screening Scale: a screening instrument for delirium. *Research and Theory for Nursing Practice*. 2003;17(1):31–50.
- Sörensen Duppils GS, Johansson I. Predictive value and validation of the NEECHAM Confusion Scale using DSM-IV criteria for delirium as gold standard. *International Journal of Older People Nursing*. 2011;6(2):133–142.
- Spedale V, Di Mauro S, Del Giorno G, Barilaro M, Villa CE, Gaudreau JD, Ausili D. Delirium assessment in hospitalized elderly patients: Italian translation and validation of the nursing delirium screening scale. *Ageing Clinical and Experimental Research*. 2017;29(4):675–683.
- van de Steeg L, IJkema R, Langelaan M, Wagner C. Can an e-learning course improve nursing care for older people at risk of delirium: a stepped wedge cluster randomised trial. *BMC Geriatrics*. 2014;14:69.
- van Velthuisen EL, Zwakhalen SM, Warnier RM, Mulder WJ, Verhey FR, Kempen GI. Psychometric properties and feasibility of instruments for the detection of delirium in older hospitalized patients: a systematic review. *International Journal of Geriatric Psychiatry*. 2016;31(9):974–989.
- Wand AP, Thoo W, Sciuriaga H, Ting V, Baker J, Hunt GE. A multifaceted educational intervention to prevent delirium in older inpatients: a before and after study. *International Journal of Nursing Studies*. 2014;51(7):974–982.
- Wong CL, Holroyd-Leduc L, Simel DL, Straus SE. Does this patient have delirium?: value of bedside instruments. *JAMA*. 2010;304(7):779–786.
- Vörösová G. Použitie NEECHAM škály na posúdenie zmätenosti. In: *Ošetrovatel'stvo - teória, výskum a vzdelávanie*. Martin: Jesseniova lekárska fakulta; 2007. p. 641–648. [cited 2019 Apr 4]. Available from: <https://www.jfmed.uniba.sk/fileadmin/jlf/Pracoviska/ustav-osetrovatelstva/Konferencia-zbornik-program/2007-konferencia-fulltext-web.pdf> (in Slovak)