# Quantifying the Severity of a Delirium Episode Throughout Hospitalization: the Combined Importance of Intensity and Duration

Sarinnapha M. Vasunilashorn, PhD[1,2], Edward R. Marcantonio, MD, SM[1,2,3], Yun Gou, MA[3], Margaret A. Pisani, MD, MPH[4], Thomas G. Travison, PhD[1,2,3], Eva M. Schmitt, PhD[3], Richard N. Jones, ScD[5], and Sharon K. Inouye, MD, MPH[1,2,3]

[1]Beth Israel Deaconess Medical Center, Boston, MA, USA; [2]Harvard Medical School, Boston, MA, USA; [3]Hebrew Senior Life, Boston, MA, USA; [4]Yale School of Medicine, New Haven, CT, USA; [5]Warren Alpert Medical School of Brown University, Providence, RI, USA.

**BACKGROUND:** The ability to determine which episodes of delirium are likely to lead to poor clinical outcomes has remained a major area of challenge.

**OBJECTIVE:** To quantify delirium severity and course over an entire hospitalization using several measures, and to evaluate their predictive validity for 30- and 90-day outcomes post-discharge.

**DESIGN:** Two prospective cohort studies.

**PARTICIPANTS:** Analysis was conducted in two independent cohorts of adult patients aged ≥70.

**MAIN MEASURES:** Nine delirium episode severity measures were examined: (1) measures reflecting delirium intensity (peak Confusion Assessment Method-Severity [CAM-S] and mean CAM-S score), (2) a measure reflecting delirium intensity and duration (sum of all CAM-S scores, sum of all CAM-S scores on delirium days only, peak CAM-S score x days with delirium), (3) measures requiring information on delirium duration and delirium at discharge (total number of delirium days, percentage of delirium days, delirium at discharge), and (4) a measure of cognitive change. Associations of the delirium episode severity measures with 30- and 90-day post-hospital outcomes (death, nursing home placement, and readmission) relevant to delirium were examined.

**KEY RESULTS:** The delirium episode severity measure that required information on both delirium intensity and duration (sum of all CAM-S scores) was the most strongly associated with 30- and 90-day post-hospital outcomes. Using this measure, the relative risk [95 % confidence interval] for death at 30-days increased across levels of sum of all CAM-S scores from 1.0 (referent) to 2.1 [0.8, 5.4] for 'low,' to 2.9 [1.2, 7.1] for 'moderate,' to 6.4 [2.9, 14.0] for 'high' (p for trend <.01).

**CONCLUSIONS:** The delirium episode severity measure that included both intensity and duration had the strongest association with important post-hospital outcomes.

KEY WORDS: delirium severity; intensity; duration; cognitive change.

## INTRODUCTION

Delirium, an acute confusional state, is a common, morbid, and costly geriatric syndrome.[1] It occurs in 30–50 % of medical patients and 27–51 % of orthopedic surgical patients, with the highest levels observed among patients in the intensive care unit[1]. Delirium after major surgery is associated with longer length of hospitalization,[2] greater nosocomial complications,[3] and higher rates of discharge to nursing homes.[4,5] Moreover, patients with post-surgical delirium have an increased risk of cognitive and functional decline, incident dementia, and a 2–20 fold increased risk of mortality,[2,6] with hospital mortality rates of 25–33 %.[5,7] Taken together, the estimated annual health care costs attributable to delirium ranges upwards of $182 billion U.S. dollars.[8]

The ability to determine which episodes of delirium are likely to lead to poor clinical outcomes has remained a major area of challenge. To date, clinicians and investigators have characterized delirium by its intensity at a single timepoint[9–14] or its total duration[4] in an effort to quantify "severity." However, previous work has not quantified the severity of an *entire episode of delirium* (such as across an entire hospitalization), which is critical for advancing our understanding of the influence of delirium on clinical outcomes. Such an exploration would require expansion of the concept of delirium severity beyond a single instantaneous rating or counting days with delirium.

Thus, the present study was designed to better quantify the severity of an entire episode of delirium during hospitalization across multiple (and combined) domains, including delirium intensity, delirium duration, or change in cognitive function over time. We examined the predictive validity of multiple delirium episode severity measures for 30- and 90-day post-hospital outcomes in two distinct cohorts of older adults. Thus, the present study was designed to better quantify the severity

of an entire episode of delirium during hospitalization across multiple (and combined) domains, including delirium intensity, delirium duration, or change in cognitive function over time. Our aim was to identify the 'best' measures of delirium episode severity by examining the predictive validity of multiple delirium episode severity measures for 30- and 90-day post-hospital outcomes in two distinct cohorts of older adults. We hypothesized that a combination measure of delirium episode severity that incorporated both intensity and duration would improve prediction of adverse post-hospital outcomes relative to any single instantaneous delirium severity measure.

## METHODS

### Study Samples

Two prospective cohort studies (surgical and medical) were examined: the Successful AGing after Elective Surgery (SAGES) Study and Project Recovery (respectively). The two samples were not combined because of baseline differences and in order to provide cross-validation for predictive validity. The ongoing SAGES Study enrolled 566 patients aged ≥70 years scheduled for elective major noncardiac surgery, including: cervical or lumbar laminectomy, total hip or knee replacement, open abdominal aortic aneurysm repair, lower extremity vascular bypass, or colectomy. Participants were recruited during June 2010 to August 2013 from two Harvard Medical School affiliated hospitals. Major inclusion and exclusion criteria have been previously published.[15] Briefly, 1052 patients were screened for eligibility based on medical record review (age ≥70 years; scheduled for major orthopedic, vascular, or general surgical procedures; having general or regional anesthesia; projected hospital stay ≥2 days). Among these patients, 318 declined an interview, 163 were ineligible (residing within 50 miles, evidence of dementia, rescheduled surgery, prior hospitalization, severe blindness or deafness), and 5 were eligible but refused participation, resulting in a final sample of 566 patients (eFigure 1).[16]

Project Recovery enrolled patients aged ≥70 years, who were admitted to the medicine service at Yale-New Haven Hospital from March 1995 to March 1998. Sample inclusion and exclusion criteria have been reported.[8,17,18] Briefly, 2434 patients were eligible based on: age ≥70 years and no delirium upon admission, but with medium to high risk for delirium. Among these patients, 1265 were excluded given their inability to complete interviews ($n = 298$), coma or terminal illness ($n = 69$), hospital stay <48 h ($n = 219$), previous enrollment ($n = 324$), or other reasons ($n = 355$). Lastly, 250 patients declined enrollment, yielding a final sample of 919 patients (eFigure 2). This reflects a combined cohort, which has been previously examined.[10,18,19]

Clinical outcomes were assessed by trained full time clinical research associates (e.g., chart abstractors and interviewers) who were blinded to the Confusion Assessment Method (CAM; described below) delirium status of the patients. A separate group of trained full-time clinical research associates conducted in-hospital CAM assessments. For SAGES, informed consent for study participation was obtained from all subjects according to procedures approved by the institutional review boards of Beth Israel Deaconess Medical Center and Brigham and Women's Hospital, the two surgical sites, and Hebrew SeniorLife, the study coordinating center, all located in Boston, MA. For Project Recovery, informed consent was obtained from the patients or a proxy for patients with significant cognitive impairment, as approved by the Yale University Institutional Review Board.

### Delirium Assessment

For both studies, delirium was determined using the CAM,[20] which was assessed daily throughout hospitalization. The CAM was scored after a 10–15 min structured interview with cognitive testing (of attention, orientation, memory) conducted by trained study staff. Delirium was present if the patient had an acute onset of change or fluctuation in mental status, inattention, and either disorganized thinking or altered level of consciousness. Study interviewers underwent intensive training and standardization.[16,21]

### Delirium Episode Severity Measures

Individual Measures: We chose to examine different types of measures based on evidence in the literature. For example, previous studies have examined delirium intensity, persistence or duration as severity indicators.[1,4] Other studies examined delirium at discharge and cognitive change as severity indicators.[22,23] Thus, we chose to examine nine individual measures of delirium episode severity that were classified into four groups of measures requiring: 1) delirium intensity only (based on the 10-item CAM-S long form measurement), 2) delirium intensity and duration, 3) measures requiring information on delirium duration and delirium at discharge, and 4) a measure of cognitive change. We considered two measures that require delirium intensity (peak CAM-S score and mean CAM-S score), along with three measures that require delirium intensity plus delirium duration (sum of all CAM-S scores; sum of all CAM-S scores, only on delirium days; and peak CAM-S score × delirium days). Peak CAM-S × delirium days was considered since it explicitly considers both the intensity and duration of delirium, whereas the sum of all CAM-S scores implicitly considers intensity and duration without including number of delirium days in its definition. Three measures required delirium diagnosis (without intensity) and delirium duration: total number of delirium days, percentage of delirium days, and delirium at discharge. The final delirium episode severity measure captured change in cognition. This was computed, in Project Recovery only, as the absolute difference between the highest and lowest Mini Mental State Examination score (purchased from Psychological Assessment Resources). For our analyses, we considered the entire episode of delirium for each patient as our unit of analysis;

thus, we did not treat each delirium day as an independent time point for these analyses. eTable 1 contains more information on these delirium episode severity measures.

Combination Measures: Two combination measures were considered to see whether predictive validity could be improved by considering several individual measures simultaneously (see Online-only Supplement for details).

## Association with Clinical Outcomes

We examined the associations among the nine delirium episode severity measures and two combination measures with relevant 30- and 90-day post-hospital outcomes in Project Recovery, where collection of clinical outcomes has been completed. These 30- and 90-day outcomes included: death, nursing home placement, and readmission. Information on death was obtained from medical records, the National Death Index, Social Security and Medicare Part A databases, and death certificates.[8,24] Nursing home placement was determined from Medicare Part A information. Hospital readmission was determined from Medicare Part A and medical record review. Since participants who die are no longer eligible for other outcomes, two hierarchical outcomes were considered to avoid inferential errors: 1) nursing home placement or death, and 2) readmission or death. We had complete information (i.e., no missing data) for our outcome measures.

Since SAGES data collection is ongoing, we investigated two currently available 30-day outcomes: nursing home placement and readmission. Only one patient died within 30-days post-discharge, and was included in these outcome groups for all analyses.

## Statistical Analysis

We conducted our analysis in two stages. First, we examined the association between each of the nine delirium episode severity measures with 30- and 90-day clinical outcomes. To aid in interpretation of the post-hospital outcomes, we categorized all delirium episode severity measures, except for delirium at discharge, into four categories defined by sample distribution quartiles and labeled: none, low, moderate, high. Patients without delirium were included in the analyses, and may have contributed severity points (i.e., not all cases had 0 points). Delirium at discharge was classified as: no delirium, delirium not at discharge, and delirium at discharge. The association of each delirium episode severity measure and each outcome was estimated using a generalized linear model with a Poisson error distribution for binary outcomes (death, nursing home residence, and readmission). Since our goal was to develop a relative ranking to compare delirium severity measures, and not to develop a predictive model for clinical outcomes following delirium, we wanted to have parsimonious adjustment, only for age, sex, and race in all models; thus, we did not control for every possible contributor to the outcomes (e.g., Charlson comorbidity index).

Two types of c-statistics were considered to examine model fit for the categorical and continuous measures of delirium episode severity. The c-statistic is widely used as a measure of predictive accuracy. It indicates how well predicted values agree with observed values and also how well the model discriminates between observations at different levels of an outcome. Values for the c-statistic can be thought of as the probability that among two people chosen at random from the sample, one with a lower value on the risk factor (delirium episode severity) and the other with a higher value of the risk factor, that the person with the higher value on the risk factor has a greater value on the outcome variable. We did not anticipate high values for our c-statistics since we did not develop explanatory models intended to maximize predictive ability for the clinical outcomes; instead we used the c-statistic as a relative metric to compare the nine delirium episode severity measures to each other. Since Project Recovery was an intervention trial, we conducted additional analyses to evaluate the potential influence of the intervention on these relationships.

## Additional Analyses

Additional analyses aimed to explore the importance of delirium intensity versus duration. We conducted a stratified analysis of the delirium episode severity measure 'peak CAM-S x delirium days' to determine whether a severe (intense) delirium over a single day was associated with poorer clinical outcomes compared to a less severe (intense) delirium over ≥2 days (see Online-Only Supplement for details).

Analyses were conducted using SAS version 9.3 (SAS Institute, Cary, NC) and Stata version 13 (StatCorp, College Station, TX).

## RESULTS

Sample characteristics of the patients are reported in Table 1. The Project Recovery and SAGES samples included older adults (mean age, 80 and 77, respectively) with fewer men than women (40 and 42 % men, respectively). A greater proportion of Project Recovery patients had a Charlson comorbidity index of ≥2 relative to SAGES patients (70 vs. 30 %, respectively). Incident delirium during hospitalization was higher in SAGES than Project Recovery patients (24 vs. 13 %, respectively).

In Table 2, we report the means and proportions of the nine delirium episode severity measures. The average levels in Project Recovery and SAGES were similar across all measures.

In Table 3, we report the association of the nine delirium episode severity measures with 30-day outcomes in Project Recovery, including death, nursing home residence, and readmission. The risk of death increased across each severity category for: peak CAM-S, mean CAM-S, sum of all CAM-S scores, and delirium at discharge. An increased risk for

**Table 1 Sample Characteristics of the Project Recovery and SAGES Studies**

| Characteristics | Project recovery | SAGES study |
| --- | --- | --- |
| | Medical population | Surgical population |
| | (*N* = 919) | (*N* = 566) |
| Age at baseline- mean years ± SD | 80.0 ± 6.5 | 76.7 ± 5.2 |
| Male—*n* (%) | 365 (40) | 236 (42) |
| Non-white race—*n* (%) | 119 (13) | 43 (8) |
| Married—*n* (%) | 332 (36) | 335 (59) |
| Education—mean years ± SD | 11.1 ± 3.5 | 15.0 ± 2.9 |
| Charlson comorbidity index—*n* (%) | | |
| 0 | 94 (10) | 260 (46) |
| 1 | 182 (20) | 139 (25) |
| 2+ | 643 (70) | 167 (29) |
| Living alone—*n* (%) | 371 (40) | 169 (30) |
| Residing in nursing home—*n* (%) | 56 (6) | 0 (0) |
| Delirium during hospitalization—*n* (%) | 115 (13) | 135 (24) |

*SAGES Successful AGing after Elective Surgery*

hospital readmission by delirium episode severity category was observed for: sum of all CAM-S scores, peak CAM-S,

**Table 2 Means and Proportions of Nine Delirium Episode Severity Measures**

| Measure | Mean (range) or n/N (%) | | | |
| --- | --- | --- | --- | --- |
| | Project sample (*N* = 919) | | SAGES study sample (*N* = 566) | |
| | Total sample | Patients with delirium | Total sample | Patients with delirium |
| *Measures requiring delirium intensity and duration* | | | | |
| 1. Sum of all CAM-S scores (points) | 10.7 (0.0, 122.0) | 34.1 (6.0, 122.0) | 9.3 (0.0, 113.0) | 22.5 (3.0, 113.0) |
| 2. Sum of all CAM-S scores, only on delirium days (points) | 2.1 (0.0, 110.0) | 17.2 (2.0 110.0) | 3.3 (0.0, 113.0) | 15.5 (0.0, 113.0) |
| 3. Peak score × days delirious (points) | 2.8 (0.0, 133.0) | 22.2 (4.0, 133.0) | 5.3 (0.0, 494.0) | 22.4 (0.0, 513.0) |
| *Measures requiring delirium intensity* | | | | |
| 4. Peak CAM-S score (points) | 3.3 (0.0, 19.0) | 8.8 (14.0, 19.0) | 4.0 (0.0, 19.0) | 8.1 (1.0, 19.0) |
| 5. Mean CAM-S score (points) | 1.7 (0.0, 13.6) | 4.1 (0.6, 13.6) | 2.3 (0.0, 12.6) | 4.1 (0.6, 12.6) |
| *Measures requiring information on delirium duration and delirium at discharge* | | | | |
| 6. Total number of delirium days (days) | 0.3 (0.0, 9.0) | 2.1 (1.0, 9.0) | 0.5 (0.0, 26.0) | 3.2 (0.0, 27.0) |
| 7. Percentage of delirium days (%) | 2.4 (0.0, 54.5) | 18.9 (1.9, 54.6) | 6.7 (0.0, 100.0) | 30.0 (0.0, 100.0) |
| 8. Delirium at discharge (%) | 31/919 (3.4 %) | 36/115 (31) | 23/566 (4 %) | 23/135 (17) |
| *Other* | | | | |
| 9. Change in cognition (points) | 4.1 (0.0, 22.0) | 7.1 (0.0, 22.0) | n/a[s] | n/a* |

*For a description of how each measure was computed, see eTable 1*
*SAGES Successful AGing After Elective Surgery*
*CAM-S Confusion Assessment Method Severity Score (range: 0–19, 19 worst)*
*\*Mini Mental State Examination not available in SAGES*

mean CAM-S, and delirium at discharge. For every measure, the intervention did not significantly influence the relationship between delirium severity and the outcomes; thus, this variable was not included in our analyses. As anticipated, the strength of associations between the delirium episode severity measures and the 90-day outcomes (eTable 2) were less pronounced than the 30-day outcomes (Table 3).

The SAGES study had generally similar results at 30 days (Table 4) as Project Recovery (Table 3). For instance, the risk of readmission increased across each severity category for: peak CAM-S, mean CAM-S, and sum of all CAM-S scores.

Table 5 lists the components of the: 1) full combination measure that includes four delirium episode severity measures (sum of all CAM-S scores, delirium at discharge, total number of delirium days, and change in cognition), and 2) alternate combination measure, which is similar to the full combination measure except for the absence of sum of all CAM-S scores. Sum of all CAM-S scores yielded the highest c-statistic across all 30-day outcomes, with one exception (nursing home residence in SAGES). Since peak and mean CAM-S were highly correlated with the sum of all CAM-S scores ($r_s = 0.8$–$1.0$; eTable 3), they were not included in the combination measures. None of the combination measures improved predictive validity compared to the individual measures (eTable 4).

## DISCUSSION

Our findings across every delirium episode severity measure indicate a "dose–response" relationship between delirium severity and clinical outcomes; thus, more severe delirium is worse than mild delirium across every measure. In particular, episodes of delirium with the highest intensity and longest duration combined lead to the worst clinical outcomes (death, nursing home placement, or readmission). The ability to quantify severity of an entire episode of delirium rather than severity at a single point in time allows us to better capture the full clinical impact of delirium. Although all nine delirium episode severity measures (based on ratings of delirium intensity, duration, or cognitive change) were predictive of 30- and 90-day outcomes, the measure that added the delirium intensity scores by CAM-S measured each day over the entire hospitalization (sum of all CAM-S scores) demonstrated the strongest association with these outcomes. It is important to note, however, that the differences between the sum of all CAM-S and other CAM-S measures (peak CAM-S, mean CAM-S) were relatively small. In situations where a delirium intensity measure is unavailable, a combination of other measures (e.g., delirium at discharge, delirium days, and change in cognition—included in the alternate combination score) represents the next best delirium episode severity measure. Moreover, our results suggest that considering both delirium intensity and duration are important to fully understanding the downstream effects of an entire episode of delirium. These findings hold important implications for clinicians and researchers interested

**Table 3 Association of Nine Delirium Episode Severity Measures with Three 30-Day Outcomes (Death, Nursing Home Residence, Readmission) in Project Recovery**

| Delirium severity | Death (N=919) | | Nursing home Residence or death (N=844) | | Readmission or death (N=844) | |
|---|---|---|---|---|---|---|
| Categories | n/N (%) | RR | n/N (%) | RR | n/N (%) | RR |
| **1. Sum of all CAM-S scores** | | | | | | |
| None 0–3 (299, 32 %) | 7/299 (2) | Ref | 24/266 (9) | Ref | 41/266 (15) | Ref |
| Low 4–6 (201, 22 %) | 10/201 (5) | 2.1 | 21/179 (12) | 1.3 | 27/179 (15) | 1.0 |
| Moderate 7–13 (191, 21 %) | 13/191 (7) | 2.9* | 45/182 (25) | 2.6* | 39/182 (21) | 1.4 |
| High 14–122 (228, 25 %) | 34/228 (15) | 6.4* | 100/217 (46) | 4.7* | 67/217 (31) | 2.0* |
| C-statistics | 0.70 (0.72) | | 0.73 (0.74) | | 0.60 (0.61) | |
| P for trend | <.01 | | <.01 | | <.01 | |
| **2. Sum of all CAM-S scores, only on delirium days** | | | | | | |
| None 0 (804, 87 %) | 39/804 (5) | Ref | 136/735 (19) | Ref | 132/735 (18) | Ref |
| Low 2–6 (36, 4 %) | 4/36 (11) | 2.4 | 11/33 (33) | 1.7* | 12/33 (36) | 2.1* |
| Moderate 7–15 (40, 4 %) | 8/40 (20) | 4.2* | 16/38 (42) | 2.1* | 13/38 (34) | 2.0* |
| High 16–110 (39, 4 %) | 13/39 (33) | 6.9* | 27/38 (71) | 3.8* | 17/38 (45) | 2.5 |
| C-statistics | 0.65 (0.65) | | 0.60 (0.60) | | 0.57 (0.57) | |
| P for trend | <.01 | | <.01 | | <.01 | |
| **3. Peak score × days delirium** | | | | | | |
| None 0 (804, 87 %) | 39/804 (5) | Ref | 136/735 (9) | Ref | 132/735 (18) | Ref |
| Low 4–7 (40, 4 %) | 5/40 (13) | 2.7* | 15/37 (41) | 2.1* | 13/37 (35) | 2.1* |
| Moderate 8–18 (38, 4 %) | 8/38 (21) | 4.3* | 14/36 (39) | 1.9* | 12/36 (33) | 1.9* |
| High 19–133 (37, 4 %) | 12/37 (32) | 6.8* | 25/36 (69) | 3.7* | 17/36 (47) | 2.7* |
| C-statistics | 0.65 (0.65) | | 0.60 (0.60) | | 0.57 (0.57) | |
| P for trend | <.01 | | <.01 | | <.01 | |
| **4. Peak CAM-S score** | | | | | | |
| None 0–1 (205, 22 %) | 4/205 (2) | Ref | 16/181 (9) | Ref | 26/181 (14) | Ref |
| Low 2 (288, 31 %) | 11/288 (4) | 1.9 | 33/263 (13) | 1.4 | 45/263 (17) | 1.2 |
| Moderate 3–4 (234, 25 %) | 19/234 (8) | 4.1* | 58/217 (27) | 2.9* | 45/217 (21) | 1.5 |
| High 5–19 (192, 21 %) | 30/192 (16) | 8.3* | 83/183 (45) | 4.8* | 58/183 (32) | 2.3* |
| C-statistics | 0.70 (0.72) | | 0.71 (0.72) | | 0.60 (0.61) | |
| P for trend | <.01 | | <.01 | | <.01 | |
| **5. Mean CAM-S score** | | | | | | |
| None 0.00–0.75 (247, 27 %) | 7/247 (3) | Ref | 22/217 (10) | Ref | 33/217 (15) | Ref |
| Low 0.76–1.30 (234, 25 %) | 7/234 (3) | 1.1 | 24/213 (11) | 11 | 33/213 (15) | 1.0 |
| Moderate 1.40–2.30 (213, 23 %) | 15/213 (7) | 2.5* | 52/198 (26) | 2.5* | 43/198 (22) | 1.5* |
| High 2.40–13.60 (225, 24 %) | 35/225 (16) | 5.6* | 92/216 (43) | 3.9* | 65/216 (30) | 2.1* |
| C-statistics | 0.70 (0.72) | | 0.70 (0.73) | | 0.60 (0.61) | |
| P for trend | <.01 | | <.01 | | <.01 | |
| **6. Total number of delirium days** | | | | | | |
| None 0 (804, 87 %) | 39/804 (5) | Ref | 136/735 (19) | Ref | 132/735 (18) | Ref |
| Low 1 (63, 7 %) | 10/63 (16) | 3.3* | 22/59 (37) | 1.9* | 20/59 (34) | 2.0* |
| Moderate 2–3 (34, 4 %) | 10/34 (29) | 6.4* | 20/33 (61) | 3.1* | 16/33 (48) | 2.8* |
| High 4–9 (18, 2 %) | 5/18 (28) | 5.5* | 12/17 (71) | 3.3* | 6/17 (35) | 1.9* |
| C-statistics | 0.65 (0.65) | | 0.60 (0.60) | | 0.57 (0.57) | |
| P for trend | <.01 | | <.01 | | 0.02 | |
| **7. Percent delirium days** | | | | | | |
| None 0 (804, 87 %) | 39/804 (5) | Ref | 136/735 (19) | Ref | 132/735 (18) | Ref |
| Low 1–12 (36, 4 %) | 7/36 (19) | 4.3* | 14/34 (41) | 2.1* | 15/34 (44) | 2.6* |
| Moderate 13–22 (40, 4 %) | 7/40 (18) | 3.5* | 18/37 (49) | 2.3* | 13/37 (35) | 2.0* |
| High 23–55 (39, 4 %) | 11/39 (28) | 6.0* | 22/38 (58) | 3.1* | 14/38 (37) | 2.1* |
| C-statistics | 0.65 (0.65) | | 0.60 (0.60) | | 0.57 (0.57) | |
| P for trend | <.01 | | <.01 | | <.01 | |
| **8. Delirium at discharge** | | | | | | |
| No delirium (804, 87 %) | 39/804 (5) | Ref | 136/735 (19) | Ref | 132/735 (18) | Ref |
| Delirium, not at Discharge (84, 9 %) | 12/84 (14) | 2.7* | 32/78 (41) | 2.0* | 24/78 (31) | 1.7* |
| Delirium at discharge (31, 3 %) | 13/31 (42) | 8.4* | 22/31 (71) | 3.7* | 18/31 (58) | 3.2* |
| C-statistics | 0.65 (0.59) | | 0.60 (0.55) | | 0.57 (0.54) | |
| P for trend | <.01 | | <.01 | | <.01 | |
| **9. Change in cognition** | | | | | | |
| None 0–1 (175, 19 %) | 11/175 (6) | Ref | 23/160 (14) | Ref | 29/160 (18) | Ref |
| Low 2–3 (279, 30 %) | 14/279 (5) | 0.8 | 42/258 (16) | 1.0 | 47/258 (18) | 1.0 |
| Moderate 4–5 (232, 25 %) | 15/232 (6) | 1.0 | 51/205 (25) | 1.6* | 50/205 (24) | 1.4 |
| High 6–22 (233, 25 %) | 24/233 (10) | 1.6 | 74/221 (33) | 2.1* | 48/221 (22) | 1.2 |
| C-statistics | 0.57 (0.57) | | 0.62 (0.62) | | 0.53 (0.53) | |
| P for trend | 0.15 | | <.01 | | 0.19 | |

*RR relative risk, Ref referent group, C-statistic for continuous measure (for categorical measure)—used as a metric for comparison across measures; high values not expected (see text for details)*
*\*p < .05*

in the care of delirious patients, since addressing both intensity and duration of delirium may be necessary to impact on its poor prognosis.

The predictive validity of the CAM-S for hospital and post-hospital outcomes has been previously reported.[10] The present study extends this work substantially by: 1) considering the

**Table 4** Association of Eight Delirium Episode Severity Measures with Two 30-Day Outcomes (Nursing Home Residence and Readmission) in the Successful Aging after Elective Surgery (SAGES) Study

| Delirium severity<br><br>Measure<br><br>Categories | Nursing home<br><br>Residence<br>or death<br><br>(N = 566)<br><br>n/N (%) | RR | Readmission<br><br>or Death<br><br>(N = 566)<br><br>n/N (%) | RR |
|---|---|---|---|---|
| 1. Sum of all CAM-S scores | | | | |
| None 0–2 (112, 20 %) | 52/112 (46) | Ref | 4/112 (4) | Ref |
| Low 3–5 (168, 30 %) | 82/168 (49) | 1.0 | 16/167 (10) | 2.7 |
| Moderate 6–11 (147, 26 %) | 85/147 (58) | 1.1 | 22/145 (15) | 4.3* |
| High 12–113 (139, 24 %) | 104/139 (75) | 1.5* | 26/132 (20) | 5.5* |
| C-statistic | 0.62 (0.63) | | 0.65 (0.66) | |
| *P* for trend | <.01 | | <.01 | |
| 2. Sum of all CAM-S scores, only on delirium days | | | | |
| None 0 (451, 80 %) | 231/451 (51) | Ref | 42/445 (9) | Ref |
| Low 3–7 (41, 7 %) | 29/41 (71) | 1.2* | 10/41 (24) | 2.6* |
| Moderate 8–15 (35, 6 %) | 26/35 (74) | 1.4* | 8/35 (23) | 2.5* |
| High 16–113 (37, 7 %) | 35/37 (95) | 1.9* | 8/33 (24) | 2.5* |
| C-statistic | 0.60 (0.60) | | 0.61 (0.60) | |
| *P* for trend | <.01 | | 0.01 | |
| 3. Peak score × days delirium | | | | |
| None 0 (450, 80 %) | 230/450 (51) | Ref | 42/444 (9) | Ref |
| Low 1–6 (29, 5 %) | 19/29 (66) | 1.1 | 8/29 (28) | 2.9* |
| Moderate 7–10 (30, 5 %) | 22/30 (73) | 1.3* | 6/30 (20) | 2.2* |
| High 10–494 (57, 10 %) | 52/57 (91) | 1.8* | 12/53 (23) | 2.4* |
| C-Statistic | 0.60 (0.60) | | 0.60 (0.60) | |
| *P* for trend | <.01 | | 0.02 | |
| 4. Peak CAM-S score | | | | |
| None 0–1 (95, 17 %) | 45/95 (47) | Ref | 3/95 (3) | Ref |
| Low 2 (149, 26 %) | 70/149 (47) | 1.0 | 14/149 (9) | 3.0 |
| Moderate 3–4 (147, 26 %) | 78/147 (53) | 1.1 | 17/142 (12) | 3.8* |
| High 5–19 (175, 31 %) | 130/175 (74) | 1.5* | 34/170 (20) | 6.3* |
| C-statistic | 0.62 (0.63) | | 0.65 (0.66) | |
| *P* for trend | <.01 | | <.01 | |
| 5. Mean CAM-S score | | | | |
| None 0.0–1.0 (178, 31 %) | 80/178 (51) | Ref | 10/178 (6) | Ref |
| Low 1.1–1.7 (106, 19 %) | 54/106 (51) | 1.1 | 14/104 (13) | 2.4* |
| Moderate 1.8–3.0 (148, 26 %) | 81/148 (55) | 1.1 | 22/145 (15) | 2.7* |
| High 3.1–12.6 (134, 24 %) | 108/134 (81) | 1.7* | 22/129 (17) | 3.0* |
| C-statistic | 0.64 (0.64) | | 0.62 (0.62) | |
| *P* for trend | <.01 | | <.01 | |
| 6. Total number of delirium days | | | | |
| None 0 (450, 80 %) | 230/450 (51) | Ref | 42/444 (9) | Ref |
| Low 1 (64, 11 %) | 46/64 (72) | 1.3* | 16/64 (25) | 2.7* |
| Moderate 2 (31, 5 %) | 27/31 (87) | 1.8* | 6/30 (20) | 2.1* |
| High 3–26 (21, 4 %) | 20/21 (95) | 1.8* | 4/18 (22) | 2.3 |
| C-Statistic | 0.60 (0.60) | | 0.60 (0.60) | |
| *P* for trend | <.01 | | 0.12 | |
| 7. Percent delirium days | | | | |
| None 0–5 (450, 80 %) | 230/450 (51) | Ref | 42/444 (9) | Ref |
| Low 6–20 (40, 7 %) | 27/40 (68) | 1.2 | 9/40 (23) | 2.4* |
| Moderate 22–34 (38, 7 %) | 30/38 (79) | 1.5* | 12/38 (32) | 3.3* |
| High 38–87 (38, 7 %) | 36/38 (95) | 1.8* | 5/34 (15) | 1.5 |
| C-Statistic | 0.60 (0.60) | | 0.60 (0.60) | |
| *P* for trend | <.01 | | 0.24 | |
| 8. Delirium at discharge | | | | |
| No delirium (431, 76 %) | 215/431 (50) | Ref | 40/425 (9) | Ref |
| Delirium, not at discharge (112, 20 %) | 87/112 (78) | 2.1* | 24/110 (22) | 2.3* |
| Delirium at discharge (23, 4 %) | 21/23 (91) | 3.6* | 4/21 (19) | 2.0 |
| C-statistic | 010.61 (0.61) | | 0.60 (0.60) | |
| *P* for trend | <.01 | | 0.15 | |

*Death was not included separately since only 1 person died by 30 days post-operation*

*The Change in Cognition measure could not be computed since the Mini Mental State Exam is not available in SAGES*

*RR relative risk, Ref referent group, C-statistic for continuous measure (for categorical measure)—used as a metric for comparison across measures; high values not expected (see text for details)*

*\*p < .05*

CAM-S in multiple ways; 2) creating additional delirium episode severity measures that include delirium diagnosis, duration, and cognitive change; 3) examining combination measures of severity that take into account the best-performing individual delirium episode severity measures; and 4) investigating the association between these individual

**Table 5 C-statistic of Individual Delirium Episode Severity Measures with 30-Day Outcomes in Project Recovery and SAGES**

| Delirium episode severity measure | C-Statistic* | | | | | Avg. rank‡ |
|---|---|---|---|---|---|---|
| | Death† | Nursing home residence (or death†) | | Readmission (or death†) | | |
| | PR | PR | SAGES | PR | SAGES | |
| **Requires delirium intensity:** | | | | | | |
| Sum of all CAM-S | 0.72 | 0.74 | 0.63 | 0.61 | 0.66 | 1.1 |
| Peak CAM-S | 0.72 | 0.72 | 0.63 | 0.61 | 0.61 | 1.7 |
| Mean CAM-S | 0.72 | 0.72 | 0.64 | 0.61 | 0.62 | 1.3 |
| **Requires information on delirium duration and delirium at discharge:** | | | | | | |
| Total number of delirium days | 0.65 | 0.60 | 0.60 | 0.57 | 0.60 | 4.2 |
| Delirium at discharge | 0.60 | 0.56 | 0.61 | 0.55 | 0.61 | 4.1 |
| Change in cognition | 0.57 | 0.62 | NA | 0.53 | NA | 4.4 |

*PR project recovery, SAGES Successful AGing after Elective Surgery, NA not available, CAM-S Confusion Assessment Method-Severity score*
*\*C-statistic: used as a metric for comparison across measures; high values not expected (see text for details). The c-statistic for the categorical measure is reported here, which coincides with the results for the categorical measures shown in Tables 3and 4*
*†Death included as outcome in PR sample, not SAGES sample*
*‡Average rank computed as: C-statistics ranked within column (high to low), divided by number of rows in column, averaged across available columns, and multiplied by 5*

and combination delirium episode severity measures with both shorter term (30-day) and longer term (90-day) outcomes. Other important strengths of the study are the inclusion of two large datasets from distinct clinical settings (a medical and surgical sample) of older adults with rich data collection that allowed for quantifying delirium episode severity.

One of our aims was to examine the relative contributions of delirium intensity or duration as predictors of 30- and 90-day outcomes. Both appear to be important, and combined may best reflect the severity of the pathologic insult or precipitating factors associated with delirium. However, the intensity and duration measures were highly correlated and very few patients exhibited high delirium intensity with few delirium days, or conversely, low delirium intensity with many delirium days (eFigures 3 and 4). Thus, we were unable to conduct the appropriate statistical analyses to evaluate their independent contributions in this study. While our results demonstrate that measures considering both delirium intensity and duration are superior for predicting the clinical impact of an episode of delirium, we were unable to evaluate whether intensity or duration considered separately was more important. Future work is needed to address this important area. While the measures based on daily measurements of CAM-S (required to capture both the intensity and duration of delirium) performed in a superior fashion, we acknowledge the additional

time and expense to obtain sequential CAM-S ratings may limit the broad applicability of this approach. Availability of brief approaches to measure CAM-S (<3–5 min)[25,26] will enhance this process. Moreover, such measures will provide the best insight into identifying patients at highest risk of poor clinical outcomes, who could benefit most from intervention. Under circumstances with limited time and resources, the use of measures that consider solely delirium intensity or duration is the next preferred measure for capturing the severity of an episode of delirium.

This study provides proof-of-concept for the importance of both delirium intensity and duration combined. Our findings align with and extend previous work that has linked delirium severity, delirium duration, and outcomes. A study of older hip fracture patients undergoing surgery reported a higher risk of death and nursing home placement after 6-months among delirious patients with severe relative to mild delirium.[27] Several studies have documented that persistent delirium (defined differently according to study as delirium 1- or 6-months following discharge) was associated with poor outcomes, including increased nursing home placement, functional decline, and mortality.[4,28,29] Our findings serve to extend the previous work by combining delirium intensity and duration measurement within the context of an entire episode of delirium and to demonstrate the predictive validity for both 30- and 90-day outcomes, and illustrating a graded dose–response relationship. Future work will be needed to describe the characteristics of patients in the highest levels of the delirium episode severity measures, including how many have subsyndromal delirium and baseline cognitive impairment or dementia.

Several caveats about this study are important to mention. First, our analyses were limited to the examination of a single intensity measure (CAM-S). It will be important to compare and examine other delirium intensity measures, such as the Memorial Delirium Assessment Scale, Delirium Rating Scale, and Delirium Index in future work. Second, the characteristics of both study populations may not be generalizable to all older adults. Future validation of our findings in other samples and settings will be important. Third, while the sum of all CAM-S scores consistently demonstrated the highest ranking by c-statistics across all of the clinical outcomes examined (with one exception), we acknowledge that the sum, peak, and mean CAM-S measures are highly correlated and their respective results are very close. One of our major points is that all of these measures require daily CAM-S scores, and thus, daily measurement of delirium severity offers the optimal approach to measuring delirium severity, when possible.

This work is innovative in advancing the conceptualization of delirium that is damaging or harmful in terms of long-term outcomes. The finding that both the intensity and duration of delirium contribute synergistically to poor outcomes may help to advance both future pathophysiologic studies and to focus

future interventions. Our findings suggest that interventions that address both the intensity of delirium and its duration will be key to improve clinical outcomes. Importantly, interventions that reduce intensity but increase duration (e.g., antipsychotics) may not have a positive impact on clinical outcomes.[30] Ultimately, delirium episode severity measures, such as the sum of all CAM-S scores, should prove useful for monitoring delirium clinically and for providing a quantifiable dose–response outcome measure for intervention trials and for pathophysiologic investigations, all vital steps to advance our understanding of this common, morbid, and costly geriatric syndrome.

*Authors' Contributions:* *Study concept and design: Vasunilashorn, Inouye, Jones*
*Acquisition, analysis, or interpretation of data: Vasunilashorn, Marcantonio, Gou, Pisani, Travison, Schmitt, Jones, Inouye*
*Drafting of the manuscript: Vasunilashorn, Inouye*
*Critical revision of the manuscript for important intellectual content: Vasunilashorn, Marcantonio, Gou, Pisani, Travison, Schmitt, Jones, Inouye*
*Statistical analysis: Vasunilashorn, Gou*
*Obtained funding: Inouye, Jones, Marcantonio*
*Administrative, technical, or material support: Vasunilashorn, Marcantonio, Gou, Pisani, Travison, Schmitt, Jones, Inouye*
*Study supervision: Marcantonio, Jones, Inouye*

*Corresponding Author:* *Sarinnapha M. Vasunilashorn, PhD; Beth Israel Deaconess Medical Center, Boston, MA, USA (e-mail: svasunil@bidmc.harvard.edu).*
**Compliance with Ethical Standards:**

**Conflict of Interest:** *The authors declare no conflicts of interest.*

## REFERENCES

1. **Inouye SK, Westendorp RG, Saczynski JS.** Delirium in elderly people. Lancet. 2014;383:911–22.
2. **Rudolph JL, Jones RN, Rasmussen LS, Silverstein JH, Inouye SK, Marcantonio ER.** Independent vascular and cognitive risk factors for postoperative delirium. Am J Med. 2007;120:807–13.
3. **Martin BJ, Buth KJ, Arora RC, Baskett RJ.** Delirium as a predictor of sepsis in post-coronary artery bypass grafting patients: a retrospective cohort study. Crit Care. 2010;14:171.
4. **Marcantonio ER, Flacker JM, Michaels M, Resnick NM.** Delirium is independently associated with poor functional recovery after hip fracture. J Am Geriatr Soc. 2000;48:618–24.
5. **Witlox J, Eurelings LS, de Jonhe JF, Kalisvaart KJ, Eikelenboom P, van Gool WA.** Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. JAMA. 2010;304:443–51.
6. **Marcantonio ER, Goldman L, Mangione CM, et al.** A clinical prediction rule for delirium after elective noncardiac surgery. JAMA. 1994;271:134–9.
7. **Inouye SK, Zhang Y, Han L, Leo-Summers L, Jones R, Marcantonio E.** Recoverable cognitive dysfunction at hospital admission in older persons during acute illness. J Gen Intern Med. 2006;21:1276–81.
8. **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;168:27–32.
9. **Adamis D, Sharma N, Whelan PJP, MacDonald AJD.** Delirium scales: a review of current evidence. Aging Ment Health. 2010;14:543–55.
10. **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;160:526–33.
11. **Schuurmans MJ, Deschamps PI, Markham SW, Shortridge-Baggett LM, Duursma SA.** The measurement of delirium: review of scales. Res Theory Nurs Pract. 2003;17:207–24.
12. **Smith MJ, Breitbart WS, Platt MM.** A critique of instruments and methods to detect, diagnose, and rate delirium. J Pain Symptom Manag. 1995;10:35–77.
13. **Trzepacz PT.** A review of delirium assessment instruments. Gen Hosp Psychiatry. 1994;16:397–405.
14. **Trzepacz PT, Baker RW, Greenhouse J.** A symptom rating scale for delirium. Psychiatry Res. 1988;23:89–97.
15. **Schmitt EM, Marcantonio ER, Alsop DC, et al.** Novel risk markers and long-term outcomes of delirium: the Successful Aging after Elective Surgery (SAGES) study design and methods. J Am Med Dir Assoc. 2012;13:1–10.
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. In press.
17. **Inouye SK, Bogardus ST Jr, Charpentier PA, et al.** A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med. 1999;340:669–76.
18. **Inouye SK, Leo-Summers L, Zhang Y, Bogardus ST Jr, Leslie DL, Agostini JV.** A chart-based method for identification of delirium: validation compared with interviewer ratings using the Confusion Assessment Method. J Am Geriatr Soc. 2005;53:312–8.
19. **Leslie DL, Zhang Y, Holford TR, Bogardus ST, Leo-Summers LS, Inouye SK.** Premature death associated with delirium at 1-year follow-up. Arch Intern Med. 2005;165:1657–62.
20. **Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Hortwitz RI.** Clarifying confusion: the Confusion Assessment Method. A new method for detection of delirium. Ann Intern Med. 1990;113:941–8.
21. **Inouye SK.** The Confusion Assessment Methods (CAM): training manual and coding guide. New Haven: Yale University School of Medicine; 2003. http://www.hospitalelderlifeprogram.org/uploads/disclaimers/Long_CAM_Training_Manual_10-9-14.pdf Accessed 7 Mar 2016.
22. **McAvay GJ, Van Ness PH, Bogardus ST Jr, et al.** Older adults discharged from the hospital with delirium: 1-year outcomes. J Am Geriatr Soc. 2006;54:1245–50.
23. **Davis DH, Skelly DT, Murray C, et al.** Worsening cognitive impairment and neurodegenerative pathology progressively increase risk for delirium. Am J Geriatr Psychiatry. 2015;23:403–15.
24. **Leslie DL, Zhang Y, Bogardus ST, Holford TR, Leo-Summers LS, Inouye SK.** Consequences of preventing delirium in hospitalized older adults on nursing home costs. J Am Geriatr Soc. 2005;53:405–9.
25. **Marcantonio ER, Ngo LH, O'Connor M, et al.** 3D-CAM: derivation and validation of a 3-minute diagnostic interview of CAM-defined delirium: a cross-sectional diagnostic test study. Ann Intern Med. 2014;161:554–61.
26. **Inouye SK.** Hospital Elder Life Program (HELP). http://www.hospital-elderlifeprogram.org/ Accessed 7 Mar 2016.
27. **Marcantonio E, Ta T, Duthie E, Resnick NM.** Delirium severity and psychomotor types: their relationship with outcomes after hip fracture repair. J Am Geriatr Soc. 2002;50:850–7.
28. **Cole MG, Ciampi A, Belzile E, Zhong L.** Persistent delirium in older hospital patients: a systematic review of frequency and prognosis. Age Ageing. 2009;38:19–26.
29. **Kiely DK, Marcantonio ER, Inouye SK, et al.** Persistent delirium predicts greater mortality. J Am Geriatr Soc. 2008;57:55–61.
30. **Page VJ, Ely EW, Gates S, et al.** Effect of intravenous haloperidol on duration of delirium and coma in critically ill patients (Hope-ICU): a randomised, double-blind, placebo-controlled trial. Lancet Respir Med. 2013;1:515–23.