



Published in final edited form as:

*Dement Geriatr Cogn Disord.* 2020 ; 49(1): 77–90. doi:10.1159/000506700.

## New Delirium Severity Indicators: Generation and Internal Validation in the Better Assessment of Illness (BASIL) study

Sarinnapha M. Vasunilashorn<sup>1,2,3</sup>, Dena Schulman-Green<sup>4</sup>, Douglas Tommet<sup>5</sup>, Tamara G. Fong<sup>1,6,7</sup>, Tammy T. Hshieh<sup>1,8</sup>, Edward R. Marcantonio<sup>1,2,6</sup>, Eran D. Metzger<sup>1,9</sup>, Eva M. Schmitt<sup>6</sup>, Patricia A. Tabloski<sup>10</sup>, Thomas G. Trivison<sup>1,6</sup>, Yun Gou<sup>6</sup>, Benjamin Helfand<sup>11</sup>, Sharon K. Inouye<sup>1,2,6,\*</sup>, Richard N. Jones<sup>5,12,\*</sup>, for the BASIL Study Team\*\*

<sup>1</sup>Harvard Medical School, Boston, MA, USA

<sup>2</sup>Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

<sup>3</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, USA

<sup>4</sup>Yale University School of Nursing, New Haven, CT, USA

<sup>5</sup>Department of Psychiatry and Human Behavior, Brown University Warren Alpert Medical School, Providence, RI;

<sup>6</sup>Aging Brain Center, Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA;

<sup>7</sup>Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA;

<sup>8</sup>Department of Medicine, Brigham and Women's Hospital, Boston, MA;

<sup>9</sup>Department of Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA

<sup>10</sup>William F Connell School of Nursing at Boston College, Boston, MA

<sup>11</sup>University of Massachusetts Medical School, Worcester, MA

<sup>12</sup>Department of Neurology, Brown University Warren Alpert Medical School, Providence, RI

**Corresponding author:** Sarinnapha M. Vasunilashorn, PhD, Department of Medicine, Harvard Medical School, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston MA 02215, Office: 617-754-1417, Fax: 617-754-1440, svaunil@bidmc.harvard.edu.

Authors' contributions

Conception and design: SMV, TGF, TTH, ERM, EMS, TGT, SKI, RNJ

Analysis and interpretation of the data: SMV, DSG, DT, TGF, TTH, ERM, EDM, EMS, PAT, TGT, YG, BH, SKI, RNJ

Critical revision of the article for important intellectual content: SMV, DSG, DT, TGF, TTH, ERM, EDM, EMS, PAT, TGT, YG, BH, SKI, RNJ

Final approval of the article: SMV, DSG, DT, TGF, TTH, ERM, EDM, EMS, PAT, TGT, YG, BH, SKI, RNJ

Provision of study materials or patients: TGF, TTH, ERM, EMS, SKI, RNJ

Statistical expertise: DT, TGT, YG, RNJ

Obtaining of funding: SKI, RNJ

Administrative, technical, or logistic support: TGF, TTH, EMS

Collection and assembly of data: DT, EMS, YG, RNJ

\*Contributed equally to this work

\*\* see Appendix A

Competing interests

The authors declare that they have no competing interests in this section.

Statements

Ethics approval and consent to participate

Signed informed consent was obtained from participants whenever possible and was otherwise obtained from a health care proxy.

Study procedures were approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center and Hebrew SeniorLife, the study coordinating center.

## Abstract

**Introduction:** Delirium is a common and preventable geriatric syndrome. Moving beyond the binary classification of delirium present/absent, delirium severity represents a potentially important outcome for evaluating preventive and treatment interventions, and tracking the course of patients. Although several delirium severity assessment tools currently exist, most have been developed in the absence of advanced measurement methodology and have not been evaluated with rigorous validation studies.

**Objective:** We aim to report our development of new delirium severity items and the results of item reduction and selection activities guided by psychometric analysis of data derived from a field study.

**Methods:** Building on our literature review of delirium instruments and expert panel process to identify domains of delirium severity, we adapted items from existing delirium severity instruments and generated new items. We then fielded these items among a sample of 352 older hospitalized patients.

**Results:** We used an expert panel process and psychometric data analysis techniques to narrow a set of 303 potential items to 17 items for use in a new delirium severity instrument. The 17-item set demonstrated good internal validity and favorable psychometric characteristics relative to comparator instruments, including the Confusion Assessment Method-Severity Score, the Delirium Rating Scale Revised 98, and the Memorial Delirium Assessment Scale.

**Conclusion:** We have more fully conceptualized delirium severity and have identified characteristics of an ideal delirium severity instrument. *These characteristics include an instrument that is relatively quick to administer; easy to use by raters with minimal training; and provides a severity rating with good content validity, high internal consistency reliability, and broad domain coverage across delirium symptoms. We anticipate these characteristics to be represented in the subsequent development of our final delirium severity instrument.*

## Keywords

delirium; delirium severity; psychometrics

---

## Background

Delirium is a clinical syndrome characterized by an acute decline in cognition, often associated with inattention, disorientation, lethargy or agitation, and perceptual disturbance [1]. Delirium is a common, serious, and morbid condition for older adults. With in-hospital mortality rates of 25–33% and annual healthcare costs in excess of >\$182 billion (2011 USD) [2] delirium has garnered increasing attention as a worldwide public health and patient safety priority [3,69,70].

Due to its potentially preventable nature, delirium is a high priority target for preventive strategies and clinical trials [4]. Thus, identification of appropriate delirium outcomes is critical to track the clinical course, recovery, and response to intervention and treatment. *Moreover, the ability to correlate delirium status with biomarkers would advance our mechanistic understanding and would enable the development of pathophysiologically-based*

*treatment approaches [71]. Additionally, severity scores may also be useful in predicting prognosis.* Delirium severity represents a nuanced and potentially more powerful framework than a binary classification of delirium as present or absent [5], and is a promising outcome for informing intervention strategies and clinical trials.

While a variety of tools for the assessment of delirium severity currently exist, most delirium severity instruments have not been developed with advanced measurement methodology or evaluated with rigorous validation studies [6]. An expert panel of delirium researchers tasked with identifying domains and indicators of delirium severity revealed several characteristics of an ideal instrument. This includes: broad domain coverage across delirium symptoms, high construct and predictive validity, yielding diagnosis by a severity rating and criteria, quick administration, and ease of use by minimally trained raters [7]. This review indicated that high quality delirium severity instruments should ultimately have immediate relevant application to clinical care and quality improvement efforts.

Through the Better Assessment of Illness (BASIL) study, we have undertaken rigorous synthesis and development activities focusing on measurement of delirium severity that builds on the characteristics identified from the expert panel [7], with the ultimate goal of developing new instruments to measure delirium severity as a multi-dimensional construct. Specifically, the goal of this paper is to report our selection of domains and selection/development of new items guided by psychometric analysis of data derived from a field study (the BASIL study).

## Methods

Our approach to the development of a new delirium severity measure was informed by consensus standards for the evaluation of health measurement tools [8, 9] and followed examples of measurement development in general health and neurological outcomes provided by PROMIS [10] and the NIH Toolbox [11]. Following the studies summarized above, we defined the construct to be measured, and identified seven broad domains (and 13 narrow domains) that are informative for describing the severity of a delirium episode at a given time (see Figure 1). For the present study, we define delirium severity as the cumulative intensity of multi-domain symptoms or behaviors associated with delirium, which is captured in a continuous, quantitative scale.

### Expert panel.

Our expert panel consisted of nine clinicians, researchers and methodological experts, who generated new items to assess delirium severity, in following recommended reporting of instrument development and testing [12]. This panel included a broad range of interdisciplinary expertise in delirium measurement, including geriatricians (SKI, ERM, TTH), a behavioral neurologist (TGF), a psychiatrist (ERM), geriatric nurse practitioner (PAT), social worker (EMS), and methodologists and epidemiologists (SMV, RNJ). This was a local expert panel, and was distinct from the external expert panel that participated in the identification and definition of content domains (c.f., [7]).

### Item generation.

Items were generated in two ways: adapted from existing instruments and generated *de novo*. To generate new items, the experts adhered to the following steps: aligning the item with the domain definitions provided by our external expert panel [7], and crafting new items specifying the domain, interview instructions (e.g., for interaction with patient, family member, or nursing staff), time reference or look-back period (e.g., past day, past hour, from pre-hospitalization condition), symptom or sign to be assessed, and coding or response categories. The list of source instruments from which items were adapted is detailed in Appendix B. If the item was generated *de novo* by the local experts based on the domain definition, we refer to those items as *generated items* (Table 1). If the item was adapted from an existing or legacy instrument or research publication, we refer to those items as *adapted items* (Table 1).

After assembling a set of 303 items, the expert panel met to come to a consensus regarding items to include in our field survey. Panelists reviewed items for unique content, clarity (e.g., avoiding double barreled questions), alignment with the definition of the domains provided by our external expert panel [7], and suitability for use among older hospitalized patients. This resulted in the identification of a set of 103 items for the assessment of delirium severity (Table 1). These items were administered in the field study (including a daily hospital interview, described in the following section). The item set included three legacy instruments for comparison purposes (the Confusion Assessment Method-Severity [CAM-S] [13], Memorial Delirium Assessment Scale [MDAS] [14], and Delirium Rating Scale [DRS-R98] [15]).

### Field study.

The BASIL study is a prospective cohort study of 352 older adults hospitalized at the Beth Israel Deaconess Medical Center (BIDMC), a large academic medical center in Boston, MA (USA) with 673 beds and over 40,000 admissions per year. The purpose of the BASIL study is to develop and test new delirium severity measures, compare them with existing measures, and examine related clinical outcomes. *A complete description of the study has been previously published [16], and sample size calculations that account for losses given expected attrition were computed prior to enrollment of patients.*

We enrolled participants between October 2015 and March 2017. Participants were identified by chart review, recruited in the hospital, and followed up to one year. We obtained permission from participating hospitalists and surgeons to approach patients for consent and further eligibility determination. Eligible patients were aged 70 or older, English-speaking, admitted or transferred to the medical or surgical services as either emergency or elective admissions, and lived within 40 miles of BIDMC to facilitate long-term follow-up. *Patients were ineligible if they were unable to participate in cognitive assessment due to legal blindness or severe deafness, reported a recent history of heavy alcohol use (more than five drinks/day for men and four drinks/day for women) or alcohol withdrawal within the last six months, had a diagnosis of schizophrenia or an active psychosis, were non-communicative, had plans for immediate discharge, or were admitted for a terminal condition based chart review.* Signed informed consent was obtained from

participants whenever possible and was otherwise obtained from a health care proxy. Study procedures were approved by the Institutional Review Boards of BIDMC and Hebrew SeniorLife, the study coordinating center.

Trained lay interviewers conducted initial evaluations within 48 hours of hospital admission, followed by daily assessments during hospitalization. The initial assessment of about 45 minutes included collection of demographics, delirium status, and other study variables. On subsequent days, delirium and delirium severity were assessed daily with face-to-face, 10–15-minute interviews. Our protocol called for the collection of up to four daily interviews per patient, followed by interviews every other day until discharge; however, if delirium was present on the prior day based on the CAM [17], the daily interviews were continued. Data were recorded in a secure web-based database management system implemented within the Research Electronic Data Capture (REDCap) platform [18].

### Data analysis.

We report item analysis conducted using item response theory (IRT) and the generalized structural equation modeling (GSEM) modeling framework [19–21]. This framework is consistent with factor analysis for response variables that are ordinal [22–26]. Our analytic steps aimed to examine patient interview data on items designed to assess the domains defined by our external expert panel [7], to simultaneously evaluate the adequacy of the conceptual domain framework to guide the analysis of patient data, and to identify those items that most accurately measured the domains identified by the external expert panel. Our approach involved multiple iterative steps using confirmatory factor analysis (CFA) methods for ordinal dependent variables [27], including bi-factor measurement models [28]. Details on how we parameterized our models for delirium severity indicators, addressed item level skew, and derived summary model fit and robust parameter estimates have been previously described [29]. Our analysis was supported by Mplus software (version 8.2, Muthén & Muthén, Los Angeles CA) executed within the R environment (version 3.5, R Core Team, Vienna, Austria) and the Mplus Automation package [30]. The multilevel modeling capabilities of the Mplus software enables the control for non-independence of observations (the clustering of observations within patient) [31].

**Item parameters, item information, and test information.**—We present our results based on estimated item parameters, including standardized factor loadings (i.e., correlations between underlying and unobserved traits) and item thresholds (i.e., levels on the “latent” [unobserved] trait that are measured by the observed items). Summary data from these parameters include item and test information functions, which summarize measurement precision across the range of the latent trait. Information in this context is similar to the concept of Fisher Information [32], which enables quantification of the degree to which an unobserved variable (e.g., delirium severity) is accounted for by an observed item (e.g., responses/observations to delirium severity item ratings). Information functions vary across the possible range of the delirium severity items and peak where the item rating provides greater information. The location of this peak occurs where the thresholds are observed (i.e., how frequently persons are rated in specific categories of the item); the height of the peaks

depends on the strength of the correlation between the level of severity and the item response. More details in Appendix C.

**Reliability.**—To characterize the reliability of our set of items, we report internal consistency reliability using McDonald’s Omega coefficient [33]. This coefficient is interpreted similar to the Cronbach’s alpha coefficient [34]. Items with coefficient values > .80 were used to justify use in our delirium severity instrument [35].

**Model fit.**—Finally, we judge the adequacy of our latent variable models using standard methods for assessing model fit. This includes three statistics: the standardized root mean square residual (SRMR), the root mean squared error of approximation (RMSEA), and the confirmatory fit index (CFI). The SRMR is the mean standardized residual among mean and covariance estimates for model-implied and observed values, and values approach 0 as model fit improves. The computation is based on residuals only, and does not incorporate either the sample size or model complexity. We infer “approximate fit” when SRMR is 0.08. We also consider the RMSEA and CFI, both of which are computed based on the model  $\chi^2$ . The model  $\chi^2$  statistic is computed based on the discrepancy of observed and model-implied mean and covariance matrices. The CFI approaches 1 as model fit improves, and values of .95 or greater are taken as indicators of good fit. Values of the RMSEA approach 0 as model fit improves, and values of < 0.5 are taken as an indicator of good fit. The CFI and RMSEA incorporate model complexity in their calculation and thereby reward model parsimony [36, 37].

## Results

### Item generation, fielding, and retention.

Table 1 summarizes the item generation and retention process after winnowing among the possible items for inclusion in the initial set (termed “Initial analysis”) analysis and the final item set (termed “Final model”). Expert panelists generated 207 items and adapted 96 items from legacy instruments, resulting in a total of 303 items for consideration. We selected 103 items for inclusion in our daily hospital interview, and among these we included the severity items from comparator delirium severity instruments (described previously [38]). The expert panel made the decision to exclude one of the domains identified from our previous expert panel process: the domain capturing *Functional/Decline or low performance* in activities of daily living.

As summarized in Table 1, our final item set retains no items from the following broad / narrow domains: *Cognitive/Cognitive impairment* (as distinct from *Cognitive/Disorganized thinking* and *Cognitive/Disorientation*), *Psychiatric-behavioral/Inappropriate behavior*, *Emotional dysregulation/Depression*, and *Emotional dysregulation/Anger, hostility*. For all of these domains, we were unable to retain generated items in our final set (“Final model”) due to poor performance characteristics in psychometric data analysis (further described in the Table 3 results [below]).



### Field study.

Sample characteristics for the field study are presented in Table 2. The sample had a mean age of 80, 58% were women, and 15% were nonwhite or Hispanic. The sample had a relatively high level of education (mean 15 years), and a relatively large proportion had a recognized or chart diagnosis of dementia (n = 101, 29%). About 1 in 5 patients (n=68, 19%) satisfied CAM-criteria on at least one of their daily assessments during hospitalization.

### Psychometric data analysis.

For our psychometric data analysis, we used all daily assessments collected for the 352 patients (1,189 total daily assessments). Twelve assessments were excluded because the patient lacked sufficient communicative capacity (e.g., intubation, coma) to have a single item rated on any of the delirium severity comparator instruments (CAM-S, DRS-R98, MDAS). This resulted in 1,177 assessments that were included in the analysis.

As summarized in Table 1, we field-tested many more items (103 items) than we included in our initial psychometric data analysis (termed “Fielded,” 43 items). Many items in the *Cognitive* domain were excluded since they were only used to score the legacy items and were not part of our included domains. Other items were omitted due to lack of variability (too few participants were rated or responses for some levels were lacking) or collinearity with other items in the analysis set. To identify items with insufficient information for Mplus weighted least squares estimation, we used empirical methods to determine the variance and covariance parameters for model and parameter estimation. We additionally explored the possibility of generating composites of items when derived from the same domain but for which there was high collinearity or limited covariance. Ultimately, we rejected all composites in preference for retaining individual items. Composites necessitate strong assumptions about the equivalence of factor loadings and level of severity assessed by individual components, both assumptions that were not satisfied in our sample [39].

Reasons for dropping items and the results of the interim psychometric analysis models were as follows. The expert panel made the decision to exclude one of the domains identified from our previous expert panel process [7]: the domain capturing *Functional/Decline or low performance* in activities of daily living. This domain was excluded based on concerns in developing health outcome measurement instruments is ensuring that the selected item set includes only effect indicators or reflective indicators [8, 9]. These are observable or rated items that assess indicators that are theoretically caused by the underlying construct we are attempting to measure (e.g., delirium severity). Effect indicators are distinct from cause or formative indicators, which are indicators that might increase the risk or level of the underlying construct. Panelists felt that functional limitations, especially low performance or decline in activities of daily living, including walking, posture, and ability to transfer, did not fit in either of these two categories. Rather, functional limitations were considered to be potential outcomes of delirium, and potentially worse in the presence of more severe delirium. Moreover, this domain lacked specificity for delirium severity, which was necessary to serve as a meaningful effect indicator.

**Missing data.**—Observer-rated items tended to have missing values on less than 1% of assessments; however, items that required a response from the patient had a greater frequency of missing data (e.g., *Did you feel your thinking was slow?* had 46/1177, or about 4% missing).

After restricting our item set to indicators for which stable covariance parameters could be estimated, 43 candidate items remained (Table 1). These included 12 items in the broad *Cognitive* domain, 3 items in the *Level of consciousness* domain, 5 items in the *Inattention* domain, 12 items in the *Psychiatric-behavioral* domain, 5 in the *Emotional dysregulation* domain, 4 items in the *Psychomotor features* domain, and 2 items in the *Functional* domain (including 2 items in the *Sleep disorder* narrow domain). We fit a bi-factor model to these 43 items. A bi-factor model includes a general factor that loads in all items, and a set of specific factors that are mutually uncorrelated and uncorrelated with the general factor. We used as a specific factor structure the broad domains identified by our previous expert panel (Figure 1, see also [7]: *Cognitive, Level of consciousness, Inattention, Psychiatric-behavioral, Emotional dysregulation, Psychomotor features, and Functional (sleep)*). The model fit ranged from poor to adequate depending on the criterion used: the CFI was 0.91 (good fit judged when CFI > .95 with a maximum of 1); the RMSEA was 0.04 (good fit when less than .05 with a minimum of 0); and the SRMR was 0.12 (good fit judged when less than 0.08, to a minimum of 0). This model included some warnings relating to collinearity among indicators.

To address the lack of fit, we pursued a set of model modifications to narrow the number of items included in the item set, dropping the less frequently endorsed of item pairs where collinearity was indicated. We removed a specific factor and reduced the number of items in a specific broad domain if the loadings in the specific factor were larger than the loading in the general factor. We attended to the balance of items in terms of broad domain representation, and retained items involving direct questioning of patients.

Our “Final (psychometric) model” included 21 items (Table 3). Model fits for a single factor and empirically derived bi-factor model for these items fits slightly better than the initial bi-factor model using 43 indicators. The best fitting model was a bi-factor model with four specific factors (*level of consciousness, Inattention, psychiatric-behavioral, emotional dysregulation*; CFI = .93, RMSEA = 0.03, SRMR = 0.13).

In Figure 2 we display the test information functions for our final analytic model that includes 21 items listed in Table 3. We also show information functions for our three comparator instruments: CAM-S, MDAS, and DRS-R98. These information functions for comparator instruments have been described [29]. We also illustrate the distribution of estimated latent trait scores in our sample with a rug plot (shown along the x-axis). The rug helps focus attention on the region of the latent trait (delirium intensity) likely to be encountered in a relatively unselected sample of older hospitalized adults. The information plots reveal that the new item set compares favorably relative to the legacy items, providing greater information and therefore more precise and reliable measurement. The high internal consistency reliability of the new 21 item set is also supported by the Omega coefficients reported in Table 3.



Final item selections were determined by balancing coverage across the domains of delirium severity, and the item information values across the range of plausibly observed latent trait values [40]. The item information functions are shown within narrow domains in Figure 3. Each panel in Figure 3 illustrates the information for the set of items within the narrow domain as colored lines; the gray lines illustrate the information curves for all of the remaining items in the set of 21 items. The information plots help to identify items that may be redundant with each other, or to make choices for inclusion acknowledging both the strength of the item measurement of delirium intensity (the height of the information function) and the location along the latent trait where the item contributes information.

For example, in the panel labeled “*level of consciousness*”, there are two items that appear redundant (*consciousness fluctuates* and *level of consciousness*). These items can be identified as redundant by considering their psychometric contribution because they have a similar height to their information curves and provide information around the same region of the latent trait. A third item -- *exaggerated startle reflex* -- does not provide much information with regard to the general delirium intensity trait, and seems more likely to be relevant to persons with more extreme delirium severity than those included in our sample.

Our recommendations for retaining the items described in Table 3 include the retention of *level of consciousness* and *exaggerated startle reflex* and exclusion of *consciousness fluctuates* in our long form set, and additionally dropping *startle reflex* for our short form set. A similar logic guided our decisions regarding recommended items for long and short form (LF and SF, respectively) versions as indicated in Table 3.

## Discussion

Using advanced psychometric methodology, this study aimed to develop new delirium severity items via reporting of item generation, analysis, and reduction activities. Our process began with a set of domains within delirium severity that were previously defined by an expert panel of clinicians and researchers [7]. These expert panelists then generated 303 potential items (>200 items generated, nearly 100 items adapted from legacy instruments). Using data from a study of prospective hospitalized older patients, we further narrowed the item set to the 43 items considered in our “Initial Analysis” (see Table 1), for which we derived a set of 17-items for inclusion in a severity instrument.

For any medical disorder, severity is a complex topic and may mean different things to different stakeholders. From a phenomenological perspective, severity may reflect the intensity of specific signs and symptoms. From a clinical perspective, severity may reflect the likelihood of an adverse outcome or the urgency for symptom treatment (e.g., agitation). For patients and their families, severity may reflect distress or impaired functioning and recovery.

This study operationalizes delirium severity from a phenomenological perspective, as informed by clinical and research experts in the field, in order to develop a severity tool that will measure the intensity of a delirium episode with sufficiently high reliability to warrant individual patient level inference (e.g., guide treatment decisions; track individual changes in

severity over time) while comparing favorably to legacy instruments (MDAS, DRS-R-98, and CAM-S).

Current delirium severity measures provide mixed coverage of delirium signs and symptoms. The most frequently cited severity instruments include (in descending order): DRS [41] (and DRS-R-98 [15]), MDAS [14], NEECHAM (the Neelon and Champagne Confusion Scale [42]), and the Delirium Symptom Interview (DSI) [43]. The DRS and MDAS measure phenomenological features of delirium. Other tools mix indicators of presumed causal agents with phenomenological features (e.g., NEECHAM) or mix diagnostic features (e.g., acute onset or fluctuation of symptoms) with the magnitude of symptoms (e.g., DRS, DSI, Delirium-O-Meter [44]). Each of these severity measures is characterized by notable strengths, as well as important limitations. The MDAS, for example, was designed to exclude diagnostic and etiologic content from the severity score, and to be delivered several times in a single day to capture the fluctuating course of delirium; a key strength of this widely used tool. However, the MDAS is designed to be rated by a clinician (similar to the DRS), and does not provide separate calibration for hyperactive and hypoactive symptoms. Building upon the foundational groundwork of current delirium severity rating scales, our study combines the best aspects of these severity tools with the identification of domains, signs, and symptoms of delirium severity from our in-depth expert panel process and modern psychometric analysis of items from a field study to improve measurement of delirium severity.

We summarize the steps that our group has taken towards the measurement of delirium severity to-date, which includes the work undergone in this current study. First, we conducted a systematic review of the medical literature on assessment of delirium severity and evaluated the quality of existing tools for measuring this construct [38]. Second, we conducted a psychometric synthesis and harmonization of the three most commonly used delirium severity instruments to calibrate the measurement of delirium severity across these legacy instruments using advanced psychometric methods [29]. Third, we conducted in-depth qualitative interviews with patients, family caregivers, and bedside nurses to ensure inclusion of important domains of delirium severity as identified by these groups [45]. Fourth, we used a modified Delphi process involving an expert, interdisciplinary panel of clinical and research experts in delirium to define salient domains of delirium severity (see Appendix A) [7]. Finally, we conducted a prospective field study [16] to evaluate new delirium severity assessment items that we developed using rigorous standards for item development and evaluation in a sample derived from a target population in which the instrument is intended to be used (current study).

The strengths of our approach include the sequence of preliminary studies, defining the domains of delirium severity, the input from an interdisciplinary expert panel for domain definitions and item selection, the evaluation and testing of items in a prospective cohort, and the application of advanced psychometric methods in the final selection process. Despite these noteworthy strengths, several limitations are worth noting. First, our item reduction procedure is not fully automated based on the results of psychometric data analysis. Instead, we used multiple iterative steps to judge the psychometric results combined with clinical insight and expertise. While this approach to reducing an item set maximizes the clinical

relevance and content validity of the proposed instrument [46], it cannot be considered completely reproducible. A different team of clinical experts might have selected different items. Second, for many of our proposed items, either the construction of the question or our restricted sample size did not allow for the evaluation of the most intense symptoms of delirium severity. As a result of this limitation, we were forced to exclude many items from the psychometric data analysis based on insufficient variability (i.e., not enough patients were rated or endorsed the presence of a particular symptom). This was true for domains that we could not retain in our final model (shaded narrow domains in Figure 1). Finally, our final models fit well by some model fit indicators, but not well by others. In particular, our standardized root mean squared residuals indicate large residuals among the estimated item covariances. This may be a consequence of the extreme skew of the indicators, or it may be a signal that the common factor model is not entirely appropriate for modeling the intensity of a delirium severity episode. These limitations will need to be addressed in future research.

This study has laid the foundation for creation of a new delirium severity measure based on advanced psychometric and clinimetric input. Next steps will include development of the final instrument with precise wording of the key items and refined scoring system.

*Subsequently, evaluation of performance characteristics, including construct validity (e.g., convergent validity) and predictive validity, are a critical future step in assessing the utility of this new delirium severity tool. These are critical steps to advance the fundamental science of delirium measurement.*

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

See Appendix A for list of BASIL study investigators

### Funding

This work was supported by grants from the National Institutes of Health [K01AG057836 (SMV), R03AG061582 (SMV), R01AG030618 (ERM), K24AG035075 (ERM), R01AG044518 (SKI/RNJ), R24AG054259 (SKI), K07AG041835 (SKI), P01AG031720 (SKI)], and the Alzheimer's Association [AARF-18-560786 (SMV)]. All funding agencies were not involved in the design of the study and collection, analysis, or interpretation of data, or writing of the manuscript.

## Abbreviations:

<b>BIDMC</b>	Beth Israel Deaconess Medical Center
<b>BWH</b>	Brigham and Women's Hospital
<b>HMS</b>	Harvard Medical School
<b>HSL</b>	Hebrew SeniorLife
<b>PI</b>	principal investigator

## References

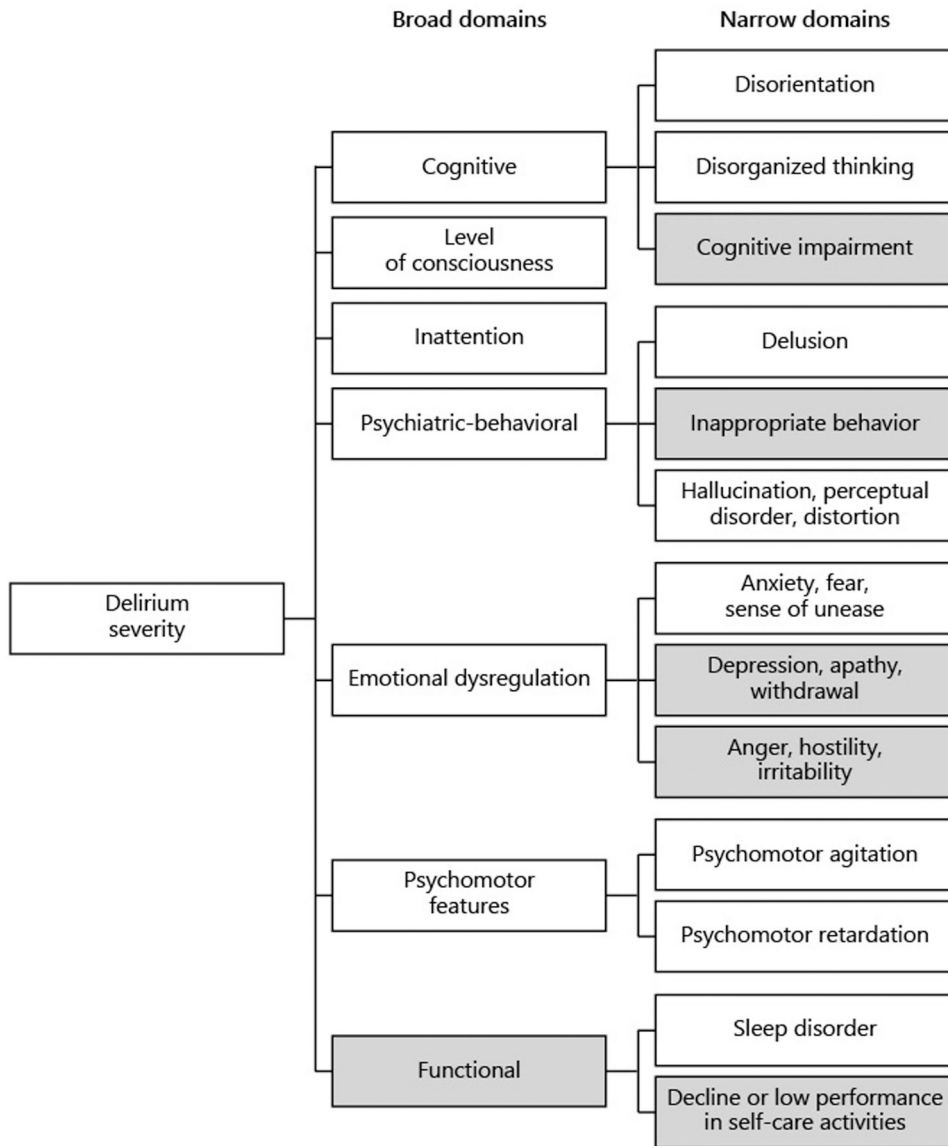
1. Engel GL, Romano J: Delirium, a syndrome of cerebral insufficiency. *J Chronic Dis* 1959, 9(2–3):260–277. [PubMed: 13631039]
2. Leslie DL, Marcantonio ER, Zhang Y, Leo-Summers L, Inouye SK: One-year health care costs associated with delirium in the elderly population. *Arch Intern Med* 2008, 168(1):27–32. [PubMed: 18195192]
3. Morandi A, Pozzi C, Milisen K, Hobbelen H, Bottomley JM, Lanzoni A, Tatzler VC, Carpena MG, Cherubini A, Ranhoff A et al.: An interdisciplinary statement of scientific societies for the advancement of delirium care across Europe (EDA, EANS, EUGMS, COTEC, IPTOP/WCPT). *BMC Geriatr* 2019, 19(1):253. [PubMed: 31510941]
4. Hayden KM, Inouye SK, Cunningham C, Jones RN, Avidan M, Davis D, Kuchel G, Khachaturian AS: Reduce the burden of dementia now. *Alzheimer's & Dementia* 2018, 14(7):845–847.
5. Altman DG, Royston P: The cost of dichotomising continuous variables. *BMJ* 2006, 332(7549):1080. [PubMed: 16675816]
6. Adamis D, Sharma N, Whelan PJP, Macdonald AJD: Delirium scales: A review of current evidence. *Aging Ment Health* 2010, 14(5):543–555. [PubMed: 20480420]
7. Schulman-Green D, Schmitt EM, Fong TG, Vasunilashorn SM, Gallagher J, Marcantonio ER, Brown CH, Clark D, Flaherty JH, Gleason A et al.: Use of an expert panel to identify domains and indicators of delirium severity. *Qual Life Res* 2019, 28(9):2565–2578. [PubMed: 31102155]
8. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HCW: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010, 63(7):737–745. [PubMed: 20494804]
9. de Vet HC, Terwee CB, Mokkink LB, Knol DL: *Measurement in medicine: a practical guide*: Cambridge University Press; 2011.
10. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M: The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years. *Med Care* 2007, 45(5):S3.
11. Cella D, Nowinski C, Peterman A, Victorson D, Miller D, Lai JS, Moy C: The Neurology Quality-of-Life Measurement Initiative. *Arch Phys Med Rehabil* 2011, 92(10):S28–S36. [PubMed: 21958920]
12. Streiner DL, Kottner J: Recommendations for reporting the results of studies of instrument and scale development and testing. *J Adv Nurs* 2014, 70(9):1970–1979. [PubMed: 24684713]
13. Inouye SK, Kosar CM, Tommet D, Schmitt EM, Puelle MR, Saczynski JS, Marcantonio ER, Jones RN: 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]
14. 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]
15. Trzepacz PT, Mittal D, Torres R, Canary K, Norton J, Jimerson N: Validation of the Delirium Rating Scale-Revised-98: comparison with the Delirium Rating Scale and the Cognitive Test for Delirium. *The Journal of neuropsychiatry and clinical neurosciences* 2001, 13(2):229–242. [PubMed: 11449030]
16. Hshieh TT, Fong TG, Schmitt EM, Marcantonio ER, D'Aquila ML, Gallagher J, Xu G, Guo YR, Abrantes TF, Bertrand SE et al.: The Better Assessment of Illness Study for Delirium Severity: Study Design, Procedures, and Cohort Description. *Gerontology* 2019, 65(1):20–29. [PubMed: 30032141]
17. 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]
18. Harris P, Taylor R, Thielke R, Payne J, Gonzalez N, Conde J: Research electronic data capture (REDCap)--A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inf* 2009, 42(2):377–381.

19. Muthén B: A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* 1984, 49(1):115–132.
20. Birnbaum A: Some latent trait models (chapter 17) In: *Statistical Theories of Mental Test Scores*. Edited by Lord F, Novick M. Reading, MA: Addison-Wesley; 1968: 397–424.
21. Mellenbergh GJ: Generalized linear item response theory. *Psychol Bull* 1994, 115:300–300.
22. Lord F, Novick M: Latent traits and item characteristic functions (Chapter 16) In: *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley; 1968: 358–393.
23. Bock RD, Lieberman M: Fitting a response model for n dichotomously scored items. *Psychometrika* 1970, 35(2):179–197.
24. Mislevy RJ: Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics* 1986, 11(1):3–31.
25. Takane Y, De Leeuw J: On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* 1987, 52(3):393–408.
26. Glockner-Rist A, Hoijsink H: The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling* 2003, 10(4):544–565.
27. Muthén B: Dichotomous factor analysis of symptom data. *Sociological Methods and Research* 1989, 18(1):19–65.
28. Holzinger KJ, Swineford F: The bi-factor method. *Psychometrika* 1937, 2(1):41–54.
29. Gross AL, Tommet D, D’Aquila M, Schmitt E, Marcantonio ER, Helfand B, Inouye SK, Jones RN, Group BS: Harmonization of delirium severity instruments: a comparison of the DRS-R-98, MDAS, and CAM-S using item response theory. *BMC Med Res Methodol* 2018, 18(1):92. [PubMed: 30200896]
30. Halquist M: Mplus Automation. R package version 04–2 2011.
31. Muthén B: Beyond SEM: General latent variable modeling. *Behaviormetrika* 2002, 29(1):81–117.
32. Fisher RA: On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character* 1922, 222(594–604):309–368.
33. Peters G-JY: The alpha and the omega of scale reliability and validity: why and how to abandon Cronbach’s alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist* 2014, 16(2):56–69.
34. Cronbach L: Test “reliability”: its meaning and determination. *Psychometrika* 1947, 12(1):1–16. [PubMed: 20293842]
35. Nunnally JC, Bernstein IH: *Psychometric theory*, 3rd edn New York: McGraw-Hill College Division; 1994.
36. Lt Hu, Bentler PM: Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal* 1999, 6(1):1–55.
37. SRMR in Mplus [<http://www.statmodel.com/download/SRMR2.pdf>]
38. Jones RN, Cizginer S, Pavlech L, Albuquerque A, Daiello LA, Dharmarajan K, Gleason LJ, Helfand B, Massimo L, Oh E et al.: Assessment of Instruments for Measurement of Delirium Severity: A Systematic Review. *JAMA Intern Med* 2019, 179(2):231–239. [PubMed: 30556827]
39. Bandalos DL: Is parceling really necessary? A comparison of results from item parceling and categorical variable methodology. *Structural Equation Modeling* 2008, 15(2):211–240.
40. Kertesz SG, Pollio DE, Jones RN, Steward J, Stringfellow EJ, Gordon AJ, Johnson NK, Kim TA, Daigle SG, Austin EL et al.: Development of the Primary Care Quality-Homeless (PCQ-H) Instrument A Practical Survey of Homeless Patients’ Experiences in Primary Care. *Med Care* 2014, 52(8):734–742. [PubMed: 25023918]
41. Trzepacz PT, Baker RW, Greenhouse J: A symptom rating scale for delirium. *Psychiatry Res* 1988, 23(1):89–97. [PubMed: 3363018]
42. Neelon VJ, Champagne MT, Carlson JR, Funk SG: The NEECHAM Confusion Scale: construction, validation, and clinical testing. *Nurs Res* 1996, 45(6):324. [PubMed: 8941300]

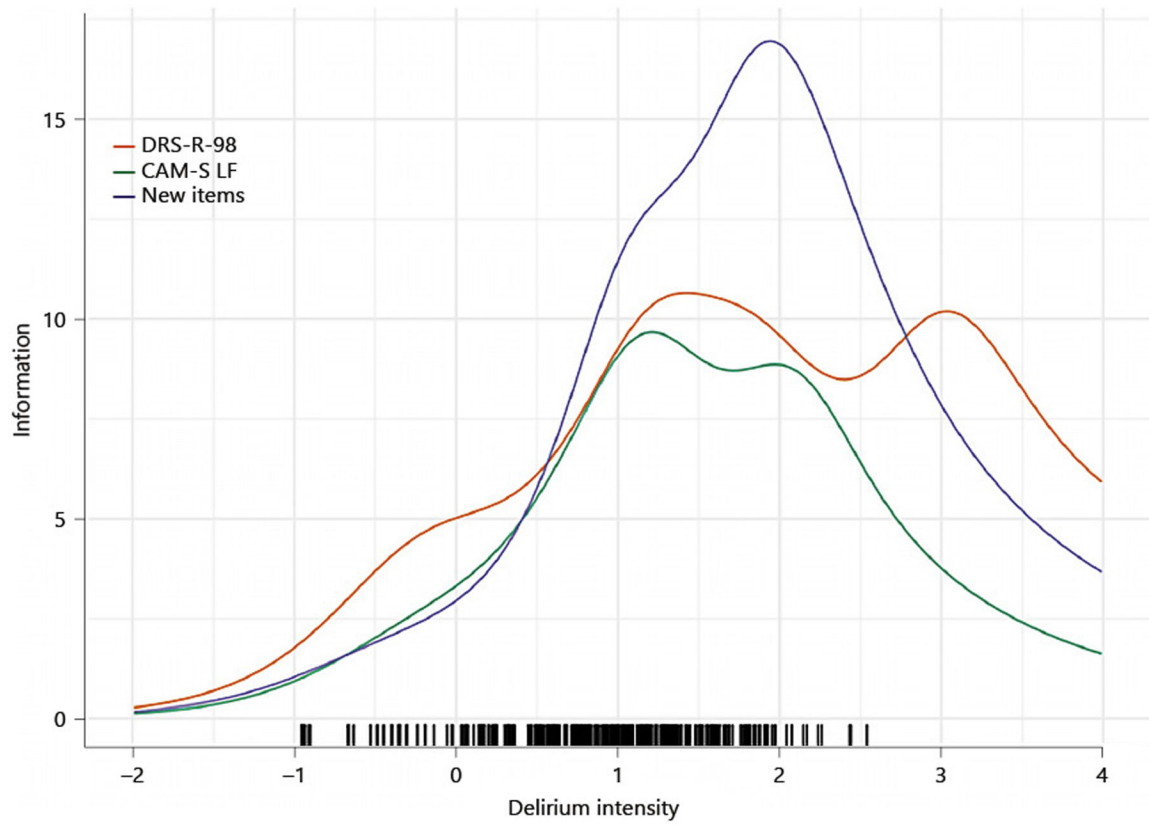
43. Albert MS, Levkoff SE, Reilly C, Liptzin B, Pilgrim D, Cleary PD, Evans D, Rowe JW: The Delirium Symptom Interview: An interview for the detection of delirium symptoms in hospitalized patients. *J Geriatr Psychiatry Neurol* 1992, 5(1):14–21. [PubMed: 1571069]
44. de Jonghe JFM, Kalisvaart KJ, Timmers JFM, Kat MG, Jackson JC: Delirium-O-Meter: a nurses' rating scale for monitoring delirium severity in geriatric patients. *Int J Geriatr Psychiatry* 2005, 20(12):1158–1166. [PubMed: 16315151]
45. Schmitt E, Gallagher J, Inouye S, Schulman-Green D: Perspectives on the delirium experience and its burden: Common themes among older patients, their family caregivers, and nurses. *Gerontologist* 2019, 59(2):327–337. [PubMed: 30870568]
46. Feinstein AR: Multi-item Instruments vs Virginia Apgar's Principles of Clinimetrics. *Arch Intern Med* 1999, 159(2):125–128. [PubMed: 9927093]
47. Katz S, Ford A, Moskowitz R, Jackson B, Jaffe M: Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association* 1963, 185(12):914–919. [PubMed: 14044222]
48. Morita T, Tsunoda J, Inoue S, Chihara S, Oka K: Communication Capacity Scale and Agitation Distress Scale to measure the severity of delirium in terminally ill cancer patients: a validation study. *Palliat Med* 2001, 15(3):197–206. [PubMed: 11407191]
49. Otter H, Martin J, Basell K, von Heymann C, Hein OV, Böllert P, Jänsch P, Behnisch I, Wernecke KD, Konertz W: Validity and reliability of the DDS for severity of delirium in the ICU. *Neurocritical care* 2005, 2(2):150–158. [PubMed: 16159057]
50. 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]
51. Meagher D, Moran M, Raju B, Leonard M, Donnelly S, Saunders J, Trzepacz P: A new data-based motor subtype schema for delirium. *The Journal of neuropsychiatry and clinical neurosciences* 2008, 20(2):185–193. [PubMed: 18451189]
52. Schuurmans MJ, Shortridge-Baggett LM, Duursma SA: The Delirium Observation Screening Scale: A screening instrument for delirium. *Research and theory for nursing practice* 2003, 17(1):31–50. [PubMed: 12751884]
53. O'Keeffe ST: Rating the severity of delirium: the Delirium Assessment Scale. *Int J Geriatr Psychiatry* 1994, 9(7):551–556.
54. Viosca E, Martínez JL, Almagro PL, Gracia A, González C: Proposal and validation of a new functional ambulation classification scale for clinical use. *Arch Phys Med Rehabil* 2005, 86(6):1234–1238. [PubMed: 15954065]
55. Giladi N, Shabtai H, Simon E, Biran S, Tal J, Korczyn A: Construction of freezing of gait questionnaire for patients with Parkinsonism. *Parkinsonism & related disorders* 2000, 6(3):165–170. [PubMed: 10817956]
56. Eastlack ME, Arvidson J, Snyder-Mackler L, Danoff JV, McGarvey CL: Interrater reliability of videotaped observational gait-analysis assessments. *Phys Ther* 1991, 71(6):465–472. [PubMed: 2034709]
57. Bergeron N, Dubois MJ, Dumont M, Dial S, Skrobik Y: Intensive Care Delirium Screening Checklist: evaluation of a new screening tool. *Intensive Care Med* 2001, 27(5):859–864. [PubMed: 11430542]
58. Jette AM, Haley SM, Coster WJ, Kooyoomjian JT, Levenson S, Heeren T, Ashba J: Late life function and disability instrument: I. Development and evaluation of the disability component. *The Journals of Gerontology Series A, Biological Sciences and Medical Sciences* 2002, 57(4):M209–216.
59. Newell AM, VanSwearingen JM, Hile E, Brach JS: The modified gait efficacy scale: establishing the psychometric properties in older adults. *Phys Ther* 2012, 92(2):318–328. [PubMed: 22074940]
60. Champagne M, Neelon V, McConnell E, Funk S: The NEECHAM Confusion Scale: Assessing acute confusion in the hospitalized and nursing home elderly. *The Gerontologist* 1987, 27(4A):473–480.



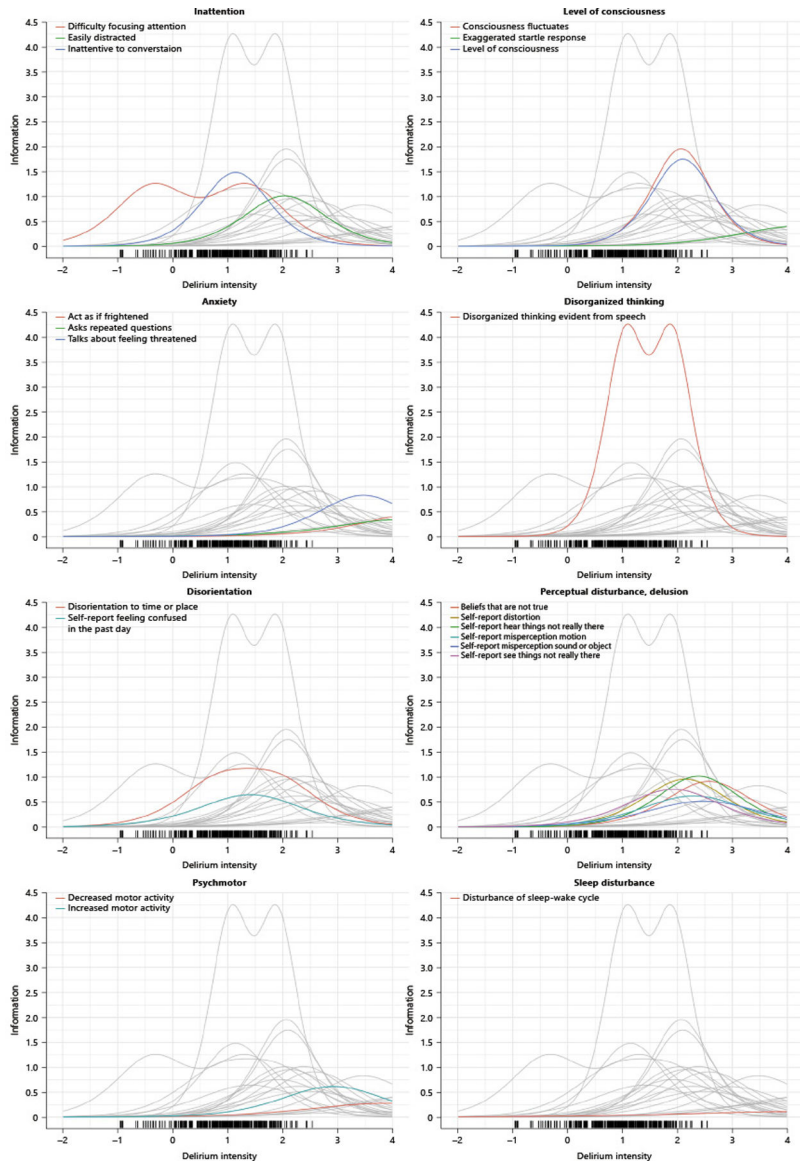
61. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J: The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology* 1994, 44(12):2308–2314. [PubMed: 7991117]
62. NINDS: User manual for the quality of life in neurological disorders (Neuro-QoL) measures (version 2). In.: National Institute of Neurological Disorders and Stroke; 2015.
63. Gaudreau JD, Gagnon P, Harel F, Roy MA: Impact on delirium detection of using a sensitive instrument integrated into clinical practice. *Gen Hosp Psychiatry* 2005, 27(3):194–199. [PubMed: 15882766]
64. Tiegés Z, McGrath A, Hall RJ, MacLulich AM: Abnormal level of arousal as a predictor of delirium and inattention: an exploratory study. *The American Journal of Geriatric Psychiatry* 2013, 21(12):1244–1253. [PubMed: 24080383]
65. Bellelli G, Speciale S, Morghen S, Torpilliesi T, Turco R, Trabucchi M: Are fluctuations in motor performance a diagnostic sign of delirium? *JAMDA* 2011, 12(8):578–583. [PubMed: 21450181]
66. Fong TG, Tulebaev SR, Inouye SK: Delirium in elderly adults: diagnosis, prevention and treatment. *Nature Reviews Neurology* 2009, 5(4):210–220. [PubMed: 19347026]
67. Godfrey A, Conway R, Leonard M, Meagher D, Ólaighin G: A classification system for delirium subtyping with the use of a commercial mobility monitor. *Gait Posture* 2009, 30(2):245–252. [PubMed: 19539473]
68. Salzman B: Gait and balance disorders in older adults. *Am Fam Physician* 2010, 82(1):61–68. [PubMed: 20590073]
69. Krogseth M, Bruun Wyller T, Engedal K, Juliebo V: Delirium is an important predictor of incident dementia among elderly hip fracture patients. *Dement Geriatr Cogn Disord* 2011, 31:63–70. [PubMed: 21212674]
70. Wahlund LO, Björln GA. Delirium in clinical practice. Experience from a Specialized Delirium Ward. *Dement Geriatr Cogn Disord* 1999, 10:389–392. [PubMed: 10473945]
71. Eshmaey M, Ledschbor-Frahn C, Guenther U, Popp J: Preoperative depression and plasma cortisol levels as predictors of delirium after cardiac surgery. *Dement Geriatr Cogn Disord* 2020, epub ahead of print doi:10.1159/000505574.



**Figure 1: Domain hierarchy of delirium severity: broad and narrow domains.**  
 Domain coverage of final item set is indicated by shading of domain boxes: gray-filled boxes are not represented in our final item set



**Figure 2:** Test information function, unidimensional model and previously published test information functions for the CAM-S, DRS-R-98, and MDAS. Better Assessment of Illness (BASIL) study (N=352).



**Figure 3.**  
Item information functions by domain

**Table 1:**

Number of items generated, items included in psychometric models, items retained in the final model.

<b>Domain</b>	<b>Gen-erated</b>	<b>Adapted</b>	<b>Fielded</b>	<b>Initial analysis</b>	<b>Final model</b>	<b>Recommended</b>
<b>Cognitive</b>	<b>25</b>	<b>36</b>	<b>43</b>	<b>12</b>	<b>3</b>	<b>3</b>
Disorganized thinking	14	5	5	3	1	1
Disorientation	5	15	16	3	2	2
Cognitive impairment	6	16	22	6	0	0
<b>Level of consciousness</b>	<b>21</b>	<b>5</b>	<b>5</b>	<b>3</b>	<b>3</b>	<b>2</b>
<b>Inattention</b>	<b>18</b>	<b>9</b>	<b>9</b>	<b>5</b>	<b>3</b>	<b>2</b>
<b>Psychiatric-behavioral</b>	<b>21</b>	<b>12</b>	<b>12</b>	<b>12</b>	<b>6</b>	<b>6</b>
Delusions	4	3	3	3	1	1
Inappropriate behavior	7	1	1	1	0	0
Perceptual disturbance, hallucination, distortion	10	8	8	8	5	5
<b>Emotional dysregulation</b>	<b>43</b>	<b>18</b>	<b>18</b>	<b>5</b>	<b>3</b>	<b>2</b>
Anxiety	10	5	5	3	3	2
Depression	19	7	7	2	0	0
Anger, hostility	14	6	6	0	0	0
<b>Psychomotor features</b>	<b>9</b>	<b>6</b>	<b>6</b>	<b>4</b>	<b>2</b>	<b>2</b>
Agitation	6	3	3	2	1	1
Retardation	3	3	3	2	1	1
<b>Functional</b>	<b>70</b>	<b>10</b>	<b>10</b>	<b>2</b>	<b>1</b>	<b>0*</b>
Sleep disorder	36	9	9	2	1	1
Decline or low performance	34	1	1	0	0	0
<b>Total</b>	<b>207</b>	<b>96</b>	<b>103</b>	<b>43</b>	<b>21</b>	<b>17</b>

Notes: Generated items were items created by local expert panel members. Adapted items were chosen from legacy items included in the field study, which were drawn from existing assessment instruments, included in whole or in part.

\* The expert panel made the decision to exclude this domain (*Functional/Decline or low performance* in activities of daily living).

**Table 2.**

Baseline characteristics of the BASIL sample (N=352)

Characteristic	Mean (SD) or N (%)	
Age, years, mean (SD)	80.3	(6.8)
Female sex, n (%)	203	(58)
Non-white race, n (%)	52	(15)
Years of education, mean (SD)	14.5	(3.0)
Married, n (%)	139	(40)
Lives alone, n (%)	135	(39)
Lives in nursing home, n (%)	13	(4)
Dementia, n (%)	101	(29)
Charlson comorbidity score, mean (SD)	2.2	(2.2)
Surgical patient, n (%)	102	(29)
Delirium, n (%)	68	(19)

Abbreviations: BASIL, Better Assessment of Illness; SD, standard deviation

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 3:**

Standardized measurement slopes (factor loadings) for unidimensional and bi-factor measurement models of delirium severity. Better Assessment of Illness (BASIL) study, N = 352 patients and n = 1,117 patient assessments.

Domain	Feature	Factor loadings			
		f	fb	sb	LF
LOC	Consciousness fluctuates	.74	.77	.69	s1
LOC	Level of consciousness (camlf4a)	.71	.72	.62	s1 *
LOC	Exaggerated startle response	.58	.62	-.37	s1 *
INA	Inattentive to conversation	.87	.81	.52	s2
INA	Difficulty focusing attention	.81	.76	.35	s2 *
INA	Easily distracted	.70	.64	.40	s2 *
DTH	Disorganized thinking evident from speech	.88	.92		*
DIS	Disorientation to time or place	.71	.74		*
DIS	Self-report feeling confused in the past day	.67	.70		*
ANX	Talks about feeling threatened	.49	.42	.79	s3 *
ANX	Acts as if frightened	.49	.40	.79	s3 *
ANX	Asks repeated questions	.44	.40	1.00	s3
PER	Self-report see things not really there	.72	.54	1.00	s4 *
PER	Self-report hear things not really there	.71	.62	.42	s4 *
PER	Self-report misperception motion	.71	.56	.52	s4 *
PER	Self-report distortion	.70	.61	.46	s4 *
PER	Beliefs that are not true	.63	.64	.08	s4 *
PER	Self-report misperception sound or object	.57	.50	.40	s4 *
PM	Increased motor activity	.60	.62		*
PM	Decreased motor activity	.44	.45		*
FUN	Disturbance of sleep-wake cycle	.46	.48		
	Omega	.94	.93		.93
	RMSEA	.04	.03		-
	CFI	.92	.95		-
	SRMR	.15	.13		-

Notes/abbreviations: f, loading and model fits in single factor model; fb, sb, loadings on general (fb) and specific (sb) factors in bifactor model; Omega refers to McDonald's coefficient for internal consistency reliability; CFI, Confirmatory Fit Index where values of 0.95 or greater are considered good; RMSEA, root mean squared error of approximation where values below 0.05 are considered good; SRMR, standardized root mean squared residual, where values below 0.08 are considered good; LF, long form indicates with "\*" whether the item is retained for recommended long form; Domain abbreviations are LOC, level of consciousness; INA, inattention; DTH, disorganized thinking; DIS, disorientation; ANX, anxiety; PER, perceptual disturbance, delusion; PM, psychomotor symptoms; FUN, functional.