



Published in final edited form as:

*Crit Care Med.* 2022 January 01; 50(1): e11–e19. doi:10.1097/CCM.0000000000005224.

## Physiological Assessment of Delirium Severity: The Electroencephalographic Confusion Assessment Method Severity Score (E-CAM-S)

Meike van Sleuwen, BSc<sup>1,2</sup>, Haoqi Sun, PhD<sup>1</sup>, Christine Eckhardt, MD<sup>1</sup>, Anudeepthi Neelagiri, MD<sup>1</sup>, Ryan A. Tesh, BSc<sup>1</sup>, Mike Westmeijer, MSc<sup>1</sup>, Luis Paixao, MD<sup>1</sup>, Subapriya Rajan, MD<sup>1</sup>, Parimala Velpula Krishnamurthy, MD<sup>1</sup>, Pooja Sikka, MD<sup>1</sup>, Michael J. Leone, MSc<sup>1</sup>, Ezhil Panneerselvam, MD<sup>1</sup>, Syed A. Quadri, MD<sup>1</sup>, Oluwaseun Akeju, MD<sup>1</sup>, Eyal Y. Kimchi, MD, PhD<sup>1</sup>, M. Brandon Westover, MD, PhD<sup>1</sup>

<sup>1</sup>Department of Neurology, Massachusetts General Hospital, Boston, MA

<sup>2</sup>University of Twente, Enschede, The Netherlands

### Abstract

**OBJECTIVES:** Delirium is a common and frequently underdiagnosed complication in acutely hospitalized patients, and its severity is associated with worse clinical outcomes. We propose a physiologically based method to quantify delirium severity as a tool that can help close this diagnostic gap: the Electroencephalographic Confusion Assessment Method Severity Score (E-CAM-S).

**DESIGN:** Retrospective cohort study.

**SETTING:** Single-center tertiary academic medical center.

**PATIENTS:** Three-hundred seventy-three adult patients undergoing electroencephalography to evaluate altered mental status between August 2015 and December 2019.

**INTERVENTIONS:** None.

**MEASUREMENTS AND MAIN RESULTS:** We developed the E-CAM-S based on a learning-to-rank machine learning model of forehead electroencephalography signals. Clinical delirium severity was assessed using the Confusion Assessment Method Severity (CAM-S). We compared associations of E-CAM-S and CAM-S with hospital length of stay and inhospital mortality. E-CAM-S correlated with clinical CAM-S ( $R = 0.67$ ;  $p < 0.0001$ ). For the overall cohort, E-CAM-S and CAM-S were similar in their strength of association with hospital length of stay (correlation

---

For information regarding this article, [mwestover@mgh.harvard.edu](mailto:mwestover@mgh.harvard.edu).

Ms. van Sleuwen, Dr. Sun, and Dr. Eckhardt are co-first authors.

Drs. Kimchi and Westover are co-senior authors.

Ms. van Sleuwen and Drs. Sun, Eckhardt, Kimchi, and Westover were involved in design of the work. Ms. van Sleuwen, Dr. Neelagiri, Mr. Westmeijer, Dr. Rajan, Dr. Velpula Krishnamurthy, Mr. Leone, and Drs. Panneerselvam, Quadri, and Westover were involved in data collection. Ms. van Sleuwen, Dr. Sun, Dr. Neelagiri, Mr. Westmeijer, Mr. Leone, and Dr. Kimchi were involved in data preparation and analysis. Ms. van Sleuwen van Sleuwen and Drs. Sun, Eckhardt, Akeju, Kimchi, and Westover were involved in drafting the work. All authors were involved in revising critically.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (<http://journals.lww.com/ccmjournal>).

= 0.31 vs 0.41, respectively;  $p = 0.082$ ) and inhospital mortality (area under the curve = 0.77 vs 0.81;  $p = 0.310$ ). Even when restricted to noncomatose patients, E-CAM-S remained statistically similar to CAM-S in its association with length of stay (correlation = 0.37 vs 0.42, respectively;  $p = 0.188$ ) and inhospital mortality (area under the curve = 0.83 vs 0.74;  $p = 0.112$ ). In addition to previously appreciated spectral features, the machine learning framework identified variability in multiple measures over time as important features in electroencephalography-based prediction of delirium severity.

**CONCLUSIONS:** The E-CAM-S is an automated, physiologic measure of delirium severity that predicts clinical outcomes with a level of performance comparable to conventional interview-based clinical assessment.

### Keywords

clinical outcomes; delirium severity; electroencephalography; machine learning

---

Delirium is an acute and fluctuating disturbance of consciousness (1), common in hospitalized patients across many medical specialties (1, 2). Delirium is associated with worse clinical outcomes (3, 4), including increased length of stay (LOS), worse functional outcomes as assessed by the Glasgow Outcome Scale, and increased mortality (5). Nevertheless, delirium remains largely underdiagnosed (6, 7). Increasing evidence shows that not only the presence of delirium but also its severity are associated with worse prognosis (8). Measuring delirium severity is important for assessing prognosis, monitoring response to treatment, and anticipating the burden of care for patients both during and after hospitalization. Currently, delirium severity is primarily assessed using clinical tools, but these involve intermittent and subjective evaluation of a dynamic, complex condition and can generate disagreement among experts (9). An automated method that quantifies the presence and severity of delirium directly based on assessment of brain physiology could enable the development of more effective treatments and prevention strategies for delirium (10).

Early studies showed that qualitative features of electroencephalography (EEG) data are associated with delirium presence (11, 12) and severity (13, 14). EEG slowing, an increase of delta (1–4 Hz) and/or theta power (4–8 Hz) or a decrease of alpha power (8–12 Hz), correlates with the presence of delirium across various types of delirium presentations (5, 11, 15). In current practice, EEGs are analyzed using visual interpretation by clinical experts rather than quantitative analysis. Limitations of visual EEG interpretation include interrater variability and the use of only a small number of relatively simple descriptive features, typically scored for their presence or absence. An automated method able to provide a quantitative assessment of the degree of EEG abnormality may provide better monitoring of delirium severity. Several prior studies address the potential of quantitative EEG analysis for delirium detection. Numan et al (16) showed that delirium could be detected using a single slowing parameter from a single EEG channel. Shinozaki et al (17) showed that a bispectral index EEG score was more strongly correlated with mortality than clinical delirium status. These findings suggest the potential of using simplified EEGs for routine screening purposes.

Here, we present the EEG Confusion Assessment Method Severity (E-CAM-S) score, an automated physiologic method for assessing the presence and severity of delirium using quantitative EEG in a large and heterogeneous patient population. We further evaluate which quantitative EEG features are most strongly associated with delirium severity. Last, we investigate whether the E-CAM-S is an independent predictor of important clinical outcomes, including hospital LOS and inhospital mortality.

## MATERIALS AND METHODS

### Study Setting and Participants

We conducted a single-center, retrospective, observational cohort study of adult inpatients who underwent EEG monitoring as a part of routine care at Massachusetts General Hospital between August 2015 and December 2019. Patients were excluded if they had a history of dementia, other intellectual disability, deafness, aphasia, or were non-English speaking. The study was conducted under a protocol (No. 2012P001929) approved by the Institutional Review Board using a waiver of written informed consent.

### Clinical Data

Delirium presence was assessed using the CAM short form (18). Delirium severity was assessed using the Confusion Assessment Method Severity (CAM-S) (19) (short form), which assigns a score between 0 and 7 (for details, see Supplemental Digital Content 1, <http://links.lww.com/CCM/G696>). Comatose patients (Richmond Agitation-Sedation Scale [RASS] score of  $-4$  or  $-5$ ) were assigned a CAM-S score of 7 (20, 21); however, all analyses were performed on both the entire cohort (nondelirious, delirious, and comatose patients) and after excluding comatose patients. LOS, inhospital mortality, and Charlson Comorbidity Index (CCI) (22) were extracted and calculated from the medical record.

### EEG Recording and Preprocessing

We calculated the E-CAM-S using only four frontal EEG channels, as forehead electrodes are amenable to application with minimal technical experience. These channels were: Fp1–Fp2, Fp1–F7, Fp2–F8, F7–F8. Details of EEG signal preprocessing are in Supplemental Digital Content 1 (<http://links.lww.com/CCM/G696>).

### Feature Extraction, Model Training, and Cross Validation

From each 6-second epoch, we extracted 298 features (summarized in Supplemental Table 1, <http://links.lww.com/CCM/G696>) (23, 24). We used 10-fold external cross validation (ECV) to evaluate model performance (Supplemental Fig. 3, <http://links.lww.com/CCM/G696>). Within each fold, we split data into training and testing data at the patient level, with 90% of EEGs ( $n = 336$ ) used for training and 10% ( $n = 37$ ) for testing. For each fold, we fit the model (including feature selection) only on the training data and measure model performance only on the held-out test data. In this process, each subject ends up being used once in testing and nine times in training. The choice of 10 for the number of folds in ECV is widely accepted for developing and evaluating machine learning models, as it achieves a favorable bias-variance trade-off (25, 26). Additional technical details about ECV are provided in Supplemental Digital Content 1 (<http://links.lww.com/CCM/G696>).

For feature selection during model training, we used a two-step approach of: 1) selecting the top k-features that showed the strongest Spearman correlation with the CAM-S score on the training data and 2) fitting a least absolute shrinkage and selection operator (LASSO)-penalized Learning-to-Rank (LTR) ordinal regression machine learning model (27) (described further below). We selected the values k and the value of the LASSO penalty parameter value using internal cross validation. Technical details are provided in Supplemental Digital Content 1 (<http://links.lww.com/CCM/G696>).

We created the E-CAM-S by training a LTR ordinal regression machine learning model (27) that attempts to produce scores (between 0 and 1) correlated with the clinical CAM-S score (0–7). The distribution of clinically assessed delirium severity scores in the entire cohort and in the noncomatose subset is shown in Supplemental Figure 1 (<http://links.lww.com/CCM/G696>). Technical details are in Supplemental Digital Content 1 (<http://links.lww.com/CCM/G696>).

### **Association of E-CAM-S With Mortality and Hospital Length of Stay**

We evaluated the association of E-CAM-S scores with inhospital mortality using multivariable logistic regression, including age, sex, and CCI as additional covariates. Age, CAM-S, and CCI were z normalized prior to model fitting. We assessed association with hospital LOS using multivariable linear regression with log-transformed LOS as the dependent variable. We performed both linear and logistic regression with four models: without any delirium information, with E-CAM-S scores included, with clinically assessed CAM-S scores, and with both E-CAM-S and CAM-S scores. Results are reported as Spearman correlations and area under the receiver operating characteristic curve (AUROC). We further compared the point-biserial correlation (PBC) between mortality and E-CAM-S and CAM-S.

### **Statistical Reporting**

Statistical reporting conventions are described in Supplemental Digital Content 1 (<http://links.lww.com/CCM/G696>). To evaluate the correlation between E-CAM-S and CAM-S, we used Spearman correlation coefficients. To evaluate the ability of E-CAM-S to discriminate between patients with versus without delirium, we used AUROC. We also compared the E-CAM-S with a previously published method (16) for assessing delirium based on the EEG, using Spearman correlations with the CAM-S and AUROC for predicting delirium presence as evaluation metrics.

## **RESULTS**

### **Patient Characteristics**

A total of 403 patients were enrolled in our study. Of these patients, 30 were subsequently excluded due to technical difficulties with the EEG, a diagnosis of dementia, missing data or because the time interval between the clinical and EEG test time was too large. Of the 373 patients analyzed, 252 (67.6%) screened positive for delirium by the CAM (67.6% of total 373 patients), and 122 were comatose (32.7% of total 373 patients). Patients with delirium were generally more ill and had worse clinical outcomes than patients without delirium

(Table 1), including longer LOS and higher rates of in-hospital mortality. Patients with delirium were older, had lower RASS scores, and higher CAM-S and CCI. The noncomatose subset differed from the entire cohort in terms of having lower CAM-S and in-hospital mortality rates, higher RASS, and shorter hospital LOS (Table 1; and Supplemental Fig. 2, <http://links.lww.com/CCM/G696>).

### Physiologic Measurement of Delirium Severity: The E-CAM-S

Example EEGs of patients with low, moderate, and high CAM-S scores are shown in Figure 1. Patients with higher delirium severity exhibited qualitatively more slowing of the EEG. EEGs were used to generate 1,192 extracted features, from which machine learning computed a corresponding E-CAM-S score, to reflect delirium severity. E-CAM-S scores successfully correlated with clinical CAM-S scores ( $R = 0.68$ ;  $p < 0.001$ ) (Fig. 2). The correlation between E-CAM-S and CAM-S scores was lower but still significant when limiting the analysis only to noncomatose patients ( $R = 0.52$ ;  $p < 0.001$ ; Supplemental Table 2 and Supplemental Fig. 4, <http://links.lww.com/CCM/G696>).

### EEG Features Predictive of Delirium Severity

We next explored which types of EEG features were most informative in predicting delirium severity. Machine learning based feature selection resulted in a retained subset of 53 features using the entire population and 23 features for the noncomatose subset. The top nine most important features are shown in Supplemental Figure 5 (<http://links.lww.com/CCM/G696>). SDS across epochs were primarily selected instead of minimum, maximum, and average values. Frequency features based on delta, theta, and alpha activity were also strong contributors to the E-CAM-S score. Other important features reflected differences in amplitude, variance, and regularity of the EEG signal in the time domain.

### Comparison of E-CAM-S to Assessment Using Single Slowing Parameter

A previous study (16) proposed a method to detect delirium using a one-channel EEG recording and a single measure of slowing, either relative delta power (1–4 Hz) or relative power from 1 to 6 Hz. Albeit in a different context, we compared how well these features correlate with CAM-S scores compared with the E-CAM-S. The results (Table 2) show that the E-CAM-S correlates more strongly with CAM-S than either the relative delta power ( $p < 0.001$ ) or the power from 1 to 6 Hz ( $p < 0.001$ ) under these conditions. In terms of predicting delirium presence, E-CAM-S also performed better than power from 1 to 6 Hz ( $p = 0.030$ ) and similarly to relative delta power ( $p = 0.070$ ) in this context.

### Association E-CAM-S With Clinical Outcomes

We next investigated the association of E-CAM-S with relevant clinical outcomes using a multivariable regression model including the E-CAM-S and other covariates (age, sex, and CCI).

**Association With Hospital Length of Stay.**—For the entire cohort, E-CAM-S was significantly associated with hospital LOS after adjusting for age, sex, and CCI (Table 3). This association was similar to that of the CAM-S score (correlation with LOS: E-CAM-S,

0.33; CAM-S, 0.41;  $p = 0.082$ ). Models with E-CAM-S and CAM-S scores also showed comparable associations with LOS in the noncomatose subset (correlation: 0.37 vs 0.42;  $p = 0.310$ ) (Table 3). The correlation of a combined model using E-CAM-S, CAM-S, age, sex, and CCI was 0.43 (0.34–0.51) and 0.46 (0.34–0.56) for the entire cohort and noncomatose subset, respectively (Table 3). This is similar to and with CI containing the correlation values obtained above for models using only CAM-S or E-CAM-S alone.

**Association With Inhospital Mortality.**—For the entire cohort, E-CAM-S was associated with in-hospital mortality after adjusting for age, sex, and CCI (Table 3). The strength of this association was similar to that of CAM-S (AUROC: E-CAM-S 0.77 [0.72–0.82] CAM-S 0.81 [0.75–0.85];  $p = 0.188$ ). Models with E-CAM-S and CAM-S scores also showed similar age, sex, and CCI adjusted associations with mortality for the noncomatose subset (AUROC: = 0.83 [0.76–0.90] vs 0.74 [0.62–0.84];  $p = 0.112$ ) (Table 3).

The AUROC of a combined model using E-CAM-S, CAM-S, age, sex, and CCI, was 0.80 (0.76–0.86), similar to and with CI containing the AUROC values obtained above for models using only CAM-S or E-CAM-S alone (Table 3). The PBC between mortality and the delirium scores showed overlapping CIs (E-CAM-S, 0.36 [0.28–0.45]; CAM-S, 0.40 [0.34–0.46]).

## DISCUSSION

In this study, we developed an automated physiologically based method to measure delirium severity using EEG, the E-CAM-S score. Our results show that the E-CAM-S, based on signals from four frontal EEG leads, reliably quantifies delirium severity and is independently associated with hospital LOS and in-hospital mortality across a wide range of acutely hospitalized adults. We also found that the strengths of the associations of clinical CAM-S and the E-CAM-S with clinical outcomes are comparable. These results establish E-CAM-S as a promising tool for physiologically based monitoring of delirium.

Our model showed strong performance for measuring delirium severity with the use of only four frontal EEG leads. We compared our results to a reduced EEG method used in a previous study, based on relative power in lower frequencies (either 1–4 Hz or 1–6 Hz) (16). E-CAM-S performed at least as well in this specific context, suggesting that using multiple features may be useful for classifying a patient's level of delirium severity. However, our method is not completely comparable to the previous study, which used a 1-minute EEG recording from different electrodes to detect the binary presence or absence of delirium. In contrast, we used a longer EEG recording timeframe from forehead electrodes to predict ordinal delirium severity.

Regression analysis with our predicted scores and potential covariates showed that both clinical CAM-S and EEG-based E-CAM-S assessment of delirium had strong associations with clinical outcomes, that is, in-hospital mortality and length of hospital stay. Both CAM-S and E-CAM-S show similarly strong associations with LOS and in-hospital mortality, reflecting that patients with higher delirium severity stay longer in the hospital and experience a higher probability of death, in keeping with previous studies (5). Regression

analysis with a combination of age, sex, CCI, and both CAM-S + E-CAM-S did not show a stronger association with clinical outcomes. E-CAM-S and CAM-S also showed overlapping PBCs, suggesting that the associations of E-CAM-S and CAM-S with mortality are similar. Thus, E-CAM-S appears to capture information regarding mortality risk similar to CAM-S rather than complementary.

Our results also provide insights into EEG features associated with delirium severity and inhospital mortality. In agreement with prior research (6, 10), our results indicate that spectral content in the delta, theta, and alpha bands are important EEG findings in delirium. Distinct from prior studies, our results highlight a new defining EEG feature of delirium: variability over time. In the E-CAM-S, the majority of selected features were based on the sd across epochs. This suggests that variation over time (between epochs) is more related to delirium severity than the average, minimum, or maximum values across epochs. This phenomenon might be related to the clinical observation that delirium can fluctuate over time.

While our results demonstrate successful automated detection of delirium severity, future steps may enhance EEG-based delirium measurement. For example, further studies may benefit from deep learning models, which often outperform conventional machine learning algorithms. Qualitative EEG analysis may also prove complementary to quantitative EEG analysis, because qualitative EEG may include features potentially missed by current quantitative methods. Further research should investigate to what extent important EEG features will be missed with quantitative EEG analysis. Additionally, we are planning a prospective study to validate the efficacy of E-CAM-S in quantifying delirium and to further clarify its prognostic value.

Our work has important limitations. Clinical CAM-S assessments were performed by multiple providers, which may have introduced interrater variability. Second, our model is best described as prognostic rather than mechanistic. The physiologic interpretation of the features selected by the model remains partly unclear, and more research to investigate the exact meaning and relationship between these features should be performed before clinical adoption is possible. Third, we performed our analysis on a heterogeneous patient population with various causes of altered mental status. Further research could be done to study different subsets of patients to relate selected features to underlying diseases. Fourth, we limited our analysis to four frontal EEG leads; additional EEG leads might improve performance. However, requiring multiple leads would reduce usability and impede our goal of achieving a simple, user-friendly method for rapid delirium detection and screening purposes. Fifth, our study was conducted at a single center. To see if our results can be generalized, external validation is required. Sixth, in this study, we assessed variability but did not include EEG reactivity. Studies have shown that absence of EEG reactivity is associated with later development of delirium among ICU patients with sepsis (28), which raises the possibility that EEG reactivity could enhance the performance of the E-CAM-S. However, our goal was to create an automated physiologic test that does not require trained human involvement, and assessment of EEG reactivity would require a standardized examination performed by a human operator. Studies of EEG reactivity have highlighted variability between experts in EEG reactivity testing among hospitals (29), which further

highlights the difficulty of utilizing reactivity in a widespread, standardized format, even with trained individuals. Seventh, E-CAM-S does not take into account the effects of concurrent drugs in a direct way. Drugs (e.g., benzodiazepines, anesthetics, opioids) are important contributors to delirium and, in some cases, have large effects on both mental status and EEG patterns. However, delirium rating scales (e.g., CAM-S) usually do not address the cause of delirium; they only quantify it. We followed the same approach in developing E-CAM-S. We feel that this approach is appropriate, as our main goal was to quantify the degree of abnormality in a patient's brain activity, regardless of the underlying cause. Nevertheless, the same elevated E-CAM-S score, like a clinical finding that a patient is in coma, may not be concerning if we know a patient is receiving anesthetic drugs but may be alarming in the absence of drugs. Thus, future work on the association between E-CAM-S and outcomes should account for the presence and type of drugs. Eighth, Figure 2 shows that E-CAM-S and CAM-S are correlated, but the relationship shows considerable spread. However, this does not necessarily mean that those cases are all false positives and negatives. CAM-S is not a "gold standard," and our primary goal is not to reproduce CAM-S but to create a severity measure that is physiologic and easy and reproducible. The E-CAM-S provides an objective assessment of severity of delirium with comparable clinical prognostic value as CAM-S, given their correlations with mortality. Nevertheless, when adapting E-CAM-S into clinical practice, it will be important to educate clinicians that E-CAM-S values (like a normal CAM-S value) can be falsely negative and must therefore be interpreted within the full clinical context of other clinical data. Last, we acknowledge that the availability of EEG in many hospitals does not currently match the ubiquity of delirium, because of the intensive clinical resources required to apply and interpret EEG. However, we anticipate that EEG will continue to become more widely available through advances in hardware that simplify data collection. Frontal EEG is increasingly accessible in a range of clinical practices, including technologies such as EEG headbands and dry EEG electrodes (30).

While standard EEG is not suitable for mass screening due to its size and the required expertise for placement and interpretation, simplified EEG devices show potential for automatic detection and routine screening of a patient's level of delirium severity. We will only be able to use such hardware through accompanying innovations in software, including algorithms similar to ours, that help streamline clinical interpretation. Even without the adoption of new EEG recording devices, we have demonstrated clinical value with only four standard frontal leads, which often can be applied without technician assistance.

## CONCLUSIONS

The E-CAM-S automatically quantifies delirium severity in patients with a variety of delirium presentations. The physiologically derived E-CAM-S and the manually assessed CAM-S show comparable strengths of associations with inhospital mortality and hospital LOS.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.



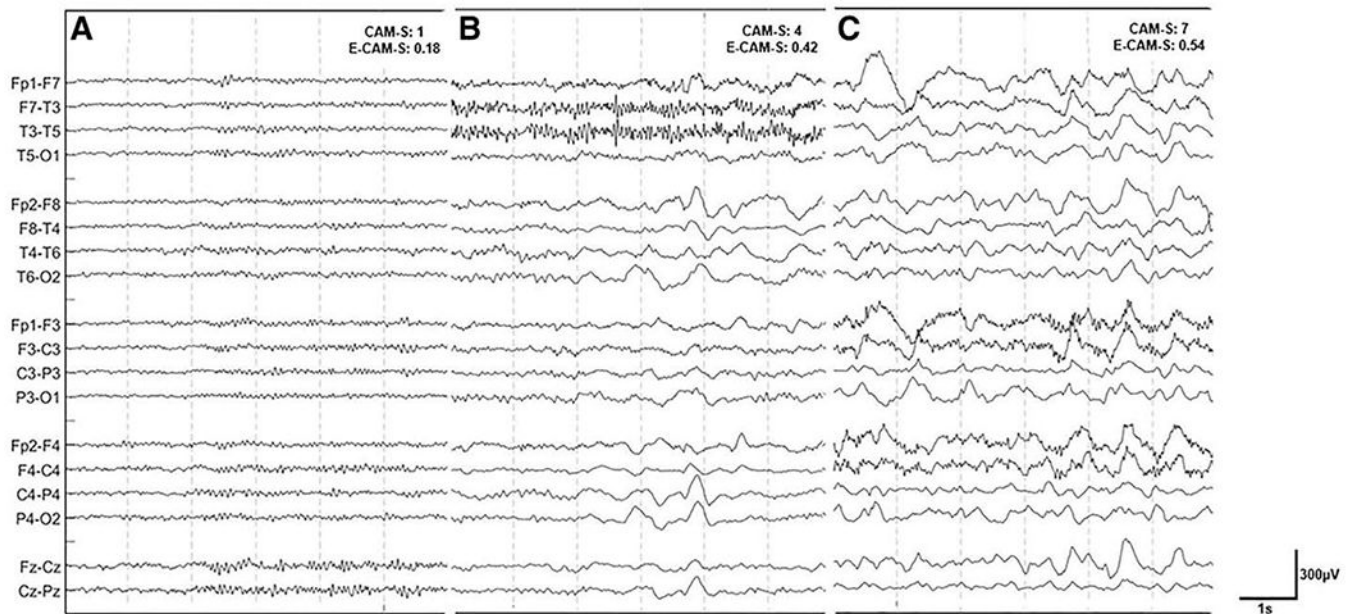
## Acknowledgments

Drs. Kimchi's and Westover's institutions received funding from the National Institutes of Health (NIH). Drs. Kimchi and Westover received support for article research from the NIH. Dr. Westover received funding from Beacon Biosignals. The remaining authors have disclosed that they do not have any potential conflicts of interest.

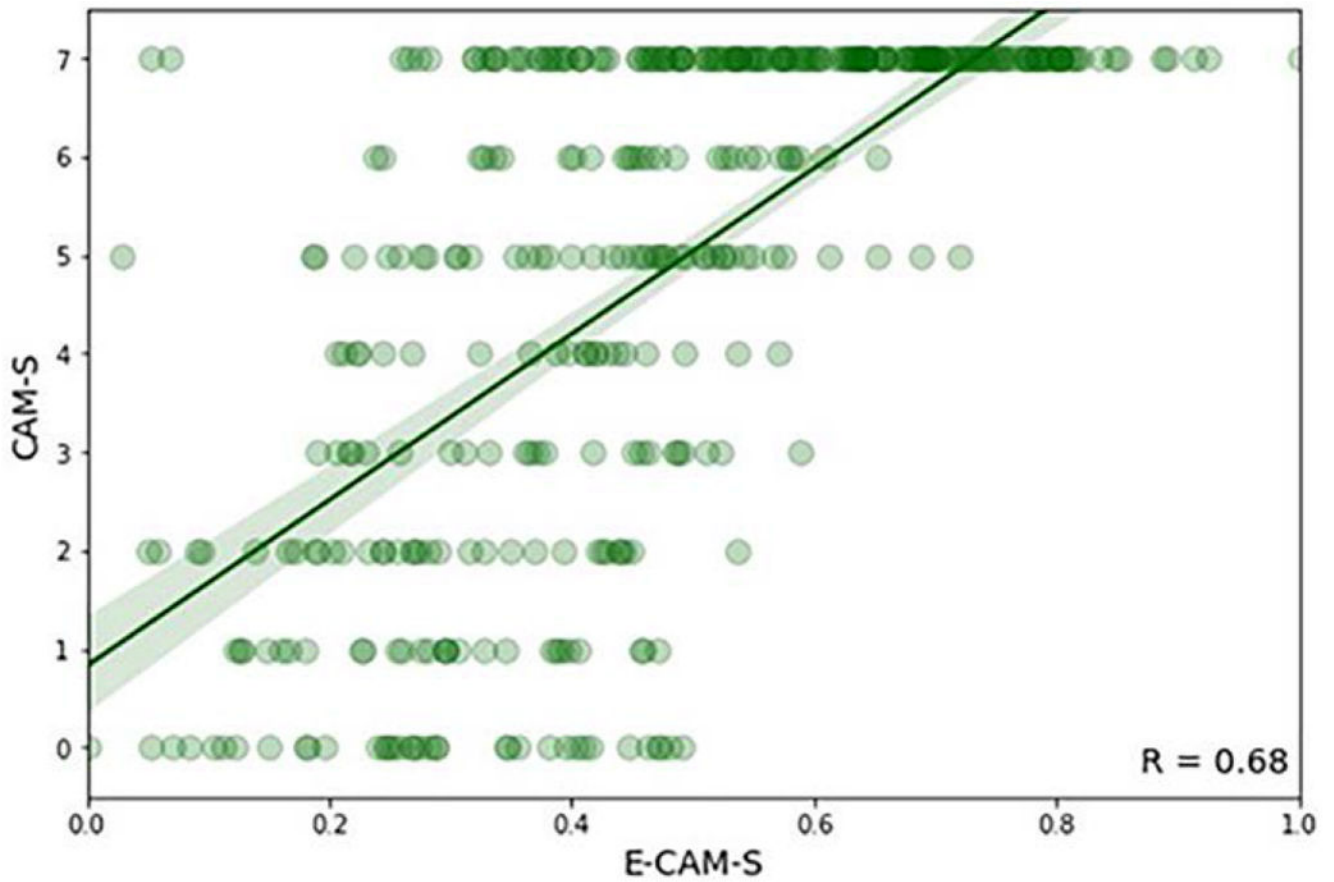
## REFERENCES

1. Cruz-Velarde JA, Gil de Castro R, Vázquez Allén P, et al. [Study of inpatient consultation for the neurological services]. *Neurologia* 2000; 15:199–202 [PubMed: 10850119]
2. Ances B: The more things change the more they stay the same: A case report of neurology residency experiences. *J Neurol* 2012; 259:1321–1325 [PubMed: 22186851]
3. Pandharipande PP, Girard TD, Jackson JC, et al. BRAIN-ICU Study Investigators: Long-term cognitive impairment after critical illness. *N Engl J Med* 2013; 369:1306–1316 [PubMed: 24088092]
4. Pisani MA, Kong SY, Kasl SV, et al. Days of delirium are associated with 1-year mortality in an older intensive care unit population. *Am J Respir Crit Care Med* 2009; 180:1092–1097 [PubMed: 19745202]
5. Kimchi EY, Neelagiri A, Whitt W, et al. Clinical EEG slowing correlates with delirium severity and predicts poor clinical outcomes. *Neurology* 2019; 93:e1260–e1271 [PubMed: 31467255]
6. Inouye SK, Westendorp RG, Saczynski JS: Delirium in elderly people. *Lancet* 2014; 383:911–922 [PubMed: 23992774]
7. Spronk PE, Riekerk B, Hofhuis J, et al. Occurrence of delirium is severely underestimated in the ICU during daily care. *Intensive Care Med* 2009; 35:1276–1280 [PubMed: 19350214]
8. Vasunilashorn SM, Fong TG, Albuquerque A, et al. Delirium severity post-surgery and its relationship with long-term cognitive decline in a cohort of patients without dementia. *J Alzheimers Dis* 2018; 61:347–358 [PubMed: 29171992]
9. Numan T, van den Boogaard M, Kamper AM, et al. Dutch Delirium Detection Study Group: Recognition of delirium in postoperative elderly patients: A multicenter study. *J Am Geriatr Soc* 2017; 65:1932–1938 [PubMed: 28497575]
10. Jones RN, Cizginer S, Pavlech L, et al. Better Assessment of Illness (BASIL) Study Group: Assessment of instruments for measurement of delirium severity: A systematic review. *JAMA Intern Med* 2019; 179:231–239 [PubMed: 30556827]
11. van der Kooi AW, Zaai IJ, Klijn FA, et al. Delirium detection using EEG: What and how to measure. *Chest* 2015; 147:94–101 [PubMed: 25166725]
12. Shafi MM, Santarnecchi E, Fong TG, et al. Advancing the neurophysiological understanding of delirium. *J Am Geriatr Soc* 2017; 65:1114–1118 [PubMed: 28165616]
13. Parsons-Smith BG, Summerskill WH, Dawson AM, et al. The electroencephalogram in liver disease. *Lancet* 1957; 273:867–871 [PubMed: 13482229]
14. Thomas C, Hestermann U, Kopitz J, et al. Serum anticholinergic activity and cerebral cholinergic dysfunction: An EEG study in frail elderly with and without delirium. *BMC Neurosci* 2008; 9:86 [PubMed: 18793418]
15. Boord MS, Moezzi B, Davis D, et al. Investigating how electroencephalogram measures associate with delirium: A systematic review. *Clin Neurophysiol* 2021; 132:246–257 [PubMed: 33069620]
16. Numan T, van den Boogaard M, Kamper AM, et al. Dutch Delirium Detection Study Group: Delirium detection using relative delta power based on 1-minute single-channel EEG: A multicentre study. *Br J Anaesth* 2019; 122:60–68 [PubMed: 30579407]
17. Shinozaki G, Bormann NL, Chan AC, et al. Identification of patients with high mortality risk and prediction of outcomes in delirium by bispectral EEG. *J Clin Psychiatry* 2019; 80:19m12749
18. 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]

19. 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; 160:526–533 [PubMed: 24733193]
20. Oldham MA, Holloway RG: Delirium disorder: Integrating delirium and acute encephalopathy. *Neurology* 2020; 95:173–178 [PubMed: 32518149]
21. Slooter AJC, Otte WM, Devlin JW, et al. Updated nomenclature of delirium and acute encephalopathy: Statement of ten societies. *Intensive Care Med* 2020; 46:1020–1022 [PubMed: 32055887]
22. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis* 1987; 40:373–383 [PubMed: 3558716]
23. Sun H, Jia J, Goparaju B, et al. Large-scale automated sleep staging. *Sleep* 2017; 40:zsx139
24. Donoghue T, Haller M, Peterson EJ, et al. Parameterizing neural power spectra into periodic and aperiodic components. *Nat Neurosci* 2020; 23:1655–1665 [PubMed: 33230329]
25. Kohavi R: A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI* 1995; 14:1137–1143
26. James G, Witten D, Hastie T, et al. *An Introduction to Statistical Learning: With Applications in R*. New York, NY, Springer, 2013
27. Burges C, Shaked T, Renshaw E, et al. Learning to rank using gradient descent. *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*. Bonn, Germany, ACM Press, 2005, pp 89–96
28. Azabou E, Navarro V, Kubis N, et al. Value and mechanisms of EEG reactivity in the prognosis of patients with impaired consciousness: A systematic review. *Crit Care* 2018; 22:184 [PubMed: 30071861]
29. Amorim E, Gilmore EJ, Abend NS, et al. EEG reactivity evaluation practices for adult and pediatric hypoxic-ischemic coma prognostication in North America. *J Clin Neurophysiol* 2018; 35:510–514 [PubMed: 30216207]
30. Singla S, Garcia GE, Rovenolt GE, et al. Detecting seizures and epileptiform abnormalities in acute brain injury. *Curr Neurol Neurosci Rep* 2020; 20:42 [PubMed: 32715371]



**Figure 1.** Electroencephalography (EEG) patterns vs. delirium severity. Examples of three EEG recordings of patients with low (A), moderate (B), and high (C) Confusion Assessment Method Severity (CAM-S) scores, along with their corresponding Electroencephalography Confusion Assessment Method Severity (E-CAM-S) scores. EEGs were processed using a notch filter at 60 Hz and a bandpass filter from 0.5 to 30 Hz.



**Figure 2.** Scatter plot of electroencephalography-based delirium severity prediction (Electroencephalography Confusion Assessment Method Severity [E-CAM-S]) versus Confusion Assessment Method Severity (CAM-S) scores. The *green line* represents a fitted regression line with 95% CI.

**TABLE 1.** Patient Characteristics Based on Confusion Assessment Method Defined Delirium and Differences for the Entire Cohort (Including Nondelirious, Delirious, and Comatose Patients) and Noncomatose Subset

<b>Quantitative Data, Median (IQR)</b>	<b>No Delirium (n = 121)</b>	<b>Delirium (n = 252)</b>	<b>p (Rank-Sum)</b>
Age (yr)	54 (40–68)	63 (54–73)	0.0021
Charlson Comorbidity Index	3 (1–5)	4 (3–6)	< 0.0001
RASS score (–5 to +4)	0 (0–0)	–3 (–5 to –1)	< 0.0001
Delirium severity (CAM-S)	1 (0–2)	7 (5–7)	< 0.0001
Length of stay (d)	6 (4–11)	15 (8–25)	< 0.0001
<b>Categorical Data, % (n)</b>	<b>No Delirium (n = 121)</b>	<b>Delirium (n = 252)</b>	<b>p (χ<sup>2</sup>)</b>
Gender (Female)	42.1% (n = 51)	42.5% (n = 107)	0.8877
Hospital Mortality	0.8% (n = 1)	31.7% (n = 80)	< 0.0001
<b>Quantitative Data, Median (IQR)</b>	<b>All Patients (n = 373)</b>	<b>Noncomatose Subset (n = 251)</b>	<b>p (Rank-Sum)</b>
Age (yr)	61 (49–72)	61 (50–72)	0.9070
Charlson Comorbidity Index	4 (2–6)	4 (2–5)	0.0630
RASS score (–5 to +4)	–1 (–4 to 0)	0 (–1 to 0)	< 0.0001
Delirium severity (CAM-S)	6 (2–7)	4 (2–6)	< 0.0001
Length of stay (d)	11 (6–21)	9 (5–16)	0.0114
<b>Categorical Data, % (n)</b>	<b>All Patients (n = 373)</b>	<b>Noncomatose Subset (n = 251)</b>	<b>p (χ<sup>2</sup>)</b>
Gender (female)	41.6% (n = 157)	43.8% (n = 110)	0.6678
Hospital mortality	21.7% (n = 81)	6.8% (n = 17)	< 0.0001

CAM-S = Confusion Assessment Method Severity, IQR = interquartile range, RASS = Richmond Agitation-Sedation Scale.

**TABLE 2.**

Association of Delirium Severity and Delirium Presence With Either the Electroencephalography Confusion Assessment Method Severity or Relative Power in Lower Frequency Bands

Evaluation Metrics	Electroencephalography Confusion Assessment Method		
	Severity	Relative 1–4 Hz Power	Relative 1–6 Hz Power
Correlation (R) with delirium severity (Confusion Assessment Method Severity)	0.68 (0.64–0.73)	0.33 (0.24–0.42)	0.36 (0.27–0.45)
Area under the receiver operating characteristic curve for nondelirious vs delirious	0.70 (0.64–0.76)	0.65 (0.59–0.70)	0.63 (0.58–0.69)

Data are reported as median (95% CIs).

**Prediction Performance of Various Multivariable Models for Predicting the Inhospital Mortality and Logarithm of Length of Stay for the Entire Cohort or the Noncomatose Subset**

**TABLE 3.**

Model	Log LOS All Patients (Correlation)	Mortality All Patients (AUROC)	Log LOS Noncomatose Patients (Correlation)	Mortality Noncomatose Patients (AUROC)
Baseline model	0.04 (-0.15 to 0.18)	0.65 (0.57-0.71)	0.16 (-0.02 to 0.30)	0.64 (0.52-0.75)
Baseline model + CAM-S	0.41 (0.32-0.50)	0.81 (0.75-0.85)	0.42 (0.30-0.53)	0.74 (0.62-0.84)
Baseline model + E-CAM-S	0.33 (0.23-0.44)	0.77 (0.72-0.82)	0.37 (0.23-0.48)	0.83 (0.76-0.90)
Baseline model + CAM-S + E-CAM-S	0.43 (0.33-0.51)	0.80 (0.76-0.86)	0.46 (0.34-0.56)	0.82 (0.76-0.89)

AUROC = area under the receiver operating characteristic curve, CAM-S = Confusion Assessment Method Severity, E-CAM-S = Electroencephalography Confusion Assessment Method Severity, log LOS = logarithm of length of stay.

The baseline model included age, sex, and Charlson Comorbidity Index, onto which was added either the clinical delirium severity measure (CAM-S) or electroencephalography-based delirium severity measure (E-CAM-S). Data are reported as median (95% CIs).