



# Möglichkeiten und Grenzen der Einschätzung des Delirrisikos: ein Scoping Review zu den Gütekriterien der Skala *Recognizing Acute Delirium As part of your Routine*

Petra Schumacher<sup>1</sup> · Marten Schmied<sup>2</sup> · Michael Schiller<sup>3</sup> · Philippe Voyer<sup>4</sup> · Gerhard Müller<sup>2</sup>

Eingegangen: 25. November 2020 / Angenommen: 28. Februar 2022  
© Der/die Autor(en) 2022

## Zusammenfassung

**Hintergrund** Die Diagnose eines Delirs beruht primär auf der Erfassung der klinischen Symptomatik mit akutem Beginn und fluktuierendem Verlauf. In der Literatur werden 5 pflegerische Screeninginstrumente zur Delirrisikoeinschätzung beschrieben, eines davon ist die Skala *Recognizing Acute Delirium As part of your Routine* (RADAR). Bisher wurde noch keine Literaturübersichtsarbeit zu den Gütekriterien der Skala durchgeführt.

**Ziel** Die Literatur zu RADAR zu sichten, um einen Überblick über die testtheoretischen und anwendungsbezogenen Gütekriterien aufzuzeigen.

**Methode** Das Scoping Review wurde in MEDLINE via PubMed sowie CINAHL und Academic Search Elite via EBSCOhost in der Zeit vom Juni 2019 bis Juli 2019 u. a. mit den Suchbegriffen *delir\**, *screening tool*, *psychometric properties* durchgeführt. Die Checkliste Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews mit ihren 9 Schritten wurde verwendet, um über die angewandte Methode dieses Scoping Review zu berichten.

**Ergebnis** Die Interrater-Reliabilität der RADAR variiert zwischen 82 und 92 % ( $\kappa=0,34-1$ ). Die konvergente Validität mit der Confusion Assessment Method liegt zwischen 36 und 85 % ( $\kappa=0,08-0,42$ ). Die prädiktive Validität wird mit einer Sensitivität von 54,4–100 % und einer Spezifität von 72–85,5 % angegeben ( $PPV=12,5-71\%$ ,  $NPV=94,2-100\%$ ). Die RADAR ist einfach zu verstehen und kann in weniger als 6 min ausgefüllt werden. Die Akzeptanz wird als hoch angegeben.

**Schlussfolgerung** Die RADAR-Skala scheint ein valides und akzeptiertes Instrument zur Delirrisikoeinschätzung zu sein. Für die deutschsprachige Version werden weitere Validierungsstudien empfohlen.

**Schlüsselwörter** Delirium · RADAR · Reliabilität · Validität · Skala

## Possibilities and limitations of assessing delirium risk: a scoping review of the scale's quality criteria *Recognizing Acute Delirium As part of your Routine*

### Abstract

**Background** A delirium diagnosis is primarily based on the recording of clinical symptoms with acute onset and fluctuating course. The international literature describes five different screening methods for delirium, one of which is the *Recognizing Acute Delirium As part of your Routine* (RADAR) scale. Although a German version of the scale is available, no literature review on its quality criteria has been published yet.

✉ Gerhard Müller  
gerhard.mueller@umit-tirol.at

<sup>1</sup> Department of Health Sciences, Institut für Pflegewissenschaft, IMC-Fachhochschule Krems, Krems, Niederösterreich, Österreich

<sup>2</sup> Department für Pflegewissenschaft und Gerontologie, Institut für Pflegewissenschaft, UMIT – Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik, Hall in Tirol, Tirol, Österreich

<sup>3</sup> Barmherzige Schwestern Krankenhaus Ried, Ried, Oberösterreich, Österreich

<sup>4</sup> Faculté des sciences infirmières, Université Laval, Québec, Kanada

**Aim** To review the literature on RADAR in order to provide an overview of its test-theoretical and application-related quality criteria.

**Method** A scoping review was conducted in MEDLINE via PubMed, CINAHL, and Academic Search Elite via EBSCO-Host between June and July 2019 using the search terms delir\*, screening tool, psychometric properties, among others. The Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews checklist with its nine steps was used to report on the method of this scoping review.

**Result** RADAR's interrater reliability varies from 82% to 92% ( $\kappa=0.34-1$ ). Its convergent validity with the confusion assessment method rates between 36% and 85% ( $\kappa=0.08-0.42$ ). The sensitivity ranges from 54.4% to 100% and the specificity from 72% to 85.5% ( $PPV=12.5-71\%$ ,  $NPV=94.2-100\%$ ). The RADAR is easy to understand and takes less than 6 min to complete. Acceptance is indicated as high.

**Conclusion** The RADAR scale seems to be an accepted and valid instrument for detecting a delirium risk. Further validation studies of the quality criteria are recommended for the German language version.

**Keywords** Delirium · RADAR · Reliability · Validity · Scale

## Einleitung

Obwohl die Symptome eines Delirs, auch als hirngorganisches Psychosyndrom oder Durchgangssyndrom bezeichnet, schon seit der Antike bekannt sind, werden sie besonders bei geriatrischen Patienten häufig übersehen oder missinterpretiert (Hewer et al. 2016; Paas 2017). Bei geriatrischen Patienten werden 30–60% der Delirsymptome nicht diagnostiziert, obwohl davon auszugehen ist, dass 11–25% der hospitalisierten Senioren bereits mit einem Delir aufgenommen werden und fast ein Drittel der Patientengruppe ein solches während des Krankenhausaufenthaltes entwickelt (Vasilevskis et al. 2012; Lechleitner 2013); noch höher ist die Prävalenz auf Intensivstationen (Thielscher et al. 2015). Dabei sind die Folgen eines Delirs weitreichend und u. U. tödlich. Neben der hohen Mortalität von 22–76% bei einem unbehandelten Delir haben Patienten einen um durchschnittlich 4,2 Tage verlängerten Krankenhausaufenthalt (Lechleitner 2013; Weinrebe et al. 2016). Auch poststationär zeigen 41% der Patienten kognitive Defizite, welche die Alltagsfähigkeiten einschränken und folglich zu einer erhöhten Inanspruchnahme von Unterstützungsleistungen führen (ÖGGG 2017). Nachdem schätzungsweise 30–40% der Delirien vermeidbar wären, sollte der frühzeitigen Risikoerfassung und gezielten Präventionsmaßnahmen eine hohe Priorität eingeräumt werden (Weinrebe et al. 2016). In der internationalen Literatur sind mehr als 20 Instrumente für ein Delirscreening beschrieben (Oh et al. 2017), von denen 5 für die Anwendung durch Pflegekräfte entwickelt wurden (Wetzlmair 2017). Die Skala Recognizing Acute Delirium As part of your Routine (RADAR) ist ein solches pflegerisches Screeninginstrument (Voyer et al. 2015).

## Die Recognizing Acute Delirium As part of your Routine

Die RADAR-Skala wurde 2015 in Kanada von Pflegewissenschaftlern und Fachkräften unter der Leitung von Voyer entwickelt (Voyer et al. 2015). Ziel war die Entwicklung einer Delirscreeningskala, die während pflegerischer Routinetätigkeiten einfach anwendbar ist. Die Skala umfasst lediglich 3 Items und kann in der Langzeit- sowie der Akutpflege bei Patienten mit oder ohne kognitive Einschränkungen eingesetzt werden (Voyer et al. 2015).

Die Items (a) *war der Patient schläfrig*, (b) *hatte der Patient Schwierigkeiten, Ihre Anweisungen zu befolgen*, und (c) *waren die Bewegungen des Patienten verlangsamt* werden während der Pflegehandlung mit *Ja/Nein* beurteilt. Um die Belastung für die Patienten möglichst gering zu halten und eine Verzerrung durch Testwiederholungen zu vermeiden, wurde die Skala so konzipiert, dass die 3 Items ohne eine direkte Patientenbefragung, vorherige Patientenkenntnis oder Einblick in die Dokumentation von den Pflegenden beantwortet werden können. Wenn ein Verhaltensmuster mit *Ja* beantwortet wird, besteht ein Delirrisiko und damit die Notwendigkeit einer weiteren Diagnostik (Voyer et al. 2015; Lohr 2017).

## Problembeschreibung

Das Vorhandensein verschiedener Instrumente zum Delirscreening macht die Entwicklung möglicher Entscheidungsstrategien in der klinischen Praxis notwendig. In der Literatur sind 5 pflegerische Screeninginstrumente zur Erfassung eines Delirrisikos beschrieben (Wetzlmair 2017). Von diesen Instrumenten verfügt die RADAR über die im Vergleich wenigsten Items und kürzeste Einschätzungszeit (Wetzlmair 2017). Ursprünglich wurde die Skala in englischer und französischer Sprache entwickelt. Im Jahr 2017

**Tab. 1** Ein- und Ausschlusskriterien der Literatursuche

	Einschlusskriterien	Ausschlusskriterien
Publikationszeitraum	Publikation ab 2015	Publikation vor 2015
Sprache	Literatur in deutscher, englischer und französischer Sprache	Andere Sprachen
Studiendesign und Publikationsart	Quantitative Studien	Qualitative Studien
Setting	Stationäre und ambulante Einrichtungen	–
Verwendungszweck	Arbeiten, welche die Güte- und Nebengütekriterien der Skala <i>Recognizing Acute Delirium As part of your Routine (RADAR)</i> untersuchen	Arbeiten, die die RADAR-Skala als Erhebungsinstrument im Rahmen einer klinischen Untersuchung verwenden

konnte sie ins Deutsche übersetzt und sprachlich an den österreichischen Kulturraum angepasst werden (RADAR-A) (Lohr 2017; Wetzlmair 2017). Bisher wurde noch keine Übersichtsarbeit zu den instrumenten- und anwendungsbezogenen Gütekriterien der RADAR verfasst.

## Zielsetzung und Fragestellung

Ziel der iterativ angelegten systematischen Literaturrecherche ist es, die Literatur zur RADAR-Skala zu sichten, um einen Überblick über die instrumenten- und anwendungsbezogenen Gütekriterien aufzuzeigen. Folgende Forschungsfragen wurden von der Zielsetzung abgeleitet:

1. Ist die RADAR ein valides und reliables Screeninginstrument zur Einschätzung des Delirrisikos?
2. Kann die RADAR als ein benutzerfreundliches Instrument bezeichnet werden?

## Methode

Das Scoping Review wurde in Anlehnung an die durch das Joanna Briggs Institute entwickelte Methodik (Peters et al. 2020) durchgeführt. Scoping Reviews geben einen schnellen Überblick über die vorhandene Literatur (Arksey und O'Malley 2005). Sie werden verwendet, um den Umfang und die Bandbreite der Forschungsaktivitäten als Vorstudie zu erfassen und Wissenslücken in der vorhandenen Literatur zu identifizieren. Das besondere Merkmal von Scoping Reviews ist die Identifizierung von Fragen und Forschungsthemen für zukünftige Forschung (Arksey und O'Malley 2005; Peters et al. 2020). Die Checkliste Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) mit ihren 9 Schritten wurde verwendet (s. Tab. 3 im Anhang), um über die angewandte Methode dieses Scoping Reviews zu berichten (Tricco et al. 2018; Peters et al. 2020).

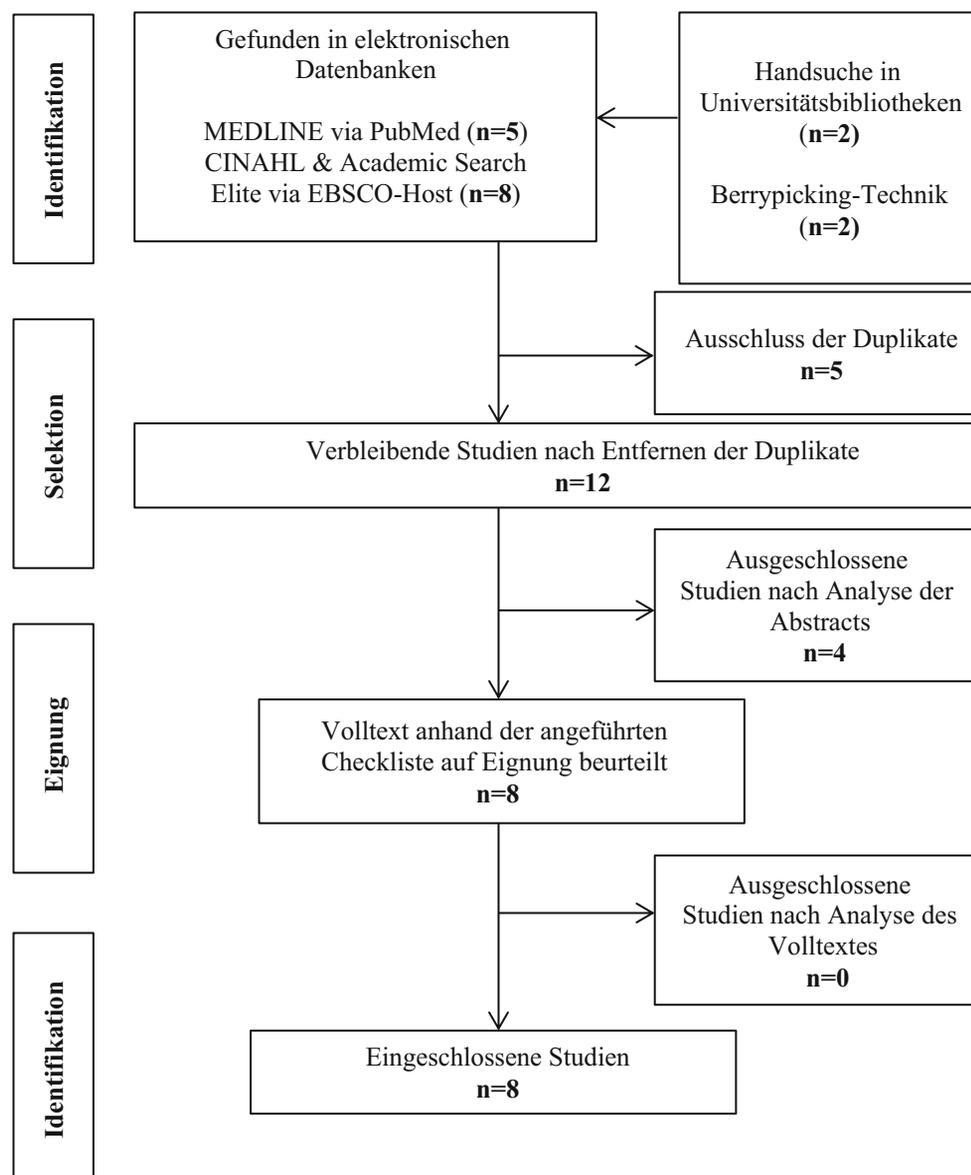
Im Juni 2019 und im Juli 2019 wurde in den Datenbanken MEDLINE via PubMed sowie CINAHL und Academic Search Elite via EBSCOhost nach geeigneter Literatur recherchiert. In den Datenbanken wurden die Suchbegriffe

*delir\**, *screening tool*, *instrument*, *test*, *scale*, *psychometric properties*, *validity*, *reliability* sowie in MEDLINE via PubMed die Medical Subject Headings [MeSH-Terms] bzw. in CINAHL via EBSCOhost die Medical Headings [MH] *Delirium*, *Risk Assessment*, *Psychometrics*, *Nonparametric Statistics* und *Reproducibility of Results* in unterschiedlichen Kombinationen mittels Bool-Operatoren zu Suchstrings verbunden. Beispielhaft wurde in MEDLINE via PubMed folgende Syntax verwendet (*delir\** OR “Delirium”[Mesh]) AND (“screening tool” OR instrument OR test OR scale OR “Risk Assessment”[Mesh]) AND (“psychometric properties” OR validity OR reliability OR “Psychometrics”[Mesh] OR “Statistics, Nonparametric”[Mesh] OR “Reproducibility of Results”[Mesh]) Filters: from 2015–2022. Zusätzliche wurden Literaturquellen durch eine Internetrecherche (Google Scholar), eine Handsuche in Universitätsbibliotheken sowie durch die Berrypicking-Technik (Bates 1989) identifiziert.

Die Literatursuche erfolgte anhand festgelegter Ein- und Ausschlusskriterien (Tab. 1). Durchgeführt wurde die Studienauswahl (Titel- und Abstract-Screening sowie Volltextscreening) von . Autoren (M.S., P.S.) unabhängig voneinander. Bei Diskrepanzen wurde eine Konsensfindung gesucht. Konnte keine erreicht werden, wurde die Abweichung durch die Entscheidung des dritten Autors (G.M.) gelöst.

Durch die iterativ angelegte Literatursuche konnten insgesamt 17 Studien identifiziert werden, die sich aus der Suche in Datenbanken ( $n = 13$ ), aus einer Handsuche ( $n = 2$ ) sowie durch die Berrypicking-Technik ( $n = 2$ ) zusammensetzten. Daraus wurden die Duplikate ( $n = 5$ ) entfernt und nach Analyse der Abstracts anhand der definierten Ein- und Ausschlusskriterien weitere Quellen ( $n = 4$ ) ausgeschlossen. Aufgrund der gewählten Methode wurde keine Bewertung der methodischen Limitation oder des Bias-Risikos der inkludierten Studien vorgenommen (Peters et al. 2020). Die Datenextraktion erfolgte in Excel 365. Aus den inkludierten Studien wurden folgende Charakteristika wie Autor(en), Jahr der Veröffentlichung, Land, Titel, Studiendesign, Studienziel, Setting, Population, Stichprobengröße, Einschluss- und Ausschlusskriterien und berichtete Ergebnisse zu instrumenten- bzw. anwendungsbezogenen Güte-

**Abb. 1** Flussdiagramm für den Scoping-Review-Prozess in Anlehnung an das PRISMA Statement (Peters et al. 2020)



kriterien extrahiert. Die eingeschlossenen Studien wurden zu deren Studiencharakteristika, instrumentenbezogenen Gütekriterien (interne Konsistenz, Interrater-Reliabilität, Inhaltsvalidität, konvergente und prädiktive Validität) sowie anwenderbezogenen Gütekriterien der RADAR-Skala (Praktikabilität, Akzeptanz) synthetisiert. Das in Abb. 1 gezeigte Flussdiagramm zeigt den Entscheidungsprozess bei der Literatursuche und Studienauswahl.

## Ergebnisse

### Studiencharakteristika

In dieses Scoping Review wurden 8 Studien eingeschlossen, aus deren Volltexten die Datenextraktion für die nach-

folgende Ergebnisdarstellung verwendet wurden. Insgesamt wurden 1891 Einschätzungen in den ausgewählten Studien mit der RADAR-Skala durchgeführt. Die Zahlen der in den Studien berücksichtigten Anwendungen reichen von wenigen ( $n=31$ ) (Bilodeau und Voyer 2017) bis zu sehr umfangreichen ( $n=514$ ) Beurteilungen (Voyer et al. 2015, 2016; Wetzlmair 2017). Im Durchschnitt sind rund 227 Einschätzungen pro Studie durchgeführt worden. Die Anzahl der in die Studien involvierten Pflegekräfte ist nicht immer dokumentiert, fluktuiert jedoch zwischen 139 (Voyer et al. 2015, 2016) und 41 (Pelletier et al. 2019) Pflegepersonen. In allen Studien waren diplomierte Gesundheits- und Krankenpflegepersonen die primäre Beurteilungsgruppe, während in 2 Studien auch Pflegeassistenten (Pelletier et al. 2019) und Ärzte (Wetzlmair 2017) miteinbezogen wurden. In 2 Studien (Voyer et al. 2015, 2016) wurden Patienten bzw. Be-

wohner in der Akut- und der Langzeitpflege ( $n=193$ ) eingeschlossen, während 3 Untersuchungen (Bilodeau und Voyer 2017; Lewallen und Voyer 2018; Pelletier et al. 2019) ausschließlich mit Klienten der Langzeitpflege ( $n=193$ ) und 3 mit Patienten im akutstationären Bereich (Voyer et al. 2017; Lohr 2017; Wetzlmair 2017) ( $n=857$ ) durchgeführt wurden.

Bezüglich der Reliabilität der RADAR wurde in einer Studie (Lohr 2017) die interne Konsistenz der Skala und in 5 Studien (Voyer et al. 2015, 2017; Bilodeau und Voyer 2017; Lohr 2017; Lewallen und Voyer 2018) die Interrater-Reliabilität ermittelt. Die Inhaltsvalidität wurde in einer Studie (Wetzlmair 2017) und die konvergente Validität wurde in 2 (Voyer et al. 2015, 2016) Studien ermittelt. Fünf Studien (Voyer et al. 2015, 2016, 2017; Bilodeau und Voyer 2017; Pelletier et al. 2019) treffen Aussagen zur prädiktiven Validität (Sensitivität, Spezifität, positive [PPV] und negative Vorhersagewerte [NPV]) der RADAR-Skala. In 5 Studien (Voyer et al. 2015; Bilodeau und Voyer 2017; Lohr 2017; Lewallen und Voyer 2018; Pelletier et al. 2019) wurden, als Indikator der Praktikabilität der RADAR-Skala, die durchschnittliche Erhebungsdauer sowie die Akzeptanz dieses Instrumentes anhand von Fragebogen erhoben. Tab. 2 gibt eine Übersicht über die eingeschlossenen Studien und deren zentralen Ergebnisse.

### Instrumentenbezogene Gütekriterien der RADAR-Skala

Eine angemessene interne Konsistenz wurde für die Items 2 und 3 berechnet (*Kuder-Richardson-Koeffizienten-20* [ $KR$ ]=0,498); eine Berechnung für Item 1 war nicht möglich, da keine Abweichung nachgewiesen wurde (Lohr 2017). Die Untersuchungen zur Interrater-Reliabilität der RADAR zeigten Beobachtungsübereinstimmungen von 82–98 % ( $\kappa=0,34$ –0,79) (Voyer et al. 2015) bzw. 94,2–99 % ( $\kappa=0,76$ –1) (Bilodeau und Voyer 2017). Lewallen und Voyer (2018) konnten eine erhebliche Übereinstimmung zwischen den Einschätzern ( $\kappa=0,63$ ) feststellen, während 2 Studien eine Gesamtübereinstimmung von 89 % ( $\kappa=0,46$ ) (Voyer et al. 2017) und 90 % (Lohr 2017) angaben.

Im Rahmen der Übersetzung der RADAR wurde die Inhaltsvalidität bei Angehörigen verschiedener Gesundheitsberufe ( $n=137$ ) getestet. Mit einem Item Content Validity Index (I-CVI) von 82–85 % sowie mit den Werten der zufallskorrigierten Übereinstimmung über alle Berufsgruppen wurden diese mit gut bewertet (Ärzte:  $\kappa=0,81$ –0,95; Pflegepersonen:  $\kappa=0,80$ –0,89) (Wetzlmair 2017). Die konvergente Validität wurde zwischen der RADAR-Skala und der Confusion Assessment Method [CAM] getestet und erreichte moderate bis zufriedenstellende Werte (52–85 % (Voyer et al. 2015) bzw. 36–85 % ( $\kappa=0,08$ –0,42) (Voyer et al. 2016)).

Weitgehend gute Ergebnisse erzielte die RADAR-Skala auch in den Untersuchungen der prädiktiven Validität, wobei das Vorhandensein eines Delirrisikos mit Ergebnissen aus der CAM und in einer Studie mit den Kriterien des DSM-IV TR (Voyer et al. 2015) verglichen wurde. Die Sensitivität der RADAR-Skala variierte, abhängig von dem beurteilten Item, von 17,6–54,4 % (Voyer et al. 2016) bzw. 65,2 % (Voyer et al. 2015) bis zu 100 % (Bilodeau und Voyer 2017; Pelletier et al. 2019), wobei der positive Vorhersagewert gering war (PPV = 12,50–19 %) (Voyer et al. 2015, 2017; Bilodeau und Voyer 2017; Pelletier et al. 2019). Je nach Item wurde die Spezifität mit Werten von 76,3 % (Voyer et al. 2015) bis 96,8 % (Voyer et al. 2016) angegeben. Wurde ein positives Delirrisiko ausschließlich mit dem Vorhandensein von Unaufmerksamkeit und veränderten Bewusstseinszuständen (*the sixth vital sign*) verglichen, erreichte die Skala einen PPV von 59,1–71 % (Voyer et al. 2016). Die negative Vorhersagewerte in den weiterführenden Studien waren hervorragend mit bis zu 81,7 % (Voyer et al. 2016) resp. 94,20 % (Voyer et al. 2015) bzw. 100 % (Bilodeau und Voyer 2017; Voyer et al. 2017; Pelletier et al. 2019).

### Anwenderbezogene Gütekriterien der RADAR-Skala

Die Praktikabilität der RADAR-Skala wurde anhand der durchschnittlichen Erhebungsdauer, die weniger als 1 min Zeit beanspruchen sollte (Voyer et al. 2015), beurteilt. Mit durchschnittlich 7,2–53 s fielen die Dauern der Beobachtung in den Studien sehr ähnlich aus (Voyer et al. 2015; Bilodeau und Voyer 2017; Lewallen und Voyer 2018; Pelletier et al. 2019), während Lohr (2017) für die Durchführung der deutschen Version eine auffallend längere durchschnittliche Dauer ( $\bar{x}=5,25$ ,  $SD \pm 3,20$  min) dokumentierte.

Um Aussagen zur Akzeptanz der RADAR-Skala zu treffen, nutzten Lewallen und Voyer (2018) denselben Fragebogen wie schon Voyer et al. (2015) und kamen mit 80–91 % gegenüber 94–99 % positiven Bewertungen zu niedrigeren Ergebnissen. Die Skala wurde als einfach aufgebaut und schnell durchführbar beurteilt (Bilodeau und Voyer 2017). Die Medikamentengabe hielten 99 % der Befragten für einen guten Beobachtungszeitraum (Voyer et al. 2015).

### Diskussion

Ziel dieses Scoping Review war, einen Überblick über die zur RADAR-Skala vorliegenden Studien und die darin erhobenen testtheoretischen Gütekriterien zu erlangen, um, wie in der Literatur empfohlen (Peters et al. 2020), Hinweise für zukünftige Forschungsarbeiten geben zu können.

**Tab. 2** Überblick zu den Ergebnissen der eingeschlossenen Studien

Autor (Jahr) Land Setting	Stichprobe	Reliabilität		Validität		Prädiktive Validität			Nebengütekriterium		
		Interne Konsistenz	Intrater- Reliabilität	Konvergenz Validität [Übereinstimmung RADAR-Items mit korrespondierenden CAM-Symptomen]	Inhaltsvalidität	Sensitivität	Spezifität	PPV	NPV	Praktika- bilität	Akzeptanz
Voyer et al. (2015) Kanada Akut- und Langzeitpflege	Pflegepersonen n = 139 Patienten/ Bewohner n = 193 RADAR-Anwen- dungen n = 541 <sup>a</sup>	-	82–98 % (κ = 0,34–0,79)	52–85 %	-	65,2 % (95 %-KI [42,7; 83,6])	76,3 % (95 %-KI [69,2; 82,5])	19 %	94,20 %	7,2 sec.	100 % „verständlich“ 99 % „genügend Hinter- grundwissen“ 99 % „Zeitpunkt günstig“ 4 % „zeitaufwendig“
Voyer et al. (2016) Kanada Akut- und Langzeitpflege	Pflegepersonen n = 139 Patienten/ Bewohner n = 193 RADAR-Anwen- dungen n = 541 <sup>a</sup>	-	-	36–85 % (κ = 0,08–0,42)	-	Abhängig von Item und Häufigkeit der Beurteilungen: 17,6–70,4 % <sup>c</sup>	Abhängig von Item und Häufigkeit der Beurteilungen: 60–96,8 % <sup>c</sup>	Abhängig von Item und Häufigkeit der Beurteilungen: 52,9–90,7 % <sup>c</sup>	Abhängig von Item und Häufigkeit der Beurteilungen: 59,1–76 % <sup>c</sup>	-	-
Bilodeau und Voyer et al. (2016) Kanada Langzeitpflege	Bewohner n = 31 RADAR Anwen- dungen n = 31	-	94,2–99 % (κ = 0,76–1)	-	-	100 %	77 %	12,50 %	100 %	22 sec.	87–100 % „einfach aufgebaut“ und „schnell durchgeführt“
Voyer et al. (2017) Kanada Akupflege	Patienten n = 57 RADAR-Anwen- dungen n = 256	-	89 % (κ = 0,46)	-	-	100 % (95 %-KI [2,5; 100]) <sup>c</sup>	89,3 % (95 %-KI [78,1; 96]) <sup>c</sup>	14,3 % <sup>c</sup>	100 % <sup>c</sup>	-	-
Lohr (2017) Österreich Akut statio- närer Bereich	Patienten/RADAR- Anwendungen n = 60 Vier Beobachter- paare n = 8	KR = 0,498 Berechnung von Items 2 und 3 (Homo- genität)	Gesamt: 90 % Itemebene: Po = 92,3–100 % ACI = 0,91–1,00 (κ = -0,025–0)	-	-	-	-	-	-	5,25 min (SD ± 3,20)	77,8 % „verständlich“ 50 % „kann Unterstützung sein“
Wetzlmair (2017) Österreich Akut statio- närer Bereich	DGKP, PA <sup>b</sup> und Ärzte n = 137 RADAR-Anwen- dungen n = 541	-	-	-	Über alle Berufsgruppen (κ = 0,82–0,85) (I- CVI = 0,82–0,85)	-	-	-	-	-	-
Lewallen und Voyer (2018) USA Langzeitpflege	Pflegkräfte n = 60 RADAR-Anwen- dungen n = 117	-	(κ = 0,63)	-	-	-	-	-	-	15 sec.	Derselbe Fragebogen wie 2014; insgesamt schlech- ter mit 80–91 %iger Akzeptanz, 20 % „zeitauf- wendig“
Pelletier et al. (2019) Kanada Langzeitpflege	Bewohner n = 45 Pflegeassistent n = 41 RADAR-Anwen- dungen n = 45	-	-	-	-	100 %	72 %	14 %	100 %	53 sec.	94–100 %ige Akzeptanz und Handhabbarkeit

<sup>a</sup>Die Studien verwenden dieselbe Stichprobe

<sup>b</sup>Diplomierter Gesundheits- und Krankenpflegepersonal, Pflegeassistenten

<sup>c</sup>Gemessen am Vorhandensein von Unaufmerksamkeit und eines veränderten Bewusstseinszustandes („sixth vital sign“)

Acht Studien aus dem Langzeit- und Akutsetting konnten in das Scoping Review aufgenommen werden.

Die vorliegenden Daten aus den Studienergebnissen sind unter Berücksichtigung der unterschiedlichen Anzahlen an eingeschlossenen Patienten und Pflegekräften in den jeweiligen Untersuchungen zu interpretieren, da sich die empfohlenen Stichprobengrößen je nach untersuchtem Gütekriterium in der Literatur unterscheiden. Die Empfehlungen für den Stichprobenumfang von Reliabilitätsstudien sind sehr unterschiedlich und reichen von 200 bis zu über 1000 in manchen Fällen (Streiner und Kottner 2014). Jedoch argumentiert Cicchetti (2001), dass sich Stichprobengrößen über 50 kaum lohnen, weil ein Zuwachs an Präzision selten auftritt. Jede Stichprobengröße unter diesem Wert muss begründet werden (Streiner und Kottner 2014). Zur Berechnung der internen Konsistenz wurde bei einer Anzahl von 2 oder 3 Items mit dichotomer Antwortmöglichkeit eine Stichprobengröße von 23 Probanden herangezogen, um Cronbachs  $\alpha$  von 0,62 zu erreichen (Peterson 1994). Die Berechnung der internen Konsistenz kann auch nach Kuder-Richardson-Formel erfolgen (Kuder und Richardson 1937), wobei die Stichprobengröße sich nicht wesentlich von Cronbachs  $\alpha$  unterscheidet. Die  $\alpha$ -Koeffizienten werden höher bei steigender Itemanzahl (Döring und Bortz 2016). Die Größe der Stichprobe ist zur Bestimmung der Interrater-Reliabilität von der gewählten Fehlerquote sowie der tatsächlichen Übereinstimmungswahrscheinlichkeit [pa] minus der zufälligen Übereinstimmungswahrscheinlichkeit [pe] abhängig und sollte möglichst gering sein (Gwet 2010). Beispielsweise würde sich bei einer Fehlerquote von 20% und einem  $pa$  minus  $pe$  von 0,4 eine Stichprobengröße von 156 ergeben (Gwet 2010). Der geschätzte Stichprobenumfang zur Testung der Validität ist variabler, weil sie keinem einheitlichen Design unterliegen. Je nach Forschungsdesign wird daher eine andere Stichprobengröße benötigt, welche mit unterschiedlichen und frei im Internet verfügbaren Softwares zur Berechnung des Stichprobenumfangs a priori bestimmt werden kann (Streiner und Kottner 2014). Beispielsweise werden für die Inhaltsvalidität mindestens 3 Bewerter empfohlen, wobei eine größere Gruppe vorzuziehen ist (Polit und Beck 2017, S. 311). In den inkludierten Studien waren nur 31 (Bilodeau und Voyer 2017) bis 193 (Voyer et al. 2015, 2016) Patienten sowie nur 8 (Lohr 2017) bis 139 (Voyer et al. 2015, 2016) rekrutierende Pflegekräfte an den Studien beteiligt. Schließlich schloss nur eine Autorin (Wetzlmair 2017) neben Pflegekräften auch Ärzte in die Beurteilung der RADAR-Skala mit ein.

Die interne Konsistenz wurde ausschließlich für die Items 2 und 3 ( $KR=0,498$ ) der deutschen RADAR-A erhoben und ist gerade noch als akzeptabel zu werten. Sie konnte für Item 1 nicht berechnet werden, da Item 1 keine Varianz aufzeigte (Lohr 2017). Der empfohlene

Cut-off-Wert für die Kuder-Richardson-Formel liegt bei über 0,5 und gilt dann als angemessen (McGahee und Ball 2009). Die Angaben zur Interrater-Reliabilität der RADAR (Voyer et al. 2015, 2017; Bilodeau und Voyer 2017; Lewallen und Voyer 2018) bzw. der RADAR-A (Lohr 2017) sind insgesamt zufriedenstellend. Allgemein werden Cohens-Kappa-Werte [ $\kappa$ ] mithilfe des Interpretationsschemas nach Landis und Koch (1977) ( $\kappa < 0,00$  keine Übereinstimmung;  $\kappa = 0,00-0,20$  sehr geringe Übereinstimmung;  $\kappa = 0,21-0,40$  geringe Übereinstimmung;  $\kappa = 0,41-0,60$  mittlere Übereinstimmung;  $\kappa = 0,61-0,80$  hohe Übereinstimmung;  $\kappa = 0,81-1,00$  sehr hohe Übereinstimmung) bewertet. Die Ausprägungen der  $\kappa$ -Werte hängen einerseits von der Datenverteilung und den Kategorien des Instrumentes und andererseits von dem Verhalten der Beobachter ab (Asendorpf und Wallbott 1979; Wirtz und Caspar 2002). Zwei Studien bezogen auch Pflegeassistenten in die Beurteilung der RADAR-Skala (Pelletier et al. 2019) bzw. RADAR-A-Skala (Wetzlmair 2017) mit ein, sie enthalten jedoch keine Angaben zur Reliabilität der Skalen in diesem Testsetting.

Die Inhaltsvalidität der RADAR-A ( $\kappa = 0,82-0,85$ ;  $I-CVI = 82-85\%$ ) wurde in einer Studie (Wetzlmair 2017) erhoben und ist als zufriedenstellend bzw. gut zu beurteilen, nachdem der berechnete  $I-CVI$  größer als 0,78 (Polit et al. 2007) und die  $\kappa$ -Werte größer als 0,74 waren (Cicchetti und Sparrow 1981). Die konvergente Validität der RADAR (36–85%;  $\kappa = 0,08-0,42$ ) wurde in 2 Studien (Voyer et al. 2015, 2016) untersucht und kann als akzeptabel (0,40–0,59) nach dem Interpretationsschema von Cicchetti und Sparrow (1981) angesehen werden. Größere Unterschiede zeigten sich in der Beurteilung der prädiktiven Validität der RADAR-Skala. Während die Sensitivität der RADAR mit 65,2% (Voyer et al. 2015) oder abhängig von dem Item und der Häufigkeit der Anwendung mit 17,6–70,4% (Voyer et al. 2016) angegeben wurde, fanden 3 Studien (Bilodeau und Voyer 2017; Voyer et al. 2017; Pelletier et al. 2019) eine Sensitivität von 100%. Somit liegen nur die letztgenannten 3 Studien über dem empfohlenen Sensitivitätswert von 80% für die Praxis (Behrens und Langer 2016, S. 237). Auffällig ist auch, dass in der Studie von Voyer et al. (2016) der positive Vorhersagewert mit 52,9–90,7% angegeben wurde, während die anderen 4 Studien diesen mit nur 12,5–19% (Voyer et al. 2015, 2017; Bilodeau und Voyer 2017; Pelletier et al. 2019) bezifferten. Andersherum ist der negative Vorhersagewert in der Studie von Voyer et al. (2016) mit 59,1–76% geringer als in den Vergleichsstudien mit 94,2% (Voyer et al. 2015) bzw. 100% (Bilodeau und Voyer 2017; Voyer et al. 2017; Pelletier et al. 2019). Beide Vorhersagewerte unterscheiden sich je nach untersuchter Stichprobe aufgrund der erhobenen Prävalenz. Somit können die diesbezüglich angeführten Werte nicht auf andere Populationen mit unterschiedlicher

Prävalenzen übertragen werden (Behrens und Langer 2016, S. 239).

Der Unterschied in der durchschnittlichen Anwendungsdauer von 5,25 min ( $SD \pm 3,20$ ) für die deutsche Version (Lohr 2017) und den 7,2–53 sec. (Voyer et al. 2015; Pelletier et al. 2019) für die englische bzw. französische Version der RADAR ist auffallend, die Ursache ist jedoch unbekannt. Schließlich hat die RADAR-Skala insgesamt gute Beurteilungen bezüglich ihrer Akzeptanz. Bei der Beurteilung der RADAR-A gab hingegen nur die Hälfte der Befragten an, die Skala könne eine Unterstützung sein (Lohr 2017). Außerdem wurde derselbe Fragebogen von Voyer et al. (2015) sowie Lewallen und Voyer (2018) verwendet, mit dem Ergebnis, dass die Akzeptanzbewertung in der Studie von Lewallen und Voyer (2018) insgesamt weniger gut ausgefallen ist.

## Schlussfolgerungen und Ausblick

Die RADAR-Skala stellt ein valides, zuverlässiges, in der pflegerischen Praxis praktikables, ressourcenschonendes und von den Durchführenden akzeptiertes Instrument zum Delirscreening bei Patienten mit und ohne kognitive Einschränkungen dar, das sowohl im Akutsetting als auch der Langzeitpflege angewendet werden könnte. Diese Empfehlung gilt unter der Einschränkung, dass es für die englische RADAR bisher keine Erhebung der internen Konsistenz und der Inhaltsvalidität gibt.

Da für die deutsche RADAR-A noch kein Wissen zur konvergenten oder zur prädiktiven Validität existiert, kann diese nicht uneingeschränkt als pflegerisches Screeninginstrument für die Praxis empfohlen werden. Überprüfungen dahingehend sollen sich auf die akutstationäre sowie auf die poststationäre Versorgung beziehen. Schließlich sollte getestet werden, ob die RADAR-A auch bei der Anwendung durch Pflegefachassistenten valide Ergebnisse liefert, und ob sich die durchschnittliche Erhebungszeit nach einer sprachlichen Anpassung der Items reduziert.

## Anhang

Section	Item	PRISMA-ScR Checklist Item	Reported on page #
<i>Title</i>			
Title	1	Identify the report as a scoping review	1
<i>Abstract</i>			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives	2–3
<i>Introduction</i>			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach	4
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives	5
<i>Methods</i>			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number	Not done
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale	6
Information sources <sup>a</sup>	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed	5
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated	5
Selection of sources of evidence <sup>b</sup>	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review	6–7

Section	Item	PRISMA-ScR Checklist Item	Reported on page #
Data charting process <sup>c</sup>	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators	6
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made	6
Critical appraisal of individual sources of evidence <sup>d</sup>	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate)	Not done
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted	8
<i>Results</i>			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram	6–7
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations	8
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12)	Not applicable
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives	8–11
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives	8–11
<i>Discussion</i>			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups	12
Limitations	20	Discuss the limitations of the scoping review process	Not done
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps	13
<i>Funding</i>			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review	Not applicable

JB1 Joanna Briggs Institute, PRISMA-ScR Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

<sup>a</sup>Where *sources of evidence* (see second footnote) are compiled from, such as bibliographic databases, social media platforms, and Web sites

<sup>b</sup>A more inclusive/heterogeneous term used to account for the different types of evidence or data sources (e.g., quantitative and/or qualitative research, expert opinion, and policy documents) that may be eligible in a scoping review as opposed to only studies. This is not to be confused with *information sources* (see first footnote)

<sup>c</sup>The frameworks by Arksey and O'Malley (2005) and Levac et al. (7) and the JBI guidance (4, 5) refer to the process of data extraction in a scoping review as data charting

<sup>d</sup>The process of systematically examining research evidence to assess its validity, results, and relevance before using it to inform a decision. This term is used for items 12 and 19 instead of „risk of bias“ (which is more applicable to systematic reviews of interventions) to include and acknowledge the various sources of evidence that may be used in a scoping review (e.g., quantitative and/or qualitative research, expert opinion, and policy document)

**Funding** Open access funding provided by UMIT - Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik GmbH

**Interessenkonflikt** P. Schumacher, M. Schmied, M. Schiller, P. Voyer und G. Müller geben an, dass kein Interessenkonflikt besteht.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link

zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

## Literatur

- Arksey H, O'Malley L (2005) Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 8(1):19–32
- Asendorpf J, Wallbott HG (1979) Maße der Beobachterübereinstimmung: Ein systematischer Vergleich. *Z Sozialpsychol* 10(3):243–252
- Bates M (1989) The design of browsing and berrypicking techniques for the online search interface. *Online Inf Rev* 13(5):407–424
- Behrens J, Langer G (2016) Evidence-based nursing and caring. Hogrefe, Göttingen
- Bilodeau C, Voyer P (2017) Radar: un outil valide pour le repérage du syndrome confusionnel aigu (delirium) en résidences intermédiaires. *NPG Neurol Psychiatr Geriatr* 17(98):144–151. <https://doi.org/10.1016/j.npg.2016.04.004>
- Cicchetti DV (2001) The precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. *J Clin Exp Neuropsychol* 23(5):695–700
- Cicchetti D, Sparrow SA (1981) Developing criteria for establishing interrater reliability of specific items: application to assessment of adaptive behavior. *Am J Intellect Dev Disabil* 86(2):127–137
- Döring N, Bortz J (2016) Forschungsmethoden und Evaluation in der Sozial- und Humanwissenschaften, 5. Aufl. Springer, Berlin, Heidelberg
- Gwet KL (2010) Sample Size Determination. [http://agreestat.com/blog\\_irr/sample\\_size\\_determination.html](http://agreestat.com/blog_irr/sample_size_determination.html)\*. Zugegriffen: 15. Aug. 2020
- Hewer W, Thomas C, Drach LM (2016) Delir beim alten Menschen. Grundlagen – Diagnostik – Therapie – Prävention, 1. Aufl. Kohlhammer, Stuttgart
- Kuder GF, Richardson MW (1937) The theory of the estimation of test reliability. *Psychometrika* 2(3):151–160
- Landis J, Koch G (1977) The measurement of observer agreement for categorical data. *Biometrics* 33(1):159–174
- Lechleitner M (2013) Verwirrheitszustände im Alter. *Osterr Arzteztg* 23(24):22–29
- Lewallen K, Voyer P (2018) Delirium knowledge improvement and implementation of the RADAR screening tool in two skilled nursing facilities. *Ann Longterm Care* 12:19–24. <https://doi.org/10.25270/altc.2018.12.00050>
- Lohr P (2017) Delir-Screening mit RADAR-A: Überprüfung der Reliabilität und Praktikabilität. Masterarbeit. UMIT – Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik, Hall in Tirol
- McGahee TW, Ball J (2009) How to read and really use an item analysis. *Nurse Educ* 34(4):166–171
- Oh ES, Fong TG, Hsieh TT, Inouye SK (2017) Delirium in older persons: advances in diagnosis and treatment. *JAMA* 318(12):1161–1174. <https://doi.org/10.1001/jama.2017.12067>
- Österreichische Gesellschaft für Geriatrie und Gerontologie (ÖGGG) (2017) Delir 2017 – Ein häufiges Syndrom im Alter – eine interdisziplinäre Herausforderung. [http://www.alterspsychiatrie.at/bilder/publikationen/expertpapiere/2017\\_Delir\\_Folder.pdf](http://www.alterspsychiatrie.at/bilder/publikationen/expertpapiere/2017_Delir_Folder.pdf). Zugegriffen: 2. Aug. 2019
- Paas S (2017) Delir in der Notaufnahme des Krankenhauses der Barmherzigen Brüder in Trier. *HBSscience* 9(1/2):28–35. <https://doi.org/10.1007/s16024-017-0310-2>
- Pelletier I, Voyer P, Cyr N, Carmichael PH (2019) Outil RADAR: pour une détection efficace des signes du delirium en CHSLD. *Perspect Infirm* 16(3):68–73
- Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H (2020) Chapter 11: Scoping reviews. In: Aromataris E, Munn Z (Hrsg) *JBIM manual for evidence synthesis* <https://doi.org/10.46658/JBIMES-20-12>
- Peterson RA (1994) A meta-analysis of Cronbach's coefficient alpha. *J Consum Res* 21(2):381–391. <https://doi.org/10.1086/209405>
- Polit DF, Beck CT (2017) *Nursing research*, 10. Aufl. Wolters Kluwer, Philadelphia, Baltimore, New York, London, Buenos Aires, Hong Kong, Sydney, Tokyo
- Polit DF, Beck CT, Owen SV (2007) Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health* 30(4):359–467. <https://doi.org/10.1002/nur>
- Streiner DL, Kottner J (2014) Recommendations for reporting the results of studies of instrument and scale development and testing. *J Adv Nurs* 70(9):1970–1979
- Thielscher S, Contreras C, Bork-Kopp R, Fani M, Gather W, Hardt R, Heckmann J, Rix E, Wöhrli P, Fellgiebel A (2015) Das Expertenforum Demenz: Delir im Krankenhaus – ein immer noch unterschätztes Problem. *Ärztebl Rheinl Pfalz* 10(12):26–28
- Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D et al (2018) PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med* 169:467–473. <https://doi.org/10.7326/M18-0850>
- Vasilevskis E, Han J, Hughes C, Ely EW (2012) Epidemiology and risk factors for delirium across hospital settings. *Best Pract Res Clin Anaesthesiol* 26(3):277–287. <https://doi.org/10.1016/j.bpa.2012.07.003>
- Voyer P, Champoux N, Desrosiers J, Landreville P, McCusker J, Monette J, Savoie M, Sylvie R, Charmichel PH (2015) Recognizing acute delirium as part of your routine RADAR: a validation study. *BMC Nurs* 14(19):1–13. <https://doi.org/10.1186/s12912-015-0070-1>
- Voyer P, Champoux N, Desrosiers J, Landreville P, McCusker J, Monette J, Savoie M, Charmichel PH, Richard H, Richard S (2016) RADAR: a measure of the sixth vital sign? *Clin Nurs Res* 25(1):9–29. <https://doi.org/10.1177/1054773815603346>
- Voyer P, Émond M, Boucher V, Carmichael PH, Juneau L, Richard H, Tuong Minh Vu T, Lee J, Bouchard G (2017) RADAR: A rapid detection tool for signs of delirium (6th vital sign) in emergency departments. *CJEN* 4(2):37–43. <https://doi.org/10.29173/cjen79>
- Weinreb W, Johannsdottir E, Karaman M, Fügen I (2016) What does delirium cost? An economic evaluation of hyperactive delirium. *Z Gerontol Geriatr* 49(1):52–58. <https://doi.org/10.1007/s00391-015-0871-6>
- Wetzlmair J (2017) Delir-Screening mit RADAR-A: Übersetzung und Bestimmung der Inhaltsvalidität. Masterarbeit. UMIT – Private Universität für Gesundheitswissenschaften, Medizinische Informatik und Technik, Hall in Tirol
- Wirtz M, Caspar F (2002) Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. Hogrefe, Göttingen

**Hinweis des Verlags** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.