REVIEW ARTICLE

Detecting Delirium: A Systematic Review of Identification Instruments for Non-ICU Settings

Benjamin K. I. Helfand, MSc,† Madeline L. D'Aquila, BA,‡ Patricia Tabloski, PhD,§ Kristen Erickson, MPH,‡ Jirong Yue, MD,¶ Tamara G. Fong, MD, PhD,‡‖ Tammy T. Hshieh, MD, MPH,‡†† Eran D. Metzger, MD,** Eva M. Schmitt, PhD,‡ Edwin D. Boudreaux, PhD,* Sharon K. Inouye, MD, MPH,‡ and Richard N. Jones, ScD†*

BACKGROUND/OBJECTIVES: Delirium manifests clinically in varying ways across settings. More than 40 instruments currently exist for characterizing the different manifestations of delirium. We evaluated all delirium identification instruments according to their psychometric properties and frequency of citation in published research.

DESIGN: We conducted the systematic review by searching Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Library, Excerpta Medica Database (Embase), PsycINFO, PubMed, and Web of Science from January 1, 1974, to January 31, 2020, with the keywords "delirium" and "instruments," along with their known synonyms. We selected only systematic reviews, meta-analyses, or narrative literature reviews including multiple delirium identification instruments.

MEASUREMENTS: Two reviewers assessed the eligibility of articles and extracted data on all potential delirium identification instruments. Using the original publication on each instrument, the psychometric properties were examined using the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) framework.

RESULTS: Of 2,542 articles identified, 75 met eligibility criteria, yielding 30 different delirium identification instruments. A count of citations was determined using Scopus for the original publication for each instrument. Each instrument underwent methodological quality review of psychometric properties using COSMIN definitions. An expert panel categorized key domains for delirium identification based on criteria from the *Diagnostic and Statistical Manual of Mental Disorders* (DSM)-III through DSM-5. Four instruments were notable for having at least two of three of the following: citation count of 200 or more, strong validation methodology in their original publication, and fulfillment of DSM-5 criteria. These were, alphabetically, Confusion Assessment Method, Delirium Observation Screening Scale, Delirium Rating Scale-Revised-98, and Memorial Delirium Assessment Scale.

CONCLUSION: Four commonly used and well-validated instruments can be recommended for clinical and research use. An important area for future investigation is to harmonize these measures to compare and combine studies on delirium. **J Am Geriatr Soc 00:1-9, 2020.**

Keywords: delirium; measurement; systematic review; psychometrics

From the *Department of Emergency Medicine, University of Massachusetts Medical School, Worcester, Massachusetts; †Departments of Psychiatry and Human Behavior and Neurology, Warren Alpert Medical School, Brown University, Providence, Rhode Island; ‡Aging Brain Center, Marcus Institute for Aging Research, Hebrew SeniorLife, Harvard Medical School, Boston, Massachusetts; §William F. Connell School of Nursing, Boston College, Boston, Massachusetts; ¶Department of Geriatrics, West China Hospital, Sichuan University, Chengdu, China; ‖Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts; **Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, Massachusetts; and the ††Department of Medicine, Division of Gerontology, Brigham and Women's Hospital, Boston, Massachusetts.

Address correspondence to *Benjamin K. I. Helfand, MSc, Department of Emergency Medicine, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655. E-mail: benjamin.helfand@umassmed.edu,

DOI: 10.1111/jgs.16879

INTRODUCTION

Delirium is a major public health problem, impacting an estimated 2.6 million older Americans annually and accounting for more than $164 billion in healthcare expenditures.[1] Delirium disproportionately affects people aged 65 and older and is associated with prolonged hospitalization, cognitive decline, and heightened risks for dementia and death.[2,3] Clinically, many cases of delirium

0002-8614/20/$15.00

go unrecognized,[4] representing missed opportunities for its prevention.[5] A study revealed that in 61% of hospitalized patients with confirmed delirium by a palliative care expert, the diagnosis was missed by the primary referring team.[6] At least in part, the lack of a unified, accepted diagnostic approach adds to the challenges of recognition.[7]

The growing awareness of the seriousness of delirium, coupled with the fact that it remains a purely clinical diagnosis, without a laboratory test, has resulted in many tools for its detection. Currently, there are more than 40 delirium instruments for different purposes (e.g., screening, diagnosis, and severity), targeting different clinical settings (e.g., intensive care unit [ICU], emergency department, medical wards) and intended for different users (e.g., psychiatrists, geriatricians, nurses). These instruments describe varying domains of delirium. This overabundance of instruments makes direct comparisons or interpretation of results across studies challenging.

Our overall goal was to examine instruments used for identification of delirium, defined as those used for screening or diagnosis. We aimed to conduct a comprehensive systematic review to identify the most commonly used and originally well-validated instruments for the identification of delirium.

METHODS

Our approach involved five steps. First, we performed a comprehensive search of the literature for reviews of delirium identification instruments from January 1, 1974, through January 31, 2020. Second, we enumerated the citations of the original publication of each instrument. Third, we evaluated the psychometric characteristics of each instrument and rated the methodological quality of the original publication of the instrument, using the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) framework.[8-10] Fourth, we used an expert panel to identify the domains of delirium critical to identification based on *Diagnostic and Statistical Manual of Mental Disorders* (DSM) criteria. Finally, the expert panel used a combination of the count of citations, the COSMIN methodological rating, and fulfillment of DSM criteria to determine the delirium identification instruments to recommend.

Our approach to conducting and reporting of this systematic review followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines and Institute of Medicine (IOM) Standards for Systematic Reviews (Supplementary Table S1).[11,12] For the systematic review, our goal was to discover as many delirium identification instruments as possible. Because the goal of the study was to identify the most frequently cited instruments, we chose the accepted approach of a review of reviews as the most effective and efficient way to achieve this goal.[13,14] Our search began in 1974, the year the DSM-III first codified delirium,[15] and was inclusive through January 31, 2020.

Data Sources and Searches

We identified articles through searches of six different databases: Cumulative Index to Nursing and Allied Health Literature (CINAHL), Cochrane Library, Excerpta Medica Database (Embase), PsycINFO, PubMed, and Web of Science. The search terms included the keywords "delirium" and "instruments," along with their known synonyms (Supplement A). We limited articles to review articles (systematic review, meta-analysis, or narrative review) with delirium as the main outcome. We required articles to include a minimum of two instruments. For any systematic review of a single instrument, we ensured the instrument was included in another selected article before exclusion. The exclusion criteria included studies exclusively examining alcohol-related delirium (delirium tremens), studies exclusively in pediatric populations, and other article types (i.e., case reports, commentaries, letters, editorials, conference abstracts), or studies where no full-text article was available. Because of the volume of citations to review by primary English-language investigators, we restricted our search to English-language articles only. Prior studies indicated that this approach does not substantially bias systematic reviews.[16] Figure 1 shows the flow diagram for the selection of articles. The articles underwent first-pass screening based on the title and abstract; then second-pass screening was conducted using the full-text article.

Title and Abstract Initial Screening

Before screening, duplicates and non–English-language articles were removed by Endnote X9 software and manual cross-check. The first-pass screening of title and abstract was completed by two independent reviewers (B.H. and M.D.) to exclude articles that did not meet eligibility criteria. Each reviewer independently reviewed the abstracts and used RAYYAN QCRI[17] software to record results, completely blinded to the other's ratings. Articles without an abstract were included in the full-text review. If the article was rated as eligible by either of the two reviewers, the article was included for full-text review. Excluded articles were assigned a single reason for exclusion: studies restricted to pediatric populations, studies using only animal models, studies in which delirium was not the outcome, not a review, or did not evaluate at least two instruments (Figure 1).

Full-Text Review

After the first-pass review, two independent reviewers (B.H. and P.T.) established final eligibility through full-text review. If the article was rated as eligible by either of the two reviewers, the article was included for data extraction. Each rater logged their results in a Google Form in a blinded fashion. Excluded articles were given a single reason for exclusion with the same options previously described. Because the goal of this step was to identify comprehensively all potential delirium identification instruments, we did not conduct an appraisal of the quality of these reviews. We used the systematic reviews, combined with hand searches of references and consultations with experts, to assure comprehensive identification. Once we had found all the instruments, the next step was to appraise the quality of the original studies of those instruments. For eligible articles, information extracted included citation, article type (systematic review, meta-analysis, narrative review), databases and dates searched, search terms, and number of studies and instruments included in the review. Finally, to minimize biased selection based on requiring

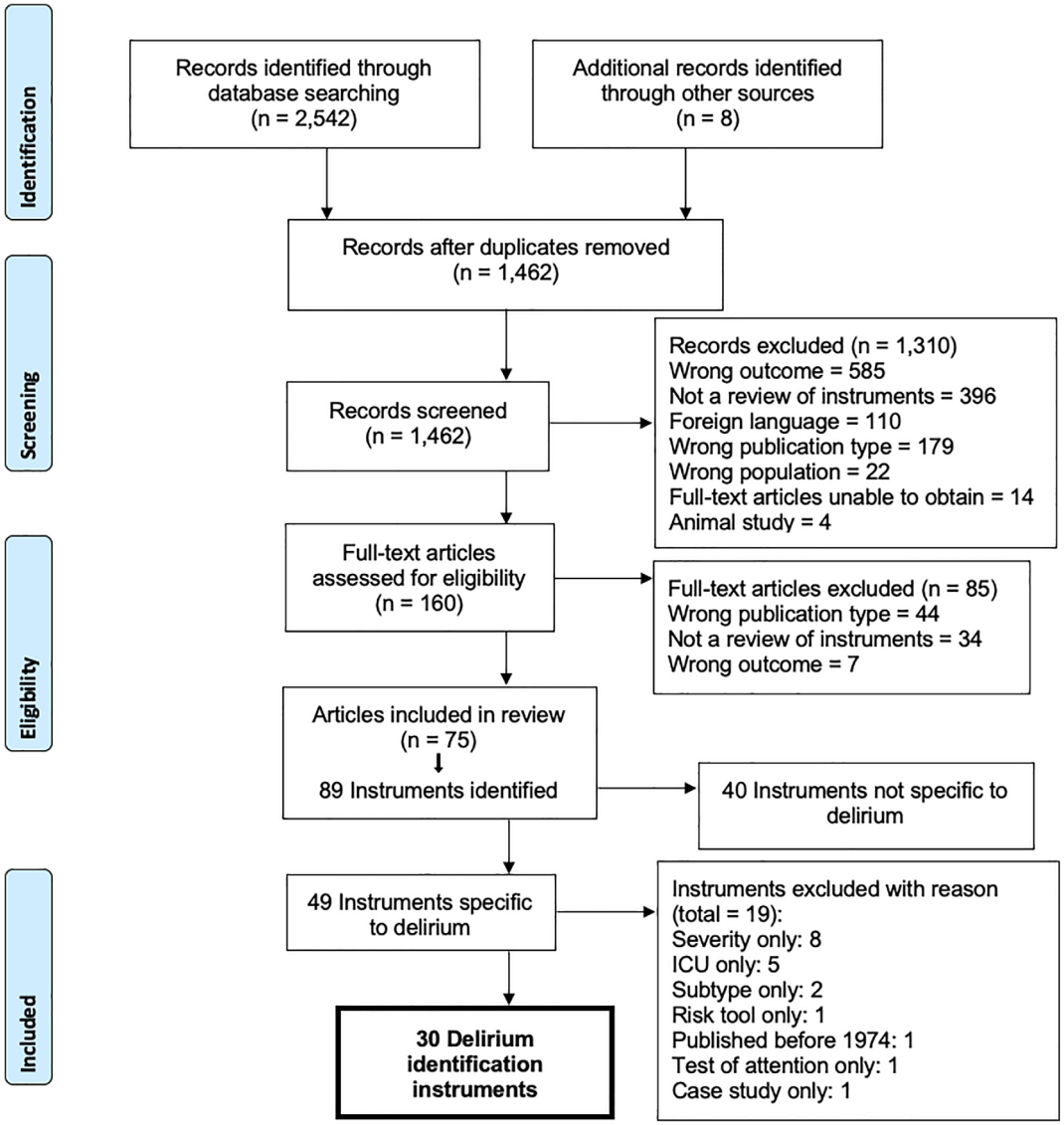

Figure 1. Systematic review flow diagram. Identification of articles occurred through six different online databases. Screening, eligibility, and inclusion were each determined by two independent raters. See text for details.

reporting in an electronic database, and as recommended by the IOM standards for systematic reviews,[12] reviewers searched the reference lists of any included articles to identify other articles to include. We augmented our electronic search with hand reviews and with queries to our experts.

Our goal at this point was to identify all potential instruments used to identify delirium. A full list of the instruments discovered from the eligible articles was presented to our expert panel. We excluded those not specific to delirium (i.e., cognitive screens, sedation instruments, dementia instruments). With the expert panel, we identified several instruments specific to delirium not found in the systematic review to bolster our final list of eligible instruments. At this stage, the experts advised excluding instruments designed solely for use in the ICU because these patients are often nonverbal, resulting in the need for unique assessments that might not be comparable with other instruments or generalizable to other settings. In addition, a systematic review of delirium identification instruments for the ICU was published in 2018.[18] Because this was a study of delirium identification instruments, we chose to additionally exclude instruments measuring only severity and subtypes (hypoactive or hyperactive).

Citation Count

We obtained the original publication for each of the eligible delirium identification instruments. The count of citations of the original publication was determined from Scopus for the date range January 1,1974, to January 31, 2020.

COSMIN-Guided Methodological Rating

Our goal for the second-stage review was to evaluate the psychometric characteristics of the instrument and the methodological quality of the original publication for each selected delirium instrument. We chose the single earliest publication for each instrument. We made an exception for the Delirium Rating Scale (DRS) and used the later study because the instrument had been revised (Delirium Rating

Scale-Revised-98 [DRS-R-98]). We rated the Confusion Assessment Method (CAM) long form and short form separately. A single publication per instrument was used to minimize bias because older instruments might have multiple validation studies. Our quality rating was based on an approach we published previously (Supplement C).[19] Our approach used the COSMIN standards of measurement properties.[8-10] The COSMIN rating was utilized to evaluate the psychometric properties of the instrument as reported in its original study. Each article was reviewed independently in a blinded fashion by at least two of three reviewers (B.H., K.E., and J.Y.) and rated according to the COSMIN framework. The assessment items include ratings of published descriptions of effect indicators, internal consistency, content validity, interrater reliability, construct/convergent validity, and criterion validity (full definitions and scoring are in Supplement C). Estimates and sample sizes for these different types of reliability and validity were recorded. The few small differences between the two independent COSMIN ratings of each article were adjudicated by a third rater (R.N.J.).

The ratings on each of the COSMIN criteria were summed and reported as a 0 to 6 score (Supplement C), using an adaptation of the COSMIN scoring procedure published previously.[19,20] For reporting on each of these categories the instruments were given one point; failure to report on these categories resulted in no points. If a category was reported, but used sample sizes fewer than 50, only a half point was assigned.

Expert Panel Review of Instruments

We assembled an interdisciplinary expert panel to determine the key domains for identification of delirium and ascertained their alignment with DSM criteria. Experts from geriatric medicine (S.K.I., T.T.H., and one anonymous), geriatric psychiatry (E.D.M.), cognitive neurology (T.G.F.), gerontological nursing (P.T.), and social work (E.M.S.) were included in the panel. Face-to-face meetings were held twice in consensus sessions following a modified Delphi approach[21,22] to adjudicate the criteria, with independent, blinded ranking assignments between meetings. We reviewed criteria enumerated in DSM-III, DSM-III-R, DSM-IV, DSM-IV-TR, and DSM-5.[15,23-26] Each individual criterion was first assigned to domain(s) identified previously.[19,27] Then the expert panel rated whether each domain was essential for delirium identification; consensus was considered achieved with agreement by six of seven (86%). The expert panel determined whether each of the 30 delirium identification instruments fulfilled DSM-5 criteria.

Subsequently, the expert panel determined the criteria for selecting the instruments to recommend. After consensus, the following criteria were selected: citation count of 200 or higher, COSMIN score 4 and above, and meeting full DSM-5 criteria. To be recommended, an instrument should meet at least two of these three criteria.

RESULTS

Results of the systematic review are shown in Figure 1. The literature review yielded 2,542 articles that were narrowed based on our exclusion criteria to 160 articles for full-text review. From full-text review, 75 articles (47%) met our inclusion criteria (Supplement B). We identified 89 total instruments. The expert panel determined 49 were specific to delirium; we excluded 19 for the following reasons: measuring severity only (n = 8); intended for ICU patients (n = 5); measuring only delirium subtypes (hypoactive or hyperactive) (n = 2); measuring only risk for developing delirium (n = 1); including only attention tests (n = 1); published before 1974 (n = 1); and case report only (n = 1) (Supplement E). Thus our study included 30 delirium-specific identification instruments developed for use in non-ICU settings (Supplement D). Of these 30 instruments, allowing for multiple categories, usage was 87% for screening, 27% for diagnosis, and 10% for severity. The most common study populations examined included medical and/or surgical wards (47%), geriatric wards (20%), emergency departments (10%), and long-term care facilities (10%). The reference standard used for each study included DSM (40%), CAM (20%), expert clinical judgment only (13%), and not described or not used (27%).

Table 1 shows characteristics of the full-text articles reviewed. There were 18 articles that mentioned at least 10 instruments. No articles were published before 1990; however, since that time, article count has risen exponentially. The 75 included articles individually reviewed between 2 and 19,000 articles.

Table 2 shows the selection criteria for all the delirium identification instruments. Four instruments stand out for satisfying most of the COSMIN framework criteria, assessing many of the DSM-5 criteria, and widespread use as evidenced by their high citation count. These were the CAM (2,685 citations, COSMIN criteria count = 4.5, full DSM-5 criteria), DRS-R-98 (499 citations, COSMIN criteria count = 4.5, full DSM-5 criteria), Memorial

Table 1. Characteristics of Articles Reviewed

Characteristic	N	%
Instruments described	75	100
2	9	12
3	4	5
4	6	8
5	13	18
6	8	11
7	7	9
8	7	9
9	3	4
10–14	8	11
15–19	4	5
≥20	6	8
Year published		
1974–1989	0	0
1990–2000	5	7
2001–2010	17	23
2011–2014	25	33
2015–2019	28	37
Article type		
Meta-analysis	5	7
Systematic review	23	30
Narrative review	47	63

Table 2. Selection Criteria for Delirium Identification Instruments Based on the Original Citation

Name of scale	Count of citations (Scopus: January 1, 1974–January 31, 2020)[a]	COSMIN score (Max = 6)	Meets DSM-5 criteria (Yes/No)
Confusion Assessment Method (CAM), long form and short form	2,909	4.5	Yes
Delirium Rating Scale-Revised-98 (DRS-R-98)	552	4.5	Yes
Memorial Delirium Assessment Scale (MDAS)	532	5	No
Delirium Observation Screening Scale (DOSS)	238	6	No
Chart Delirium Identification (CHART-DEL)	216	3.5	No
Neelon and Champagne Confusion Scale (NEECHAM)	207	5	No
Delirium Symptom Interview (DSI)	204	4	No
Confusion Assessment Method Emergency Department (CAM-ED)	176	2.5	No
4 “A”s test (4AT)	168	4	No
Delirium Triage Screen (DTS)	117	4	No
Brief Confusion Assessment Method (bCAM)	117	4	No
3-Minute Diagnostic Assessment (3D-CAM)	98	4	Yes
Saskatoon Delirium Checklist (SDC)	97	2	No
Single Question in Delirium (SQiD)	64	2	No
Nursing Home-Confusion Assessment Method (NH-CAM)	58	2	Yes
Family-Confusion Assessment Method (FAM-CAM)	48	3.5	Yes
Clinical Assessment of Confusion-A (CAC-A)	40	3.5	No
Recoverable Cognitive Dysfunction Scale (RCDS)	34	2	No
Modified Confusion Assessment Method for the Emergency Department (mCAM-ED)	28	3	No
Delirium Diagnostic Tool-Provisional (DDT-Pro)	27	3.5	No
Confusion Rating Scale (CRS)	26	4	No
Bedside Confusion Scale (BCS)	25	2.5	No
Nursing Delirium Screening Scale (Nu-DESC)	24	2.5	No
Recognizing Acute Delirium as Part of Your Routine (RADAR)	24	4	No
Visual Analog Scale for Acute Confusion (VAS-AC)	22	3	No
Inter Resident Assessment Instrument Acute Care (interRAI AC)	13	4	No
Simple Query for Easy Evaluation of Consciousness (SQeeC)	10	4	No
Informant Assessment of Geriatric Delirium Scale (I-AGeD)	9	4.5	No
Clinical Assessment of Confusion-B (CAC-B)	NA	3.5	No
Organic Brain Syndrome (OBS)	NA	NR	NR

Abbreviations: COSMIN, Consensus-based Standards for the Selection of Health Measurement Instruments; DSM, *Diagnostic and Statistical Manual of Mental Disorders*; NA, not attainable; NR, no rating.

[a]Descending order by count of citations.

Delirium Assessment Scale (MDAS) (492 citations, COSMIN criteria count = 5, partial DSM-5 criteria), and the Delirium Observation Screening Scale (DOSS) (212 citations, COSMIN criteria count = 6, partial DSM-5 criteria).

Figure 2 shows the domain coverage of the CAM, DOSS, DRS-R-98, and MDAS. Domains covered by each instrument were classified as fulfilling DSM-5 criteria, other DSM diagnostic criteria, or other associated features. They are listed in descending order by number of total domains covered, with the DRS-R-98 assessing 13 domains, the CAM long form assessing 11 domains, the MDAS assessing 10 domains, and the DOSS assessing 9 domains. The CAM short form overlaps with the CAM long form and was excluded from this analysis. For the DSM-5 criteria, all instruments included core criteria of inattention, disorientation, and cognitive impairment; however, two instruments (MDAS and DOSS) did not include acute onset and fluctuating course. In other DSM criteria, all four overlapped with the same domains on four of six criteria (disorganized thinking, psychomotor agitation, psychomotor retardation, and hallucinations), and all but the DRS-R-98 included altered level of consciousness. Only the DRS-R-98 included organic etiology.

Table 3 compares the CAM, DOSS, DRS-R-98 and MDAS. These instruments had the highest citation count and COSMIN score. We also show the number of DSM-5 criteria and delirium identification domains met by each of the top four instruments. Table 3 provides additional information about these instruments including time for completion, qualifications of the raters, and evidence of construct and criterion validity. Notably, each of the instruments used a reference standard

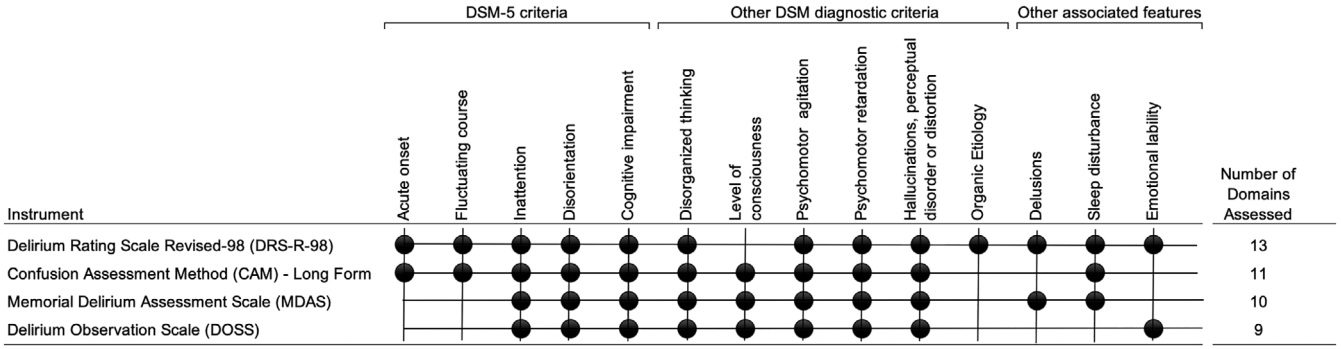

Figure 2. Domain coverage of four recommended delirium instruments. Each dot represents a domain covered by an item on the instrument as determined by the expert panel.

Table 3. Comparison of Four Recommended Delirium Instruments (Alphabetical Order)

Delirium instrument, year of publication, (sample size)	Recommendedtime to complete	Qualifications of raters, original study	Construct validity[a]	Criterion validity[b]	COSMIN rating (best = 6)	Citations, Scopus	DSM-5 criteria fulfilled, n	Domains covered, n
Confusion assessment method (CAM), 1990 (N = 56)	10–15 min (long form), 3–5 min (short form)	Trained lay or clinical raters	r = .64 with MMSE r = .59 withstory recall r = .82 with VAS-C r = .66 digit span	DSM-III-R criteria by psychiatrist	4.5	2,909	5/5	11
Delirium ObservationScale (DOSS), 2003 (N = 92)	<5 min	Nurses without specialized training	r = .60–.79 with MMSE r = .63 with CAM r = .33–.74 with IQCODE	DSM-IV criteria by geriatrician	6	238	3/5	9
Delirium Rating Scale-Revised-98 (DRS-R-98), 2001 (N = 26)	20–30 min (scoring), following about 1 h (gathering information from nurse, family, chart)	Psychiatrically trained clinicians	r = .41 with CTD	DSM-IV criteria by referring service physician	4.5	552	5/5	13
Memorial Delirium AssessmentScale (MDAS), 1997 (N = 30)	10–15 min (scoring), following 15–30 min (interview, information from nurse, family, chart)	Trained clinicians	r = .91 with MMSE r = .89 with CGR r = .88 with DRS	DSM-III-R or DSM-IV criteria by psychiatrist	5	532	3/5	10

Abbreviations: CGR, Clinician's Global Rating; CTD, Cognitive Test for Delirium; DRS, Delirium Rating Scale; DSM, *Diagnostic and Statistical Manual*; COSMIN, Consensus-based Standards for the Selection of Health Measurement Instruments; MMSE, Mini-Mental State Examination; VAS-C, Visual Analog Scale for Confusion.

[a]Construct validity represents a test of correlations with other instruments of the same construct, in this case delirium identification. For r, >.7 indicates a strong relationship, >.5 indicates a moderate relationship, and >.3 indicates a weak relationship.

[b]Criterion validity represents the reference standard assessment used.

delirium diagnosis by a physician based on DSM criteria. Full details of the review of COSMIN criteria and other details for each instrument are described in Supplementary Tables S2 and S3.

DISCUSSION

The ability to identify delirium accurately is important to provide optimal clinical care. Moreover, to advance the

field, it is critical to have reliable approaches for delirium identification. We identified 30 delirium identification instruments used in non-ICU settings. We evaluated several aspects of each instrument including citation count, satisfaction of COSMIN criteria for the evaluation of health measurement instruments, and expert panel guidance regarding the coverage of DSM-5 criteria for delirium. Based on our systematic review combined with an expert panel process, we recommend (in alphabetical order) the CAM, DOSS, DRS-R-98, and MDAS as frequently used and well-validated instruments to identify delirium that are at least partially consistent with the current diagnostic framework (DSM-5) for delirium.

Each of these instruments identifies delirium somewhat differently, assessing different domains. Each was designed for use by different users in varying clinical settings. Thus, the choice in selecting an instrument to identify delirium should be guided by these factors along with logistical considerations for the intended clinical or research application. Although different instruments may be preferred for clinical versus research uses, both settings seek approaches to maximize reliability, validity, and minimize costs and burden of assessment. However, in the clinical setting, users often prioritize expediency that may be counterbalanced by suboptimal diagnostic accuracy.

For the selected instruments, to assist nurses in rapid delirium identification during each shift, the DOSS provides a brief rating (<5 minutes) with minimal training. Although the ratings gather important information assessing clinical progress, an experienced clinician is required to confirm and establish diagnoses. Use of the DRS-R-98 may be preferred by skilled psychiatrically trained clinicians because it provides detailed ratings and has been used in phenomenological delirium studies. However, the administration of the DRS-R-98 is time consuming (20–30 minutes) and comparatively labor intensive. The MDAS is scored with or without additional tests such as the Mini-Mental State Examination.[28] However, all three of these instruments have no built-in diagnostic algorithm and use cutpoints to identify delirium. Thus, a delirium diagnosis can be achieved with multiple different domains.

The CAM can be rated by trained lay interviewers, nurses, or physicians. Scored according to a diagnostic algorithm, the CAM aligns with the DSM-5 diagnostic criteria. There are two forms, a short form that allows rapid assessment (<5 minutes) and a long form (10–15 minutes) to help establish diagnoses in clinical and research applications. The availability of two different forms may offer advantages for large-scale clinical applications or studies. The CAM has been integrated into numerous electronic medical record systems. The CAM short form is widely used as a reliable screening instrument,[29-31] but it does not cover as many domains as the other selected instruments.

Our work extends the findings of two previous reviews. Adamis and colleagues used extensive search strategies to define the features of 24 different delirium instruments including their psychometric properties,[32] rated on a scale from +++ to −. Their review did not utilize a uniform approach to characterize psychometric properties reported across studies. They recommended the CAM, DRS, MDAS, and Neelon and Champagne Confusion Scale (NEECHAM) due to their robustness and ease of use. Our work extends this article by updating the search and instruments included over the past decade and providing a more systematic approach to scoring psychometric and methodologic properties. Subsequently, van Velthuijsen and colleagues used an extensive search strategy to find 28 different delirium instruments.[33] Any study that described psychometric properties of delirium identification instruments was included. The studies were restricted to those that included reference standard delirium diagnoses made by a physician using the DSM, editions III, IV, or 5 or the International Classification of Diseases, 10th Revision. Their quality assessment was guided by Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)[34] that assesses four domains including patient selection, index test, reference standard, and flow and timing. The psychometric properties included in their review included sensitivity and specificity, interrater reliability, and internal consistency reliability. They recommended the CAM and the Nursing Delirium Screening Scale (Nu-DESC), and the DOSS, DRS-R-98, and CAM-Intensive Care Unit (CAM-ICU) were mentioned. Our study extends this previous work by considering citation counts, aligning the instruments with DSM criteria, and addressing other aspects of validity.

The present study has several strengths. We used rigorous approaches including PRISMA and IOM guidelines to guide our comprehensive systematic review. We included a count of citations of the original publication of each instrument, along with methodological quality ratings based on the COSMIN approach. We used an expert panel process to determine the domains for delirium identification and applied them to each instrument item. A major strength includes our review of every DSM delirium criterion since the original codification of delirium in DSM-III. By reviewing each version, we were able to identify an inclusive consensus listing of domains pertinent to delirium identification. This allowed for each version of DSM to be included, many of which served as the reference standards in the original publication. We further aligned each of our recommended instruments with the diagnostic criteria of the current DSM-5. We followed IOM guidelines to ensure instruments were not missed by including hand searches and consulting with experts about other potential instruments to include.[12]

Several limitations deserve comment. First, there is a potential bias because one of the authors (S.K.I.) is a creator of four delirium identification instruments found in our review (CAM, Chart Delirium Identification [CHART-DEL], Family-Confusion Assessment Method [FAM-CAM], and 3-Minute Diagnostic Assessment [3D-CAM]). Additionally, the coauthors (E.D.M and R.N.J.) are creators of the 3D-CAM. We minimized bias by not including any of these coauthors in the direct COSMIN review of any instruments. Second, restricting the COSMIN review to the original publication of each instrument poses another potential limitation. It is possible that had we probed the literature for validation studies for each instrument, we could have amassed more evidence for each instrument. Third, we understand that using citation count could potentially bias toward older instruments; however, this was only one of three criteria that the expert panel selected to rank the quality of the instruments. The other two, COSMIN score and

DSM-5 criteria, would not be biased by the age of the instrument. Fourth, we only considered the presence or absence of a validity or reliability assessment in an original instrument publication as a marker of the rigor of the original presentation. Our ranking may have been more precise if we had incorporated actual values of statistics used in the evaluation. However, not all studies reported all or the same statistics, used samples representative of different populations, and used different reference standards. These differences led us to take a very coarse approach to ranking the rigor of the original publication. Fifth, for reasons described earlier, we did not include instruments developed for ICU patients. We acknowledge that this systematic review is not generalizable to the ICU setting. Finally, the ability to distinguish delirium in persons with underlying dementia is an area of paramount importance for future investigation. Future work will be needed to rate and rank delirium identification instruments for their ability to differentiate delirium and dementia or to identify delirium superimposed on dementia.

This study provides a broad overview of delirium identification instruments. We found numerous instruments used in different clinical settings by different raters. We were unable to recommend a single instrument for universal use. However, we found four instruments that are widely used and were well validated in their original publications with a wide range of clinical and research applications. The study helped refine the construct of delirium through alignment of the delirium assessment items, DSM diagnostic criteria, and other previously identified delirium domains. Although many studies have been published using different delirium identification instruments, comparing these studies is difficult due to the measurement heterogeneity. An important area for future investigation will be to harmonize these measures, which may help to compare results across studies and to combine results from existing studies to form large data sets exploring pathophysiology and treatment. We hope this work will help unify the field around delirium identification and lay a foundation to advance the field.

ACKNOWLEDGMENTS

This work is dedicated to the memory of Joshua B.I. Helfand, Steven F. Hamilton, and Daniel S. Snyder.

Financial Disclosure: This study was supported in part by Grant Nos. R01AG044518 (S.K.I./R.N.J.), R24AG054259 (S.K.I.), P01AG031720 (S.K.I.) from the National Institute on Aging, and Grant No. T32GM107000 (B.K.I.H.) (MD/PhD [M.S.T.P.]) from the Training Program at the University of Massachusetts Medical School.

Conflict of Interest: The authors have declared no conflicts of interest for this article.

Author Contributions: Benjamin K.I. Helfand and Richard N. Jones had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Concept and design*: Helfand, Jones, Inouye, and Boudreaux. *Acquisition, analysis, or interpretation of data*: All authors. *Drafting of the manuscript*: Helfand and Jones. *Critical revision of the manuscript for important intellectual content*: All authors. *Statistical analysis*: Helfand and Jones. *Obtained funding*: Jones and Inouye. *Administrative, technical, or material support*: Helfand, Jones, Schmitt, Inouye, and Boudreaux. *Supervision*: Helfand, Jones, Inouye, and Boudreaux.

Sponsor's Role: The funding sources had no role in the design, conduct, or reporting of this study.

REFERENCES

1. Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in older persons: advances in diagnosis and treatment. JAMA. 2017;318(12):1161-1174.
2. Breitbart W, Gibson C, Tremblay A. The delirium experience: delirium recall and delirium-related distress in hospitalized patients with cancer, their spouses/caregivers, and their nurses. Psychosomatics. 2002;43(3):183-194.
3. MacLullich AM, Hall RJ. Who understands delirium? Age Ageing. 2011;40(4):412-414.
4. Inouye SK, Foreman MD, Mion LC, Katz KH, Cooney LM Jr. Nurses' recognition of delirium and its symptoms: comparison of nurse and researcher ratings. Arch Intern Med. 2001;161(20):2467-2473.
5. Inouye SK, Bogardus ST Jr, Charpentier PA, et al. A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med. 1999;340(9):669-676.
6. de la Cruz M, Fan J, Yennu S, et al. The frequency of missed delirium in patients referred to palliative care in a comprehensive cancer center. Support Care Cancer. 2015;23(8):2427-2433.
7. Neufeld KJ, Nelliot A, Inouye SK, et al. Delirium diagnosis methodology used in research: a survey-based study. Am J Geriatr Psychiatry. 2014;22(12):1513-1521.
8. De Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide. Cambridge, UK: Cambridge University Press; 2011.
9. Mokkink LB, Terwee CB, Patrick DL, et al. COSMIN Checklist Manual. Amsterdam, The Netherlands: University Medical Center; 2012.
10. Mokkink LB, Terwee CB, Stratford PW, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Qual Life Res. 2009;18(3):313-333.
11. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
12. Eden J, Levit L, Berg A, Morton S. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: National Academies Press; 2011.
13. Peters JL, Sutton AJ, Jones DR, Rushton L, Abrams KR. A systematic review of systematic reviews and meta-analyses of animal experiments with guidelines for reporting. J Environ Sci Health Part B. 2006;41(7):1245-1258.
14. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol. 2011;11(1):15.
15. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association; 1980.
16. Morrison A, Polisena J, Husereau D, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care. 2012;28(2):138-144.
17. Rayyan—a web and mobile app for systematic reviews [article 210]. *Syst Rev*. 2016;210. doi:10/1186/s13643-016-0384-4.
18. Gélinas C. Delirium assessment tools for use in critically III adults: a psychometric analysis and systematic review. Crit Care Nurse. 2018;38(1):38-50.
19. Jones RN, Cizginer S, Pavlech L, et al. Assessment of instruments for measurement of delirium severity: a systematic review. JAMA Intern Med. 2019;179(2):231-239.
20. Terwee CB, B ML, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012;21(4):651-657.
21. Kimchi EY, Hshieh TT, Guo R, et al. Consensus approaches to identify incident dementia in cohort studies: systematic review and approach in the successful aging after elective surgery study. J Am Med Direct Assoc. 2017;18(12):1010-1018.e1.
22. Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. Manag Sci. 1963;9(3):458-467.
23. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 3rd ed. Washington, DC: American Psychiatric Association; 1987.
24. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 1994.

25. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC: American Psychiatric Association; 2000.
26. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington, DC: American Psychiatric Society; 2013.
27. Schulman-Green D, Schmitt EM, Fong TG, et al. Use of an expert panel to identify domains and indicators of delirium severity. Quality of Life Research. 2019;28(9):2565-2578.
28. Breitbart W, Rosenfeld B, Roth A, Smith MJ, Cohen K, Passik S. The memorial delirium assessment scale. J Pain Symptom Manage. 1997;13(3):128-137.
29. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment methoda new method for detection of delirium. Ann Intern Med. 1990;113(12):941-948.
30. Wei LA, Fearing MA, Sternberg EJ, Inouye SK. The confusion assessment method: a systematic review of current usage. J Am Geriatr Soc. 2008;56(5):823-830.
31. Shi Q, Warren L, Saposnik G, Macdermid JC. Confusion assessment method: a ystematic review and meta-analysis of diagnostic accuracy. Neuropsychiatric disease and treatment. 2013;9:1359-1370.
32. Adamis D, Sharma N, Whelan PJ, Macdonald AJ. Delirium scales: a review of current evidence. Aging Ment Health. 2010;14(5):543-555.
33. van Velthuijsen EL, Zwakhalen SMG, Warnier RMJ, Mulder WJ, Verhey FRJ, Kempen G. Psychometric properties and feasibility of instruments for the detection of delirium in older hospitalized patients: a systematic review. Int J Geriatr Psychiatry. 2016;31(9):974-989.
34. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529-536.

SUPPORTING INFORMATION

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

Supplementary Appendix S1: Supporting Information