Measurement Instruments for Delirium Severity: A Systematic Review

Richard N. Jones, ScD[1,2], **Sevdenur Cizginer, MD**[3], **Laura Pavlech, DVM, MSLS**[4], **Asha Albuquerque, BA**[5], **Lori A. Daiello, PharmD, ScM**[2], **Kumar Dharmarajan, MD MBA**[6], **Lauren J. Gleason, MD MPH**[7], **Benjamin Helfand, MSc**[1,2,8], **Lauren Massimo, PHD CRNP**[9], **Esther Oh, MD PhD**[10], **Olivia I. Okereke, MD MS**[11,12,13], **Patricia Tabloski, GNP-BC, PhD**[14], **Laura Rabin, MD**[15], **Jirong Yue, MD**[16], **Edward R. Marcantonio, MD, SM**[17,20], **Tamara G. Fong, MD, PhD**[5,13,20], **Tammy Hshieh, MD, MPH**[5,18,20], **Eran Metzger, MD**[19,20], **Kristen Erickson, MPH**[5], **Eva M. Schmitt, PhD**[5,17,20], and **BASIL Study Group**

[1]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University

[2]Department of Neurology, Warren Alpert Medical School of Brown University, Rhode Island Hospital, Providence RI

[3]Department of Medicine, Division of Geriatrics and Palliative Medicine, Warren Alpert Medical School of Brown University, Providence, RI

[4]Hirsh Health Sciences Library, Tufts University, Boston MA

[5]Aging Brain Center, Institute for Aging Research, Hebrew SeniorLife, Boston MA

[6]Center for Outcomes Research & Evaluation, Yale University School of Medicine, New Haven CT

[7]Section of Geriatrics and Palliative Medicine, Department of Medicine, University of Chicago Medicine, Chicago IL

[8]University of Massachusetts Medical School, Worcester MA

[9]Frontotemporal Degeneration Center, University of Pennsylvania School of Medicine, Philadelphia PA

[10]Johns Hopkins University School of Medicine, Baltimore, MD

[11]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, Boston MA

[12]Department of Psychiatry, Massachusetts General Hospital, Boston MA

[13]Department of Neurology, Beth Israel Deaconess Medical Center, Boston MA

[14]Connell School of Nursing, Boston College, Chestnut Hill, MA

[15]Department of Psychology, Brooklyn College and the Graduate Center of CUNY, Brooklyn, NY

[16]Department of Geriatrics, West China Hospital, Sichuan University, Chengdu, China

[17]Department of Medicine, Beth Israel Deaconess Medical Center, Boston MA

[18]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, MA

[19]Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston MA

[20]Harvard Medical School, Boston MA

Abstract

Importance: Measurement of delirium severity has been recognized as highly important for tracking prognosis, monitoring response to treatment, and estimating burden of care both during and after hospitalization. Rather than simply rating delirium as present or absent, the ability to quantify its severity will enable development and monitoring of more effective treatment approaches for delirium.

Objective: This study had 3 major goals: to present a comprehensive review of delirium severity instruments; to conduct a methodologic quality rating of the original validation study of the most commonly used instruments; and to select a group of top-rated instruments.

Evidence Review: Using key words, subject headings, and full text approaches, we conducted a systematic review of the following databases, CINAHL, EMBASE, PsycINFO, PubMed, and Web of Science, from January 1, 1974 through March 31, 2017. Inclusion criteria were original articles assessing delirium severity and utilizing a delirium-specific severity instrument. Final listings of articles were supplemented with hand searches of reference listings to assure completeness. At least 2 reviewers independently completed each step of the review process: article selection, data extraction, and methodologic quality assessment of relevant articles using a validated rating scale. All discrepancies between raters were resolved by consensus.

Findings: From 9,409 articles identified, 228 underwent full text review, and we identified 42 different instruments in studies of delirium severity. Eleven of the 42 tools were multi-domain, delirium-specific instruments providing a quantitative rating of delirium severity, and these underwent a methodologic quality review. Applying pre-specified criteria related to frequency of use, methodologic quality, construct or predictive validity, and broad domain coverage, an expert panel used an iterative modified Delphi process to select 6 final high-quality instruments meeting these criteria.

Conclusions and Relevance: We identified 6 varied instruments with a broad range of clinical applications to be of high quality—the Confusion Assessment Method, Confusional State Examination, Delirium-O-Meter, Delirium Observation Scale, Delirium Rating Scale, and Memorial Delirium Assessment Scale. These measures will enable accurate measurement of delirium severity to improve clinical care for this common and devastating condition. We hope this work will stimulate increased usage, and head-to-head comparison of these instruments.

Keywords

Delirium; delirium severity; measurement instruments; systematic review; predictive validity; methodological review

Introduction

Delirium is a common, serious, and often preventable complication among older adults. An estimated 12 million older Americans experience delirium each year,[1] at a cost of over $164 billion (2011) in annual healthcare expenditures.[2] Delirium is distressing to patients and families,[3] prolongs hospital stays, delays rehabilitation, and increases risks for dementia and death.[1] Despite its importance for patient safety and public health, delirium is often unrecognized by clinicians, and effective treatments remain elusive.[1] Moreover, presentation of delirium is heterogeneous and multifaceted, and measurement of delirium and its severity pose unique challenges.

The time is right to advance measurement of delirium severity: it is important, impactful, and efforts to stratify risk, target treatment, and monitor for outcomes are already possible and feasible with existing approaches. Measurements of delirium severity should play an important role in the advancement of clinical care and research for persons with delirium.[4] Delirium severity ratings are directly associated with clinical outcomes, and thus, provide powerful prognostic tools for clinical care.[4,5] These instruments provide sensitive, continuous measures to track change over time, and thus, can provide finely-grained information on the earliest onset of symptoms or response to treatment. Clinically, delirium severity instruments are useful to track clinical course and recovery, provide meaningful prognostic information, and help assess patient and caregiver needs after discharge. Severity measures can help to gauge the burden of clinical care, providing an effective means to identify safe-staffing levels in the hospital or homecare setting, and also provide data to evaluate the impact of delirium severity on healthcare delivery and costs. The recognition that more severe cases of delirium can lead to long-term cognitive decline[6] has highlighted the importance of rapid recognition of more severe cases and the heightened need for tools to provide reliable serial monitoring over the entire course of delirium. These measures therefore represent powerful outcome measures for clinical trials and prognostic studies. Moreover, such continuous measures can advance statistical approaches and maximize power of delirium studies. Importantly, delirium severity measures are also essential for studies of pathophysiology, since correlating severity and its association with biomarkers or other indicators may shed light on important mechanistic relationships.

Currently, many measures of delirium severity are in active use, and little is known about their comparative characteristics. Comparison of measures is particularly difficult because these instruments were created for different purposes (screening, diagnosis, and severity rating), targeted particular clinical settings and users, and captured different features or behaviors. While many systematic reviews of delirium instruments exist, [7–15] none have focused specifically on delirium severity.

For this study, we defined delirium severity as the cumulative intensity of multi-domain symptoms or behaviors associated with delirium, and define delirium severity instruments as those capturing these symptoms or behaviors on a continuous, quantitative scale. We had three goals for the present study. The first was to present a comprehensive review of delirium severity instruments identified from a systematic literature review from 1974 to 2017. The second was to evaluate the psychometric performance characteristics of the most commonly used delirium severity instruments, applying COSMIN (Consensus-based Standards for the selection of health Measurement INstruments[16,17]) ratings to the instrument's original published report. The third was to select top-rated delirium severity instruments based on frequency of use, COSMIN rating, evidence of construct and/or predictive validity, and breadth of coverage of symptom domains reflecting delirium severity.

Methods

Our approach was informed by systematic literature review guidelines (PRISMA)[18] and specific recommendations for the evaluation of health outcome measures.[19] Our second stage review was conducted on the primary sources of research instruments (i.e., the original published article) identified in the systematic review, followed by a quality review using the COSMIN framework.[16,17] Finally, an expert panel of 7 interdisciplinary delirium experts reviewed all evidence to select the top-rated instruments.

Stage 1 Systematic Literature Review

For the systematic review, our goal was to identify comprehensively the measures used to operationalize delirium severity, and to describe how delirium severity is defined and used in these studies. We started our search in 1974, since work on the third revision to the American Psychiatric Association's Diagnostic and Statistical Manual (DSM-III) was initiated that year, marking a major reconceptualization of the clinical features and definition of delirium. Our searches were updated twice and were inclusive through March 31, 2017.

Data Sources and Searches—We identified articles by pooling results of two comprehensive searches in 5 databases: CINAHL, EMBASE, PsycINFO, PubMed, and Web of Science. Search 1 focused on identification of articles with the keywords "delirium" and "severity", and search 2 focused on identifying articles with a text keywords or subject headings relevant to "delirium" and "tests" or "measures". Adding "intensity" to the search terms did not yield any additional instruments. Exclusion criteria were studies not focused on delirium or delirium severity, studies focused solely on alcohol withdrawal delirium, case reports or editorials, duplicates, studies in children, conference abstracts, other (e.g., non-clinical abstracts, unpublished dissertations, books), studies not published in English or where full text articles were unavailable. The flow diagram for selection of articles appears

in Figure 1. The specific search strategies utilized appear in Appendix. Identified articles underwent an initial screening based on title and abstract. The final eligibility determination was based on a review of the full text, followed by a data extraction phase, detailed below.

Title and abstract initial screening.: This step was completed by four reviewers (MD, PT, RJ, SC) to identify duplicates and exclude manuscripts that did not meet criteria. Each article was first reviewed independently by two reviewers, then results were compared and any discrepancies were resolved by consensus of all reviewers.

Full text review and data extraction.: Following initial screening, full text was reviewed for a final eligibility by a group of 8 reviewers (AA, BH, JY, LD, LG, LM, PT, RJ). Each article was reviewed independently by two reviewers. If either of the two reviewers rated the article as eligible, the article was included for data extraction. One of the reviewers (or one of RJ, SC) subsequently extracted information from eligible articles, including citation, study setting [intensive care, hospital service (medicine, surgery), rehabilitation, long-term care, residential care, community setting, emergency room, or other]; sample size; name and citation for instruments used to measure delirium (up to 3); name and citation for other measures of cognition or behavioral symptoms; description of how delirium severity was defined; and specification of how delirium severity was used in the study (i.e., outcome, main predictor, covariable, descriptor, or other). Our primary goal at this stage was to identify all potential instruments used to assess delirium severity.

To assure comprehensive identification of studies and to avoid potentially biased selection based on requiring reporting in our specified electronic databases, we followed recommended approaches from the Institute of Medicine's (IOM) standards for systematic reviews.[20] Thus, we augmented our electronic searches with hand reviews of reference lists in eligible studies, prior published reviews of delirium instruments, and queries to our expert panel to identify any delirium severity instruments that might have been missed.

Stage 2 COSMIN-Guided Methodologic Quality Assessment

Our second stage review was focused on evaluating the methodological quality of the initial published study for the selected delirium severity instruments. Rating only one validation study put each instrument on a level playing field, and minimized potential bias favoring earlier published instruments which were more likely to have multiple published validation studies. The original validation study of each instrument was selected; however, in 2 cases instruments were later revised (DRS-R98, Delirium Index) and in these 2 cases, the single later validation study was used. To be eligible, the instrument was required to utilize numeric ratings of delirium severity or intensity of delirium symptoms. Many studies defined delirium severity in terms of duration only (e.g., days of delirium), without a numeric rating of severity of symptoms; these measures were excluded from the second stage review since measures of intensity have shown superior performance for prediction of clinical outcomes.[21]

We utilized the COSMIN standards to rate the methodologic quality of measurement properties of the instruments as reported in the original published article for each instrument. Two of 6 reviewers (EO, KD, LR, OO, RJ, SC) reviewed each manuscript independently,

and extracted and rated information according to the COSMIN framework. Briefly, the COSMIN assessment items included ratings of the published descriptions of content validity; internal consistency; and construct, concurrent, and predictive validity of the instrument in the initial article. We also collected information on the intended sample for the instrument; level of education or professional certification suggested or required for the raters; length of the instrument (number of items); and time for administration. A third rater (RJ) adjudicated the few minor discrepancies between the two independent COSMIN ratings of each article.

We summarized the quality of reporting as a 0–6 scale, using an adaptation of the COSMIN scoring procedure, described by Terwee et al.[22] Full scoring details for the 6 reliability and validity criteria are included in the Appendix.

Expert Panel Ratings and Synthesis of Delirium Severity Instruments

We assembled a local interdisciplinary expert panel to review the results of the COSMIN-guided review of the 11 selected instruments, and to select a recommended set of delirium severity instruments. The panel included experts from general internal medicine (ERM), geriatric medicine (SI, TH), geriatric psychiatry (EDM), cognitive neurology (TF), gerontological nursing (PT), and social work (ES). The panel met face-to-face 4 times in consensus sessions to adjudicate the instruments, with independent, blinded ranking assignments between meetings. All procedures followed a modified Delphi approach.[23] The panel agreed *a priori* on the following selection criteria for the instruments: (1) used in at least 2 or more articles in our systematic review to assure usage in at least one additional study beyond its original validation; (2) a rating of 3.5 or greater on COSMIN criteria; (3) strong evidence of construct and/or predictive validity from the original validation study; and (4) broad domain coverage of 9 or more of 16 possible delirium symptom domains. The COSMIN rating of 3.5 or greater (of 6 criteria) was selected by the expert panel to exclude the lowest quality articles. In terms of domain representation, the cut-point of 9 or more was selected, since this is the minimum number needed to yield a scale reliability (McDonald's Omega) of 0.90 for a Rasch measurement model, considered a minimum standard for patient-level outcome measures.[24] To assess domain coverage, 3 panel members were assigned to review independently the domain coverage of each instrument, and any results without complete agreement were adjudicated in two consensus conferences with all panel members. The expert panel had two additional consensus sessions to select the final top-rated instruments. As of the present date, there were no published head-to-head comparisons of any of the 11 instruments.

Results

Systematic review and identification of studies using delirium severity measures.

Results of the systematic review are presented in Figure 1. We initially identified 9,409 articles. After excluding studies not meeting our criteria, 228 articles remained which underwent full-text review with data extraction.

Characteristics of the 228 manuscripts reviewed are presented in Table 1. Although the search spanned 43 years from 1974 to 2017, more than half of the articles were published since 2010. About half (N=116) involved a medical setting, and together with surgical and intensive care settings accounted for about 95% of the manuscripts. While the majority of studies defined delirium severity by quantitative scoring of delirium symptoms (134/228, 59%), others used duration of delirium, other clinical features or cognitive scores, clinical outcomes, or multiple approaches. Citations for the 228 articles reviewed are available at: https://deliriumnetwork.org/measurement/severity-instruments-sr-downloads/.

Selection of delirium-specific severity instruments.

Of the 228 articles reviewed in full text, we identified 42 delirium-specific instruments used for rating delirium and/or delirium severity with most articles using more than one instrument (Table 1). Most manuscripts (65% of 228) used more than one delirium instrument, of which some were not used to quantify severity. The identified instruments are presented in descending order of frequency in Table 2. The 3 most commonly used instruments were the Confusion Assessment Method (CAM)[25] including the CAM-S (109/228 articles, 48%), followed by the Delirium Rating Scale including the DRS-R98[26,27] (101/228 articles, 44%), and the Memorial Delirium Assessment Scale[28] (MDAS) (44/228, 21%). None of the remaining instruments were used in more than 5% of articles reviewed. Only 38 (17%) of all articles did not include *any* of the top 3 instruments. Excluded instruments at this stage (see Table 2 footnote) included case identification instruments, single domain measures, instruments that were not delirium-specific, cognitive tests, or measures used only in a single published study in the systematic review. It is important to note that the 3D-CAM-S[29] and CAM-ICU-7[30] were excluded at this stage because they did not include broad domain coverage across more than 9 domains and because they did not appear in more than one article in our systematic review.

COSMIN methodologic review of delirium severity instruments.

Of the 12 delirium-specific, multi-domain instruments identified in the Stage 1 review, 11 were included in the COSMIN review. The inclusion criteria for the COSMIN review were that the instrument provided a total score or summary rating of delirium features, and was broadly inclusive of the multiple domains of delirium symptoms. Most of the articles used one of the 11 identified instruments (176/228, 77%). The CAM-ICU was excluded at this stage because it did not provide a numerical rating of severity in the studies identified.

The most commonly used instrument across studies (Table 2) was the Confusion Assessment Method[25] (CAM). While the original instrument was not proposed as a multiple-domain quantitative summary of delirium severity, the more recent CAM-S severity score[5] met our criteria and is therefore included in our second stage COSMIN review. The Delirium Index[31], another severity score derived from the CAM, is also included. Thus, COSMIN ratings were completed on 13 articles for 11 instruments. Two validation studies each were reviewed for the Delirium Index and revision[31,32] and for the Delirium Rating Scale (DRS) and its revision (DRS-R-98[26]). We included only the most recent validation study in each case for our final COSMIN adjudication. A summary of the results of the Stage 2 COSMIN reviews is provided in the Appendix.

The most common methodologic problem in the validation studies was inadequate sample size: 8 of the 11 manuscripts used small samples (n<50) in at least one aspect of assessing reliability or validity. The most commonly missed COSMIN criteria were assessments of criterion or external validity (3 failed to report, 3 rated as fair, 5 rated as good) and assessments of internal consistency reliability (1 failed to report, 6 rated as fair, 4 rated as good). Only one study failed to report on inter-rater reliability.

Expert panel ratings of delirium severity instruments.

Domain coverage of all instruments, as adjudicated by the expert panel, is shown in Figure 2. The expert panel selected 6 final instruments (Table 3) that met all selection criteria based on frequency of use (i.e., 2 or more publications); methodologic quality (i.e., COSMIN score 3.5 or higher); strong evidence of construct or predictive validity; and broad domain coverage (i.e., 9 or more domains). Table 3 includes logistic considerations (i.e., time for completion, qualifications of raters), whether the instrument yields a delirium diagnosis by criteria (not by cut-point alone, which provides for a dual-purpose instrument), as well as a numerical severity rating, and details of the methodologic review (i.e., COSMIN rating, construct and predictive validity, and domain coverage). The relative cost estimate provides a qualitative comparative estimate of the cost for application of each tool, as determined by a combination of the instrument administration time and required level of the training and clinical experience of the rater. Two instruments, the Delirium-O-Meter (DOM) and Delirium Observation Scale (DOS) had the shortest administration times (<5 minutes), while the DRS-R-98 had the longest time (20–30 minutes). Only two instruments, the CAM-S and DRS-R-98, provided delirium diagnosis by criteria. The Confusional State Examination (CSE) and DOM covered the broadest number of symptom domains (n=12 of 16). Of the 6 instruments, the CAM-S was the only one originally designed to be rated by lay interviewers (as well as clinicians), and demonstrated to have predictive validity for a range of clinical outcomes in the original study.

Conclusions

Given the importance of delirium severity, identifying accurate and reliable approaches to measurement is crucial to advance the field and ultimately improve patient care. In our systematic review of 228 articles, we identified 6 delirium severity instruments that met pre-specified criteria for frequency of use, methodologic quality, construct validity, and broad domain coverage. Each of these instruments represents an important contribution to the toolkit of delirium severity instruments. Overall, the CAM[25] (including the CAM-S[5]), the DRS[33] (including DRS-R98[26]), and the MDAS[28] were the most commonly used instruments for delirium severity identified in our study; the 3 additional instruments were the Confusional State Examination, Delirium-O-Meter, and Delirium Observation Scale.

The selection of a specific delirium severity instrument for clinical or research purposes should be guided by the goals of use and logistical constraints. Each of the 6 instruments has unique strengths and limitations, and several potential scenarios for their usage are provided here. For instance, for ratings of delirium severity by floor nurses on each shift, the DOM and DOS provide brief (<5 minute) ratings requiring minimal training; however, these

ratings, while providing valuable information regarding trajectory and velocity of a patient's progress, would require confirmation by an experienced clinicians before a delirium diagnosis can be established. Given its detailed ratings by skilled psychiatrically trained clinicians, the DRS has been widely used for phenomenological studies of delirium; however, the ratings can be time-consuming (20–30 minutes) and may not be feasible for widespread clinical use. For studies requiring both a delirium diagnosis and severity rating, the CAM-S might be preferred. The CAM-S can also be rated by trained lay interviewers or nurses, which may pose advantages for large-scale clinical applications or studies. For studies utilizing the Mini-Mental State Examination (MMSE), the MDAS provides severity ratings based on MMSE items. Finally, if broad domain coverage is a priority, particularly with inclusion of symptoms of behavioral or emotional dysregulation (e.g., lability, anxiety, depression), the CSE or DOM might be considered.

While delirium severity measures have primarily been utilized in research to date, high quality severity measures can have immediate, highly relevant applications in clinical care and quality improvement efforts. For example, given the preponderance of evidence, patients identified with severe delirium should be prioritized for non-pharmacologic interventions to mitigate their symptoms, and flagged for special follow-up monitoring given their heightened risk for long-term cognitive decline. These patients are also likely to be high utilizers, who would benefit from case management or specialized pathways in current healthcare systems.

The strengths of the current study include the rigorous approaches applied to the comprehensive systematic review augmented by hand searches following IOM guidelines, the methodologic rating based on the COSMIN approach, and rigorous expert panel processes for selection of the final top-rated instruments by pre-specified criteria. The final result includes 6 high quality, multi-dimensional and flexible instruments to serve highly varied uses. This study is comprehensive and inclusive, and serves to demonstrate the wide spectrum of instruments in current usage.

Several limitations deserve comment. First, it is important to acknowledge that one of the authors (SI) was the creator of the CAM-S, and several authors participated in its validation (ES, EM, RJ). Steps were taken to minimize potential bias throughout the process; as one example, SI was not involved in the initial selection and COSMIN ratings; and only involved in the final expert panel process. To further minimize bias, all interdisciplinary experts had equal votes in the final rankings and consensus was required on all decisions. Second, different search strategies or screening procedures may have identified different delirium severity instruments. However, we minimized this possibility by using the IOM recommendations of hand reviews of bibliographies from articles and consultations with experts. Another potential limitation is that the COSMIN review was based on original report only, and using all published validation studies may have yielded differing results. However, allowing multiple validation studies favors earlier published instruments; thus, we chose to include only one article per instrument to place each instrument on a level playing field. COSMIN rates only the quality of reporting, not the face or construct validity of the instruments. Thus, innovative and useful approaches to quantifying delirium severity might have been presented in publications that did not meet the rigorous reporting guidelines. An

additional caveat is our choice to focus the COSMIN review on instruments assessing 9 or more domains. We acknowledge that more lenient inclusion thresholds may have led to choosing instruments with fewer domains; however, our current threshold allowed us to achieve our goal of broad multi-domain representation. All of the instruments identified required some verbal response from patients. While some can still be rated in nonverbal patients, the final list of instruments did not include any specific to the non-verbal patient, those with disturbed arousal, or in the intensive care unit setting. Standardizing scoring across instruments can be challenging, and may require detailed scoring and training instructions. Future work will be needed to validate these instruments in persons with dementia. Finally, recently published instruments were at a disadvantage for inclusion, since they might not have had the opportunity to be used in 2 or more studies. Thus, this study will require updating as the field continues to evolve.

This study allowed us to more fully conceptualize delirium severity and to identify characteristics of an ideal instrument. These characteristic include quick to administer, easy to use by raters with minimal training, yields diagnosis by criteria as well as providing a severity rating, high construct and predictive validity, and broad domain coverage across delirium symptoms. While this study did not allow us to identify a *single* best instrument and provide a recommendation for universal use, we present 6 varied instruments with a broad range of clinical applications. Based on the strengths of each instrument, we have provided suggestions for use in specific clinical and research settings. These targeted uses will strengthen and enable more consistent and accurate measurement of delirium severity to improve clinical outcomes resulting from this devastating condition and advance the science of delirium research. We hope this study will stimulate more head-to-head comparison studies of these instruments, as well as pragmatic guidance to translate them into clinical practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Abbreviations:

BIDMC	Beth Israel Deaconess Medical Center

BWH	Brigham and Women's Hospital
HMS	Harvard Medical School
HSL	Hebrew SeniorLife

References

1. MacLullich AMJ, Hall RJ. Who understands delirium? Age Ageing 2011;40(4):412–414. [PubMed: 21636556]

2. Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in Older Persons: Advances in Diagnosis and Treatment. JAMA 2017;318(12):1161–1174. [PubMed: 28973626]

3. 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. [PubMed: 12075033]

4. Eubank KJ, Covinsky KE. Delirium severity in the hospitalized patient: time to pay attention. Ann Intern Med 2014;160(8):574–575. [PubMed: 24733202]

5. Inouye SK, Kosar CM, Tommet D, et al. The CAM-S: development and validation of a new scoring system for delirium severity in 2 cohorts. Ann Intern Med 2014;160(8):526–533. [PubMed: 24733193]

6. Vasunilashorn SM, Fong TG, Albuquerque A, et al. Delirium Severity Post-Surgery and its Relationship with Long-Term Cognitive Decline in a Cohort of Patients without Dementia. J Alzheimers Dis 2018;61(1):347–358. [PubMed: 29171992]

7. De J, Wand AP. Delirium Screening: A Systematic Review of Delirium Screening Tools in Hospitalized Patients. Gerontologist 2015;55(6):1079–1099. [PubMed: 26543179]

8. Mariz J, Costa Castanho T, Teixeira J, Sousa N, Correia Santos N. Delirium Diagnostic and Screening Instruments in the Emergency Department: An Up-to-Date Systematic Review. Geriatrics 2016;1(3):22.

9. Morandi A, McCurley J, Vasilevskis EE, et al. Tools to detect delirium superimposed on dementia: a systematic review. J Am Geriatr Soc 2012;60(11):2005–2013. [PubMed: 23039270]

10. Carin-Levy G, Mead G, Nicol K, Rush R, van Wijck F. Delirium in acute stroke: screening tools, incidence rates and predictors: a systematic review. J Neurol 2012;259(8):1590–1599. [PubMed: 22234842]

11. Carvalho JPLM, de Almeida ARP, Gusmao-Flores D. Delirium rating scales in critically ill patients: a systematic literature review. Revista Brasileira de Terapia Intensiva 2013;25(2):148–154. [PubMed: 23917980]

12. Schenning KJ, Deiner SG. Postoperative delirium: a review of risk factors and tools of prediction. Current Anesthesiology Reports 2015;5(1):48–56.

13. Tamune H, Yasugi D. How can we identify patients with delirium in the emergency department?: A review of available screening and diagnostic tools. Am J Emerg Med 2017;35(9):1332–1334. [PubMed: 28571901]

14. Martins S, Simoes M, Fernandes L. Elderly delirium assessment tools review. Current Psychiatry Reviews 2012;8(2):168–174.

15. Adamis D, Sharma N, Whelan PJP, Macdonald AJD. Delirium scales: A review of current evidence. Aging & Mental Health 2010;14(5):543–555. [PubMed: 20480420]

16. 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. [PubMed: 19238586]

17. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol 2010;10(1):22. [PubMed: 20298572]

18. 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. [PubMed: 19621072]

19. Klem M, Saghafi E, Abromitis R, Stover A, Dew MA, Pilkonis P. Building PROMIS item banks: librarians as co-investigators. Qual Life Res 2009;18(7):881–888. [PubMed: 19548118]

20. Morton S, Berg A, Levit L, Eden J. Finding what works in health care: standards for systematic reviews National Academies Press; 2011.

21. Vasunilashorn SM, Marcantonio ER, Gou Y, et al. Quantifying the Severity of a Delirium Episode Throughout Hospitalization: the Combined Importance of Intensity and Duration. J Gen Intern Med 2016;31(10):1164–1171. [PubMed: 27259291]

22. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. 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. [PubMed: 21732199]

23. 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 Dir Assoc 2017;18(12):1010–1018 [PubMed: 28927945]

24. Nunnally J, Bernstein I. Psychometric therapy New York, NY: McGraw-Hill 1994.

25. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: The Confusion Assessment Method. A new method for detection of delirium. Ann Intern Med 1990;113(12):941–948. [PubMed: 2240918]

26. Trzepacz PT, Mittal D, Torres R, Kanary K, Norton J, Jimerson N. Validation of the Delirium Rating Scale-Revised-98: comparison with the Delirium Rating Scale and the Cognitive Test for Delirium. The Journal of neuropsychiatry and clinical neurosciences 2001;13(2):229–242. [PubMed: 11449030]

27. Trzepacz PT, Dew MA. Further analyses of the Delirium Rating Scale. Gen Hosp Psychiatry 1995;17(2):75–79. [PubMed: 7789787]

28. Breitbart W, Rosenfeld B, Roth A, Smith MJ, Cohen K, Passik S. The Memorial Delirium Assessment Scale. Journal of Pain Symptom Management 1997;13(3):128–137. [PubMed: 9114631]

29. Vasunilashorn SM, Guess J, Ngo L, et al. Derivation and Validation of a Severity Scoring Method for the 3-Minute Diagnostic Interview for Confusion Assessment Method--Defined Delirium. J Am Geriatr Soc 2016;64(8):1684–1689. [PubMed: 27374833]

30. Khan BA, Perkins AJ, Gao S, et al. The Confusion Assessment Method for the ICU-7 Delirium Severity Scale: A Novel Delirium Severity Instrument for Use in the ICU. Crit Care Med 2017;45(5):851–857. [PubMed: 28263192]

31. McCusker J, Cole MG, Dendukuri N, Belzile E. The delirium index, a measure of the severity of delirium: new findings on reliability, validity, and responsiveness. J Am Geriatr Soc 2004;52(10):1744–1749. [PubMed: 15450055]

32. McCusker J, Cole M, Bellavance F, Primeau F. Reliability and validity of a new measure of severity of delirium. Int Psychogeriatr 1998;10(4):421–433. [PubMed: 9924835]

33. Trzepacz PT, Baker RW, Greenhouse J. A symptom rating scale for delirium. Psychiatry Res 1988;23(1):89–97. [PubMed: 3363018]

Key Points

Question:

To identify high-quality delirium severity instruments for clinical care and research.

Findings:

Using rigorous systematic review methodology, we identified 42 instruments in 228 studies of delirium severity. Eleven of the 42 were multi-domain, delirium-specific instruments providing a quantitative rating of severity. Applying pre-specified criteria related to frequency of use, methodologic quality, construct validity, and broad domain coverage, an expert panel selected 6 final high-quality instruments.

Meaning:

The Confusion Assessment Method (CAM-S), Confusional State Examination, Delirium-O-Meter, Delirium Observation Scale, Delirium Rating Scale, and Memorial Delirium Assessment Scale were selected as recommended measures for accurate measurement of delirium severity.
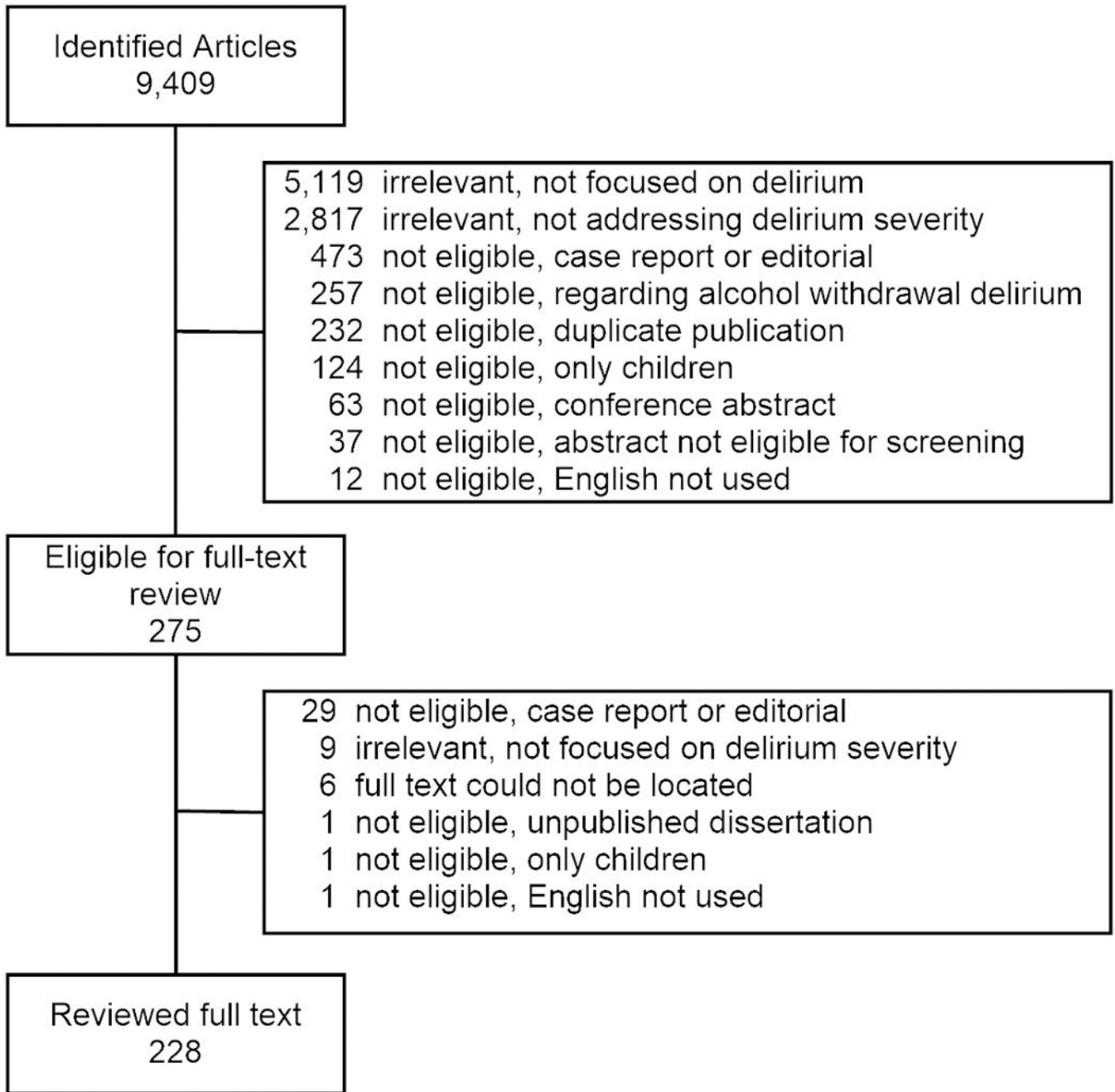
Figure 1. Selection and exclusion of articles for full text review.
Following a comprehensive, systematic review, titles and abstracts were reviewed to determine any exclusion criteria by 2 independent raters, see text for full details. Reasons for exclusion are depicted; 228 articles underwent full text review.

Instrument	Inattention & Consciousness		Cognitive			Emotional Dysregulation					Perceptual dysfunction and delusions		Psycho-motor features		Functional		Number of Domains Assessed
	Inattention	Level of consciousness	Disorganized thinking	Disorientation	Cognitive impairment	Emotional lability	Anxiety, including fear and sense of unease	Depression, including apathy and withdrawal	Anger, including irritability and hostility	Inappropriate behavior [in the hospital]	Hallucinations, perceptual disorder or distortion	Delusions	Psychomotor agitation	Psychomotor retardation	Sleep disturbance	Decline or low performance	
Confusion State Examination (CSE)	●	●	●	●		●	●		●		●	●	●	●	●		12
Delirium-O-Meter (DOM)	●	●	●	●			●	●		●	●	●	●	●	●		12
Delirium Rating Scale (DRS-R-98)	●	●	●	●	●	●				●	●	●	●	●			11
Delirium Observation Scale (DOS)	●	●	●	●	●	●					●		●	●	●		10
Memorial Delirium Assessment Scale (MDAS)	●	●	●	●	●						●	●	●	●	●		10
Confusion Assessment Method (CAM-S)	●	●	●	●	●						●		●	●	●		9
Delirium Index (DI, DIDX)	●	●	●	●	●						●		●	●			8
Agitation Distress Scale						●				●	●	●	●	●	●		7
NEECHAM	●	●	●	●									●	●		●	7
Recoverable Cognitive Dysfunction Scale (RCDS)	●	●	●		●									●			5
Communication Capacity Scale		●	●														2
Number of instruments assessing domain	9	9	10	8	6	4	2	1	2	3	8	5	9	10	6	1	

Figure 2. Domain coverage of 11 multi-item delirium severity instruments.
The domains represented are a descriptive compilation of the items from the identified instruments. Specific domain definitions and coverage decisions were adjudicated by an expert panel, see text for full details. All domain coverage decisions were based on review of item and response option content by two or more independent experts, with discrepancies adjudicated at a consensus conference of all experts. Representation of a domain in the instrument is noted with a black dot; either partial or full coverage of a domain met criteria for including the domain by the expert panel. Specific domain definitions are available upon request (RNJ).

Table 1.

Characteristics of manuscripts reviewed

Characteristic	N	%
Number of manuscripts (n, %)	228	100
Number of delirium instruments used (n, %)		
1	79	35
2	100	44
3	39	17
4	10	4
Year published (n, %)		
1974-1977	0	0
1978-1990	3	1
1991-2000	20	9
2001-2010	61	27
2011-2013	57	25
2014-2016	77	34
2017	10	4
Setting (n, %)		
Medical	116	51
Surgical	63	28
Intensive care unit	35	15
Long-term care	13	6
Palliative care	11	5
Rehabilitation	13	6
Emergency room	7	3
Community	2	1
How delirium severity used (n, %)		
Outcome	51	22
Main predictor	47	21
Descriptor	38	17
Covariate	15	7
Other or not specified	86	38
How severity defined (n, %)		
Count of delirium symptoms or features	134	59
Duration of delirium	66	29
Other clinical features or cognitive scores	12	5
Clinical outcome(s)	5	2
Other (e.g., multiple)	2	1
Unclear or not specified	12	5

Table 2.

Most common instruments[*] used in 228 studies of delirium severity, 1978–2017[*]

Delirium-Specific Instrument	N (%)			Used for Diagnosis	Used for Severity
Confusion Assessment Method (CAM) or CAM-S	109	48		Y	Y
Delirium Rating Scale (DRS) or DRS-R-98	101	44		Y	Y
Memorial Delirium Assessment Scale	47	21		Y	Y
Delirium Index	20	9		---	Y
Confusion Assessment Method for the Intensive Care Unit (CAM-ICU)	12	5		Y	No, days only
Reversible Cognitive Dysfunction Scale	4	2		Y	Y
Delirium Observation Scale	4	2		Y	Y
Neelon and Champagne Confusion Scale	3	1		Y	Y
Delirium-O-Meter	3	1		Y	Y
Confusional State Evaluation	2	1		---	Y
Communication Capacity Scale	2	1		---	Y
Agitation Distress Scale	2	1		---	Y

[*] Instruments included on this table are delirium-specific measures that cover more than one domain, and appear in more than 1 published study. The following measures were excluded from further review: measures used for diagnosis or screening only (e.g., Delirium Symptom Interview, Delirium Diagnostic Tool-Provisional, Delirium Detection Score, NUDESC, tachistoscope); measures which are single-item ratings of global severity (e.g., Global Clinical Impression - Severity scale, Breitbart's Clinician's Global Rating for Delirium); measures that assessed a single domain or feature of delirium (e.g., Richmond Agitation Sedation Scale (RASS), Riker Sedation-Agitation Scale (SAS) , the Glasgow Coma Scale (GCS), and Fainsinger consciousness score, Observational Scale of Level of Arousal (OSLA), DelAPP, Bush Francis Catatonia Rating Scale (BFCRS), Delirium Motor Subtype Scale (DMSS), and Delirium Motoric Checklist (DMC); measures ratings distress not severity (e.g., Delirium Experience Questionnaire); diagnostic criteria (e.g., DSM, ICD); other measures that were not delirium-specific (e.g., Ease of Ward Management Scale; NCI Common Terminology for Adverse Events); measures that were only used in a single study in the systematic review (e.g., 3DCAM-CAM-S, CAM-ICU-7, Confusion Rating Scale, and Delirium Assessment Scale).

Table 3.

Comparison of 6 top-rated delirium severity instruments (alphabetical order)[*]

Delirium Severity Instrument, year of publication, (Sample size)	Recommended Time to Complete	Qualifications of Raters (original study)	Provides Diagnosis by Criteria[†]	COSMIN rating, (best=6)	Construct Validity[‡]	Prediction of clinical outcomes	No. domains covered	Relative Cost Estimate[§]
Confusion Assessment Method (CAM-S), 2014 (N = 919 + 300)	10–15 minutes (long form)	Trained lay or clinical raters	Y	5	r=.64 with MMSE r=.64-.80 with global confusion rating	Y	9	$
Confusional State Examination (CSE), 1997 (N =51)	30 minutes	Trained nurse, psychologist or physician	N	5.5	r=.87 with MMSE r=.79 with CGR	N	12	$$
Delirium-O-Meter (DOM), 2005 (N = 92)	<5 minutes	Nurses without specialized training	N	4.5	r=.83 with MMSE r=.87 with DRS	N	12	$$
Delirium Observation Scale (DOS), 2003 (N = 92)	< 5 minutes	Nurses without specialized training	N	6	r=.60-.79 with MMSE r=.63 with CAM r=.33-.74 with IQCODE	N	10	$$
Delirium Rating Scale (DRS-R-98), 2001 (N=26)	20–30 minutes (scoring), following ~1 hour (gathering information from nurse, family, chart)	Psychiatrically trained clinicians	Y	3.5	r=.43 with MMSE	N	11	$$$
Memorial Delirium Assessment Scale (MDAS), 1997 (N = 30)	10–15 minutes (scoring), following15–30 minutes (interview, information from nurse, family, chart)	Trained clinicians	N	5	r=.91 with MMSE r=.88 with DRS	N	10	$$

*
All instruments included in this table are available for free (no licensing cost) and can be accessed online. Specific copyright information on each is available at: https://deliriumnetwork.org/measurement/delirium-info-cards/ Selection criteria for final instruments: (1) used in at least 2 or more articles; (2) COSMIN rating of 3.5 or greater; (3) strong evidence of construct and/or predictive validity; and (4) broad domain coverage of 9 or more domains. Only one validation study per instrument was reviewed for this study, either original validation or first available study after adaptation, see text for details. All of these instruments scored severity using counts of either number of domains, intensity of each domain, or both on a continuous scale.

†
Diagnosis by specific pre-specified criteria, not by cut-point alone

‡
r = correlation coefficient, with >0.7 indicating a strong relationship; >0.5 indicating a moderate relationship; and >0.3 indicating a weak relationship.

§
Relative cost determined from administration time and required level of clinical training of interviewer (e.g., clinician assigned higher relative cost than lay rater)