



Published in final edited form as:

J Am Geriatr Soc. 2016 August ; 64(8): 1684–1689. doi:10.1111/jgs.14234.

Derivation and Validation of a Severity Scoring Method (3D-CAM-S) for the 3-Minute Diagnostic Interview for CAM-defined Delirium

Sarinnapha M. Vasunilashorn, PhD^{a,b,c,*}, Jamey Guess, MS^{a,*}, Long Ngo, PhD^{a,b}, Donna Fick, PhD, RN^{c,d}, Richard N. Jones, ScD^{c,f}, Eva Schmitt, PhD^c, Cyrus M. Kosar, MA^e, Jane S. Saczynski, PhD^{c,g}, Thomas G. Trivison, PhD^{b,c,h}, Sharon K. Inouye, MD, MPH^{a,b,c,h,†}, and Edward R. Marcantonio, MD, SM^{a,b,c,h,†}

^aDivision of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, Boston, Massachusetts ^bHarvard Medical School, Boston, Massachusetts ^cAging Brain Center, Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts ^dCollege of Nursing and College of Medicine, Department of Psychiatry, Pennsylvania State University, University Park, Pennsylvania ^eDepartment of Health Services, Policy, and Practice, Brown School of Public Health, Providence, RI ^fDepartment of Psychiatry and Human Behavior, Warren Alpert Medical School, Brown University, Providence, Rhode Island ^gDepartment of Pharmacy and Health Systems Sciences, Northeastern University, Boston MA ^hDivision of Gerontology, Beth Israel Deaconess Medical Center, Boston, Massachusetts

Abstract

Background/Objectives—Delirium severity confers additional prognostic information beyond diagnosis, and is useful for monitoring the course of delirium and its response to treatment. Our objective was to derive and validate a method for scoring delirium severity using the 3D-CAM, a recently validated, brief, structured diagnostic interview for Confusion Assessment Method (CAM)-defined delirium, and to demonstrate its agreement with the CAM Severity (CAM-S) short form (SF) measure as the reference standard.

Design—Derivation and validation analysis in a prospective cohort study.

Address correspondence to: Sarinnapha Vasunilashorn, PhD, Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center, 1309 Beacon Street, Brookline, MA 02446, Tel: (617) 754-1424, Fax: (617) 754-1440, svasunil@bidmc.harvard.edu.

*Co-first authors

†Co-senior authors

Conflict of Interest: The authors have no conflicts of interest to disclose.

Author Contributions: Study concept and design: Guess, Vasunilashorn, Ngo, Fick, Jones, Schmitt, Kosar, Saczynski, Trivison, Inouye, Marcantonio

Acquisition of subjects and/or data: Ngo, Jones, Schmitt, Kosar, Trivison, Inouye, Marcantonio

Analysis and interpretation of data: Guess, Vasunilashorn, Ngo, Fick, Jones, Schmitt, Kosar, Saczynski, Trivison, Inouye, Marcantonio

Preparation of manuscript: Guess, Vasunilashorn, Ngo, Fick, Jones, Schmitt, Kosar, Saczynski, Trivison, Inouye, Marcantonio

Sponsor's Role: None of the sponsors were involved in the design, methods, subject recruitment, data collection, analysis, or preparation of the paper.

Note: A revised version of the 3D-CAM instrument (Version 3.0 or higher) to be used for creating a 3D-CAM-S severity score using the new methods presented here is now available at www.hospitalelderlifeprogram.org

Setting—Two academic medical centers.

Participants—Patients age ≥ 70 undergoing major elective non-cardiac surgery enrolled in the Successful Aging after Elective Surgery Study (566 patients).

Measurements—The sample was randomly divided into a derivation dataset (N=377) and an independent validation dataset (N=189). These datasets were used to: 1) develop a severity scoring method using the 3D-CAM based on the 4-item CAM-S SF (3D-CAM-S), and 2) evaluate agreement between the 3D-CAM-S and the traditional CAM-S SF using weighted kappa statistics.

Results—A method for scoring severity using 3D-CAM items was developed that achieved high agreement with the CAM-S SF in the derivation dataset: kappa=0.94 (95% confidence interval [CI] 0.93-0.95). Importantly, the 3D-CAM-S achieved nearly identical agreement in the independent validation dataset: kappa=0.93 (95% CI 0.92-0.95). Moreover, 100% of the 3D-CAM-S scores were within one point of the CAM-S SF score in both datasets. The 3D-CAM-S also strongly predicts clinical outcomes.

Conclusion—Our newly developed method for scoring delirium severity using the 3D-CAM (the 3D-CAM-S) has excellent agreement with the CAM-S SF. This new methodology enables clinicians and researchers using the 3D-CAM for surveillance to simultaneously measure delirium severity and monitor its course by tracking changes over time. The 3D-CAM-S expands the utility of the 3D-CAM as an important tool for delirium recognition and management.

Keywords

delirium; delirium severity; Confusion Assessment Method

INTRODUCTION

Delirium is a common, morbid, and costly geriatric syndrome associated with numerous poor outcomes among older patients^{1,2,3}. To address this important problem, accurate diagnosis is critical, and concurrent measurement of delirium severity confers additional prognostic information beyond diagnosis, and allows clinicians to monitor the course of delirium and its response to treatment⁴. From the research standpoint, delirium severity can serve as an important predictor or outcome measure for clinical trials, prognosis studies, and pathophysiologic investigations.

The Confusion Assessment Method (CAM)⁵ has been widely used to diagnose delirium, and a scoring system for delirium severity based on the CAM (the CAM-S) has been developed and demonstrated to have predictive validity for important clinical outcomes⁶. The CAM has gained widespread acceptance for identification of delirium. For optimal use, the CAM should be rated based on formal cognitive testing; however, the exact test used has been left to the discretion of the user. The most common tests currently utilized to score the CAM include the Short Portable Mental Status Questionnaire⁷, the Mini-Cog⁸, and the Montréal Cognitive Assessment (MoCA)⁹. None of these tests were designed specifically to operationalize the CAM, and the cognitive test chosen may influence both the time required and accuracy of CAM scoring. The CAM-S has two forms: long and short (LF and SF). The CAM-S LF assesses the severity of 10 features, while the CAM-S SF considers only the four

features in the CAM diagnostic algorithm^{5,6}. Both the CAM-S LF and SF have shown evidence of validity for predicting important clinical outcomes. Due to its brevity, the CAM-S SF may be preferred for clinical purposes.

The 3-Minute Diagnostic Interview for CAM-defined Delirium (3D-CAM) is a brief 20-item (10 cognitive testing items, 10 interviewer observations) assessment based on the CAM algorithm. It was derived from a 160-item structured delirium assessment using Item Response Theory (IRT) and model selection techniques¹⁰, as described previously^{11,12}. As opposed to other cognitive screening tests, the 3D-CAM items were selected specifically to operationalize the CAM algorithm. The 3D-CAM has been prospectively validated in a cohort of 201 general medicine patients, achieving a sensitivity of 95% and a specificity of 94% relative to a clinical reference standard¹¹. While accurate for diagnosis, there is currently no method for measuring delirium severity using the 3D-CAM. Therefore, the specific aims of the present study were: 1) to develop an objective method for scoring delirium severity using the 3D-CAM based on the CAM-S SF, called the 3D-CAM-S, and 2) to determine the agreement between the 3D-CAM-S and the CAM-S SF in an initial derivation and independent validation dataset comprised of older adults undergoing major elective surgery. The development of a delirium severity scoring algorithm for the previously validated 3D-CAM¹¹ will enable this brief 3 minute interview to also provide information on the severity of delirium, which is currently only available using longer assessments, including the CAM-S and the Memorial Delirium Assessment Scale (MDAS)^{13,14}.

METHODS

Study population

The Successful Aging after Elective Surgery (SAGES) study is an ongoing prospective cohort study of older adults undergoing major scheduled non-cardiac surgery. The study design and methods have been described in detail previously^{15,16}. In brief, eligible participants were age 70 years and older, English speaking, scheduled to undergo elective surgery at two Harvard-affiliated academic medical centers and had an anticipated length of stay of at least 3 days. Eligible surgical procedures were: total hip or knee replacement, lumbar, cervical, or sacral laminectomy, lower extremity arterial bypass surgery, open abdominal aortic aneurysm repair, and open or laparoscopic colectomy. Exclusion criteria included evidence of dementia, delirium, hospitalization within 3 months of screening, terminal condition, legal blindness, severe deafness, history of schizophrenia or psychosis, and history of alcohol abuse or withdrawal. A total of 566 patients met all eligibility criteria and were enrolled between June 18, 2010 and August 8, 2013.

Written informed consent for study participation was obtained from SAGES participants using procedures approved by the institutional review boards of Beth Israel Deaconess Medical Center and Brigham and Women's Hospital, the two study hospitals, and Hebrew SeniorLife, the study coordinating center, all located in Boston, Massachusetts.

Delirium Assessments

Delirium assessments were conducted daily in the SAGES study from postoperative day 1 through discharge¹⁵ by a team of highly trained and standardized interviewers, who underwent regular inter-rater reliability assessments. The CAM long form was completed based on a standardized cognitive assessment, which included 10 orientation items, short-term recall, digit span forwards and backwards, months of the year backwards, and days of the week backwards. Delirium was scored as present/absent based on the CAM algorithm⁵, which requires the presence of 1) acute or fluctuating course, 2) inattention, and either 3) disorganized thinking or 4) altered level of consciousness. The CAM-S SF was created based on severity rating of these 4 CAM diagnostic features⁶. Acute change or fluctuating course was scored 0 points if absent or 1 if present. The remaining three CAM diagnostic features were scored as 0 points for no symptoms, 1 point for mild symptoms, and 2 points for marked symptoms. The CAM-S SF score represents a sum of points assigned for the four features (range 0-7; 7=most severe). The 3D-CAM was not directly administered to SAGES participants; however, all of its component items were contained within the SAGES delirium assessments, which were used to simulate 3D-CAM assessments using this subset of items.

Derivation of the 3D-CAM-S scoring method

The SAGES sample of 566 patients was randomly divided into an initial derivation dataset (377 patients, 1369 assessments), and a validation dataset (189 patients, 692 assessments). Using the derivation dataset, we derived a scoring method for each of the 4 CAM diagnostic features such that the 3D-CAM-S score aligned with the CAM-S SF score. For Feature 1 (Acute Change or Fluctuating Course), the 3D-CAM score mirrored that of the CAM-S: 0 (absent) or 1 (present). For Feature 2 (Inattention) and Feature 3 (Disorganized Thinking) we used an empirical approach that incorporated both observational items and cognitive testing results, as described below. For Feature 4 (Altered Level of Consciousness), we used an *a priori*-based approach described below to assign 3D-CAM-S scores.

Feature 1—For Feature 1, the 3D-CAM contains 3 observational items and 3 direct patient questions. The 3D-CAM-S score for Feature 1 mapped directly to the CAM SF score, such that if any of these items were “positive”, Feature 1 would be present (1 point). If none of the items were “positive”, Feature 1 would be absent (0 points) (see Figure 1 for details).

Features 2 and 3—For CAM features 2 and 3, we used classification trees to prioritize and divide observational and cognitive testing items within the 3D-CAM to maximize discrimination between a CAM-S SF rating of 0, 1 or 2¹⁷. For Feature 2 (Inattention), the 3D-CAM contains 2 observational items and 4 cognitive testing items. If neither of the 2 observational items was present, then Feature 2 was scored 0. If 1 or more observational items were present, Feature 2 was rated 1 (“mild”) if the patient answered 0 to 2 cognitive items incorrectly, and rated 2 (“marked”) if the patient answered 3 or 4 cognitive items incorrectly (Figure 1). For Feature 3 (Disorganized Thinking), the 3D-CAM contains 3 observational items and 3 cognitive testing items. If none of the observational items was present, Feature 3 was scored 0. If 1 or more observational items were present, Feature 3 was rated 1 (“mild”) if the patient answered 0 or 1 cognitive item incorrectly and rated 2 (“marked”) if the patient answered 2 or 3 cognitive items incorrectly (Figure 1).

Feature 4—For CAM Feature 1 (Altered Level of Consciousness), we used an *á priori*-based approach to replicate the CAM-S score from the 3D-CAM items. 3D-CAM Feature 4 consists of only 2 observational items, the first assessing decreased states of arousal, and the second assessing increased arousal. Question 1 of 3D-CAM Feature 4 was split into 2 parts: patients received 1 point for being sleepy and 2 points for being stuporous or comatose. For question 2 of the 3D-CAM Feature 4, patients received 1 point for hypervigilance. Patients received 0 points if both 3D-CAM Feature 4 items were scored as normal.

Once the score for each feature was determined, the overall 3D-CAM-S score was computed as the sum of the severity scores for each Feature, yielding a final score of 0-7, 7 most severe. Using the same approach as had been used for the CAM-S⁶, we categorized the 3D-CAM-S scores into none, mild, moderate, and severe categories based on the sample distribution.

3D-CAM-S and CAM-S Agreement, Distributions

We determined agreement between the overall 3D-CAM-S and CAM-S SF scores using weighted kappa statistics in both the derivation dataset and independent validation dataset. To better understand 3D-CAM-S performance, we used the validation dataset to examine agreement within each of the four CAM Features, and to determine reasons for misclassification. We used scatter plots to examine the distribution of both 3D-CAM-S and CAM-S SF scores relative to the presence or absence of delirium. Finally, we performed a sensitivity analysis assessing the effect of clustering of delirium assessments within patients, calculating a weighted average of patient-specific kappa statistics with weights given as 1/(variance of the patient-specific kappa).

All analyses were performed using SAS version 9.3 (Cary, NC) and R v3.0.2 with the party package v1.0-14 (Vienna, Austria).

3D-CAM-S Validation Using Clinical Outcomes

To examine the predictive validity of the 3D-CAM-S, we measured its association with three short-term clinical outcomes: hospital length of stay (LOS), institutional discharge, and 30-day hospital readmission (See Supplemental Materials for details).

RESULTS

Cohort characteristics and prevalence of delirium

The 566 enrolled patients yielded 2061 hospital delirium assessments with a median of 3 assessments per patient (range 1-14). On average, patients were age 76.7 (standard deviation [SD] 5.2), 58% were female, and 93% were non-Hispanic white. The sample characteristics were similar in the derivation and validation datasets (Supplementary Materials, Table S-1).

Agreement of 3D-CAM-S and CAM-S SF overall scores

In the derivation dataset, the 3D-CAM-S scores were equal to the CAM-SF scores for 94% of the assessments with a weighted kappa of 0.94 (95% CI 0.93-0.95) (Supplementary Materials, Table S-2). All but 1 assessment of the 3D-CAM-S scores (1368/1369 \approx 100%)

were within one point of the CAM-S SF score. When overall scores were categorized as none (0), mild (1), moderate (2) or severe (3-7), 97% were concordant (weighted kappa 0.97 [95% CI 0.96-0.98]).

In the independent validation dataset, the agreement between the 3D-CAM-S and CAM-S overall scores was also very high; the 3D-CAM-S scores were equal to the CAM-S SF scores for 93% of the assessments with a weighted kappa of 0.93 (95% CI 0.92-0.95) (Table 1). One hundred percent (692/692) of the 3D-CAM derived CAM-S scores were within one point of the CAM-S SF scores. When the overall scores were categorized into none, mild, moderate, and severe as described above, 96% were concordant (weighted kappa 0.96 [95% CI 0.94-0.97]).

Agreement of 3D-CAM-S and CAM-S SF feature scores

The 3D-CAM-S Features 1 and 4 (Acute Change or Fluctuating Course and Altered Level of Consciousness, respectively) mapped directly onto CAM-S SF, classifying 100% of the severity ratings correctly. The 3D-CAM-S algorithm for Feature 2 (Inattention) correctly classified 94% of the CAM-S severity ratings in the validation set (N=692) and the 2 measures had a weighted kappa of 0.89 (95% CI 0.86-0.92) indicating high agreement. The 3D-CAM-S algorithm for Feature 3 (Disorganized Thinking) correctly classified 99% of the ratings with a weighted kappa of 0.90 (95% CI 0.85-0.95). For more details about misclassifications of the 3D-CAM-S relative to the CAM-S SF, see Supplemental Materials.

3D-CAM-S and CAM-S SF distributions

Supplementary Materials Figure S-1 and Figure 2 demonstrate that the distributions of the 3D-CAM-S and CAM-S SF in the derivation and validation datasets (respectively) were very similar. Both measures show excellent separation based on delirium status, with delirious patients having higher scores (4-7), and non-delirious patients having lower scores (0-2). Overall scores of 3 showed a mix of patients with and without delirium for both 3D-CAM-S and CAM-S SF (those with an overall score of 3 without delirium likely had subsyndromal delirium).

Effect of clustering assessments within patients

When we examined the effect of clustering of assessments within patients in the validation dataset, this yielded a patient-specific kappa of 0.92 (95% CI: 0.87-0.96), which was very similar to the weighted kappa reported above considering assessments independently.

3D-CAM-S and clinical outcomes

The 3D-CAM-S was significantly associated with hospital LOS, institutional discharge, and 30 day readmission, with an increasing trend in prevalence for each outcome with increasing delirium severity (see Supplemental Materials for detailed results).

DISCUSSION

We report development of a new, simple method for scoring delirium severity using the 3D-CAM¹⁰, the 3D-CAM-S. We observed a very high agreement between the 3D-CAM-S and

previously validated CAM-S SF, both in the sample used to derive the method and in an independent validation dataset. This new method enhances the 3D-CAM, a standardized, 3-minute interview for CAM-defined delirium, by providing a severity measure in addition to the previously validated determination of delirium presence or absence. The value of the current paper is demonstrating that delirium severity can be validly measured after a brief, structured 3-minute interview (the 3D-CAM). This will enable clinicians and researchers to use the 3D-CAM both for delirium identification and to monitor delirium severity status over time, such as for assessing response to treatment.

The CAM and CAM-S provide both delirium diagnosis and measure severity, and the 3D-CAM and 3D-CAM-S now also provide this advantage. Other instruments that provide information for both delirium diagnosis and severity measurement require substantial completion time (e.g., Delirium Rating Scale [DRS]^{18,19}, Delirium Symptom Interview [DSI]²⁰), require a clinician rater (e.g., DRS), and/or use a cutpoint for delirium diagnosis from a continuously scored severity measure, which results in both false positives and false negatives (e.g., DRS¹⁹ and MDAS¹⁴). The 3D-CAM/3D-CAM-S severity score takes an average of 3 minutes to administer¹¹, can be reliably conducted by non-clinically trained research assistants in addition to non-expert clinicians, and uses a CAM-based diagnostic approach rather than a cutpoint, improving diagnostic specificity.

Our study had a number of strengths. First, we derived and validated the 3D-CAM-S scoring method in a well-characterized older surgical population with rigorous delirium assessments and few missing values. Second, similar to the 3D-CAM diagnostic instrument, the 3D-CAM-S is based on objective scoring and is relatively non-interviewer dependent, resulting in greater accessibility to healthcare professionals and requiring less training. Third, we were able to generate very high rates of agreement between the 3D-CAM-S and CAM-S SF using our easily operationalized algorithm. Fourth, this paper operationalizes delirium severity using the 3D-CAM, a brief 3 minute interview. Other delirium severity measures, such as the MDAS and DRS-98, require a longer assessment. Several studies have demonstrated that delirium severity confers additional prognostic information beyond delirium diagnosis^{6,21}. Severity measures are also useful for monitoring the status of patients with established delirium over time. In the Supplemental Materials, we show that the 3D-CAM-S is also predictive of hospital LOS, institutional discharge, and 30-day readmission. Fifth, if 3D-CAM assessments have been collected previously, the scoring algorithm presented in our current study enables scoring of delirium severity retrospectively, which is not possible using current instruments that require interviewer severity ratings. Finally, our use of a split sample approach showing that an independent validation dataset yielded similar results to the initial derivation dataset underscores the robustness of our findings.

We also note some study limitations. First, the restricted enrollment based on age < 70, absence of dementia, and patients undergoing scheduled major elective surgery could potentially limit the generalizability of our findings. The 3D-CAM-S will require additional validation in medical patients and in patients with dementia. Second, repeated assessments from the same patients were considered independent. However, sensitivity analyses that accounted for clustering within patients indicated nearly identical results. Third, we used a split sample approach for independent validation, examining 3D-CAM-S in a cohort

separate from the one in which it was derived. However, we acknowledge that it would strengthen this work to validate our methods in an entirely different cohort, which will be done in future work. Fourth, in this study we used the CAM-S to validate the 3D-CAM-S. In future studies, the 3D-CAM-S should be examined against other severity measures, such as the MDAS or DRS-98. Finally, the 3D-CAM was derived from a subset of items administered as part of the SAGES CAM-based delirium assessment. Future studies in which the 3D-CAM is administered independently will be beneficial in confirming our results.

In conclusion, the new 3D-CAM-S delirium severity score provides a new dimension to the 3D-CAM assessment with minimal additional effort. The 3D-CAM-S score is derived from objective measures within the 3D-CAM and is highly correlated with the CAM-S SF score. The ability to generate a delirium severity score from the brief 3D-CAM may prove useful in clinical and research settings, and may be incorporated into studies aimed to refine delirium prognosis, improve delirium treatment, and improve outcomes of older patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute on Aging T32AG023480 (Vasunilashorn), K01AG033643 (Saczynski), P01AG031720 (Inouye), R01AG044518 (Inouye/Jones), K07AG041835 (Inouye), R01AG030618 (Marcantonio), K24AG035075 (Marcantonio); the National Heart Lung and Blood Institute U01HL105268 (Saczynski); the National Institute of Nursing Research R01NR011042 (Fick); and the Charles A. King Trust Postdoctoral Research Fellowship Program, Bank of America, N.A., Co-Trustee (Vasunilashorn). Dr. Inouye holds the Milton and Shirley F. Levy Family Chair. The authors gratefully acknowledge the contributions of the patients, family members, nurses, physicians, staff members, and members of the Executive Committee who participated in the Successful Aging after Elective Surgery (SAGES) Study.

REFERENCES

1. Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. *Lancet*. 2014; 383:911–922. [PubMed: 23992774]
2. Marcantonio ER. Postoperative delirium: A 76-year-old woman with delirium following surgery. *JAMA*. 2012; 308:73–81. [PubMed: 22669559]
3. Marcantonio ER, Goldman L, Mangione CM, et al. A clinical prediction rule for delirium after elective noncardiac surgery. *JAMA*. 1994; 271:134–139. [PubMed: 8264068]
4. Eubank KJ, Covinsky K. Delirium severity in hospitalized patient: Time to pay attention. *Ann Intern Med*. 2014; 160:574–575. [PubMed: 24733202]
5. Inouye SK, van Dyck CH, Alessi CA, et al. Clarifying confusion: The confusion assessment method. A new method for detection of delirium. *Ann Intern Med*. 1990; 113:941–948. [PubMed: 2240918]
6. Inouye SK, Kosar CM, Tommet D, et al. The CAM-S, a new scoring system for delirium severity: Association with clinical outcomes. *Ann Intern Med*. 2014; 150:526–533. [PubMed: 24733193]
7. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc*. 1975; 23:433–441. [PubMed: 1159263]
8. Robinson TN, Raeburn CD, Tran ZV, et al. Postoperative delirium in the elderly. Risk factors and outcomes. *Ann Surg*. 2009; 249:173–178. [PubMed: 19106695]

9. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005; 53:695–699. [PubMed: 15817019]
10. Harvey RJ, Hammer AL. Item response theory. *Couns Psychol.* 1999; 27:353–383.
11. Marcantonio ER, Ngo LH, O'Connor M, et al. 3D-CAM: Derivation and validation of a 3-minute diagnostic assessment for delirium. *Ann Intern Med.* 2014; 161:554–561. [PubMed: 25329203]
12. Yang FM, Jones RN, Inouye SK, et al. Selecting optimal screening items for delirium: An application of item response theory. *BMC Med Res Methodol.* 2013; 13:8. [PubMed: 23339752]
13. Breitbart W, Rosenfeld B, Roth A, et al. The memorial delirium assessment scale. *J Pain Symptom Manage.* 1997; 13:128–137. [PubMed: 9114631]
14. Shyamsundar G, Raghuthanman G, Rajkumar AP, et al. Validation of memorial delirium assessment scale. *J Crit Care.* 2009; 24:530–534. [PubMed: 19327962]
15. Schmitt EM, Marcantonio ER, Alsup DC, et al. Novel risk markers and long-term outcomes of delirium: The Successful Aging after Elective Surgery (SAGES) Study Design and Methods. *JAMDA.* 2012; 13:1–10. [PubMed: 22119549]
16. Schmitt E, Saczynski J, Kosar C, et al. The Successful Aging after Elective Surgery (SAGES) Study: Cohort description and data quality procedures. *J Am Geriatr Soc.* 2015; 63:2463–2471. [PubMed: 26662213]
17. Torsten H, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A conditional inference framework. *J Computer Graph Stat.* 2006; 15:651–674.
18. Trzepacz PT, Baker RW, Greenhouse J. A symptom rating scale for delirium. *Psychiatry Res.* 1988; 23:89–97. [PubMed: 3363018]
19. Trzepacz PT, Mittal D, Torres R, et al. Validation of the Delirium Rating Scale-revised-98: Comparison with the delirium rating scale and the cognitive test for delirium. *J Neuropsych Clin Neurosci.* 2001; 13:229–242.
20. Albert MS, Levkoff SE, Reilly C, et al. The delirium symptom interview: An interview for the detection of delirium symptoms in hospitalized patients. *J Geriatr Psychiatry Neurol.* 1992; 5:14–21. [PubMed: 1571069]
21. Marcantonio ER, Ta T, Duthie E, et al. Delirium severity and psychomotor types: Their relationship with outcomes after hip fracture repair. *J Am Geriatr Soc.* 2002; 50:850–857. [PubMed: 12028171]

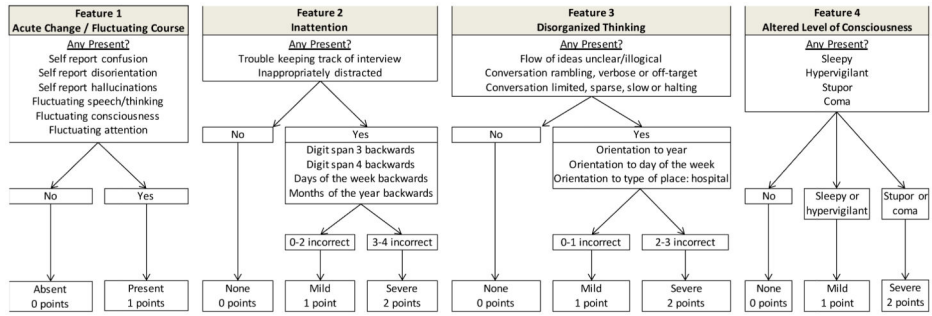


Figure 1. Scoring procedure for creating a severity score from the 3D-CAM based on the CAM-S SF

Abbreviations: 3D-CAM=3-Minute Diagnostic Interview for CAM-defined Delirium;
CAM-S SF=Confusion Assessment Method-Severity Short Form

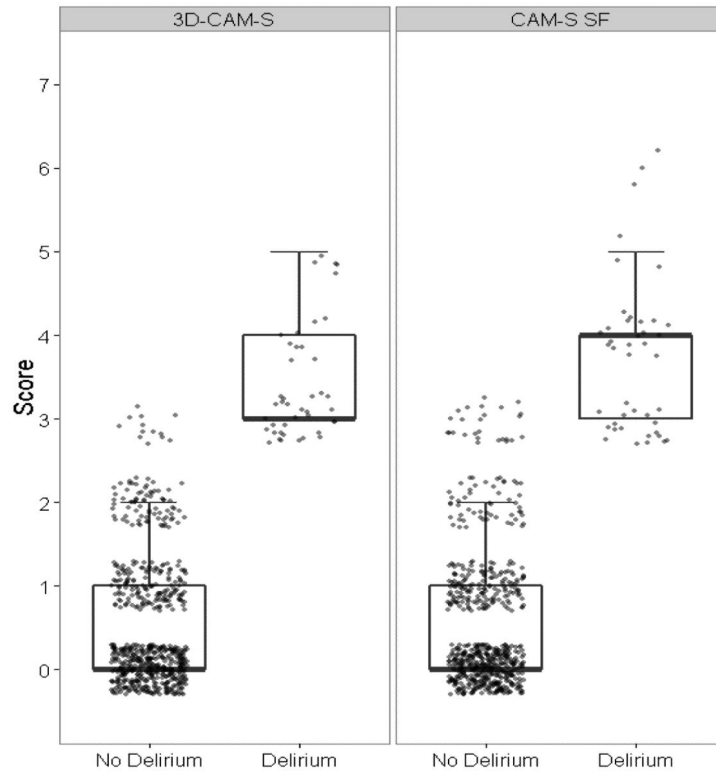


Figure 2. Distribution of the 3D-CAM derived CAM-S (the 3D-CAM-S) and CAM-S short form (SF) delirium severity scores in the validation dataset

(N=692 assessments, 43 with delirium)

Abbreviations: 3D-CAM=3-Minute Diagnostic Interview of CAM-defined Delirium; CAM-S = Confusion Assessment Method-Severity

Boxplots show the median, 25th and 75th percentiles. Whiskers extend to the highest value within 1.5 times the interquartile range. Note that in all of these, the lower whisker overlaps with the 25th percentile, and the median overlaps with the 25th or 75th percentiles.

3D-CAM derived CAM-S (the 3D-CAM-S) compared to the CAM-S short form scores for assessments in the validation dataset (N=692 assessments)

Table 1

	CAM-S Short Form Score	3D-CAM-S Score							Total	
		None	Mild	Moderate	Severe					
		0	1	2	3	4	5	6	7	
None	0	403	0	0	0	0	0	0	0	403
Mild	1	0	154	12	0	0	0	0	0	166
Moderate	2	0	2	51	1	0	0	0	0	54
Severe	3	0	0	14	31	0	0	0	0	45
	4	0	0	0	10	6	2	0	0	18
	5	0	0	0	0	3	0	0	0	3
	6	0	0	0	0	0	3	0	0	3
	7	0	0	0	0	0	0	0	0	0
Total		403	156	77	42	9	5	0	0	692

Weighted kappa for 8-level severity score = 0.93 (0.95% CI: 0.92-0.95)

Weighted kappa for 4-level severity score = 0.96 (0.95% CI: 0.94-0.97)

Weighted kappa for 8-level severity score, accounting for clustering at the patient level= 0.92 (95% CI: 0.87-0.96),

Abbreviations: 3D-CAM=3-Minute Diagnostic Interview of CAM-defined Delirium; CAM-S = Confusion Assessment Method-Severity