Association of an Electroencephalography-Based Risk Score With Seizure Probability in Hospitalized Patients

Aaron F. Struck, MD; Berk Ustun, PhD; Andres Rodriguez Ruiz, MD; Jong Woo Lee, MD, PhD; Suzette M. LaRoche, MD; Lawrence J. Hirsch, MD; Emily J. Gilmore, MD; Jan Vlachy, MS; Hiba Arif Haider, MD; Cynthia Rudin, PhD; M. Brandon Westover, MD, PhD

IMPORTANCE Continuous electroencephalography (EEG) use in critically ill patients is expanding. There is no validated method to combine risk factors and guide clinicians in assessing seizure risk.

OBJECTIVE To use seizure risk factors from EEG and clinical history to create a simple scoring system associated with the probability of seizures in patients with acute illness.

DESIGN, SETTING, AND PARTICIPANTS We used a prospective multicenter (Emory University Hospital, Brigham and Women's Hospital, and Yale University Hospital) database containing clinical and electrographic variables on 5427 continuous EEG sessions from eligible patients if they had continuous EEG for clinical indications, excluding epilepsy monitoring unit admissions. We created a scoring system model to estimate seizure risk in acutely ill patients undergoing continuous EEG. The model was built using a new machine learning method (RiskSLIM) that is designed to produce accurate, risk-calibrated scoring systems with a limited number of variables and small integer weights. We validated the accuracy and risk calibration of our model using cross-validation and compared its performance with models built with state-of-the-art logistic regression methods. The database was developed by the Critical Care EEG Research Consortium and used data collected over 3 years. The EEG variables were interpreted using standardized terminology by certified reviewers.

EXPOSURES All patients had more than 6 hours of uninterrupted EEG recordings.

MAIN OUTCOMES AND MEASURES The main outcome was the average risk calibration error.

RESULTS There were 5427 continuous EEGs performed on 4772 participants (2868 men, 49.9%; median age, 61 years) performed at 3 institutions, without further demographic stratification. Our final model, 2HELP52B, had an area under the curve of 0.819 and average calibration error of 2.7% (95% CI, 2.0%-3.6%). It included 6 variables with the following point assignments: (1) brief (ictal) rhythmic discharges (BIRDs) (2 points); (2) presence of lateralized periodic discharges, lateralized rhythmic delta activity, or bilateral independent periodic discharges (1 point); (3) prior seizure (1 point); (4) sporadic epileptiform discharges (1 point); (5) frequency greater than 2.0 Hz for any periodic or rhythmic pattern (1 point); and (6) presence of “plus” features (superimposed, rhythmic, sharp, or fast activity) (1 point). The probable seizure risk of each score was 5% for a score of 0, 12% for a score of 1, 27% for a score of 2, 50% for a score of 3, 73% for a score of 4, 88% for a score of 5, and greater than 95% for a score of 6 or 7.

CONCLUSIONS AND RELEVANCE The 2HELP52B model is a quick accurate tool to aid clinical judgment of the risk of seizures in critically ill patients.

Copyright 2017 American Medical Association. All rights reserved.

Published online October 9, 2017.
Continuous electroencephalography (cEEG) provides real-time monitoring of brain function in hospitalized patients. The use of cEEG is expanding, motivated by reports showing a high incidence of subclinical seizures in encephalopathic patients with conditions ranging from sepsis to traumatic brain injury.1-3

Features of EEG reported as factors associated with seizures include epileptiform and periodic discharges.4 However, to our knowledge, no study has examined how these factors affect seizure risk jointly, that is, it is unknown how seizure risk changes when several patterns occur simultaneously.

We propose a simple scoring system for seizure risk that we refer to as the 2HELPS2B score. Our tool provides a joint assessment of seizure risk from cEEG observations and history of seizures, and it allows physicians to make accurate, risk-calibrated probabilities by hand. We expect our tool to help physicians identify patients in need of continued cEEG monitoring and who are likely to benefit from interventions.

### Methods

#### Patients

Following institutional review board approval at Emory University, Brigham and Women's Hospital, and Yale University, institutions prospectively entered participant data into an anonymized database.5 Waiver of consent was granted because of minimal risk to patients. The database includes reports of clinical information and findings on cEEG greater than or equal to 6 hours. The cEEG findings were coded using American Clinical Neurophysiology Society standardized terminology.6 Clinical variables were collected as described in Lee et al.5 Patients admitted for elective epilepsy monitoring were excluded. Data from 5427 cEEG sessions on 4772 different patients were collected. All investigators entering patient data had to undergo a module explaining the patterns and an examination demonstrating mastery of the material. This method has been shown to have high interrater reliability.7 Seizures are not defined in the American Clinical Neurophysiology Society terminology, but most clinicians used the modified Young et al8 criteria to define seizures. Both electrographic and electroclinical seizures were included.

#### Data Set Creation

We considered 24 candidate variables for inclusion in risk models (Table 1). Posterior dominant rhythm; brief (ictal) rhythmic discharges (BRD); reactivity; sporadic (nonperiodic and non-rhythmic) epileptiform discharges; history of seizure, generalized rhythmic delta activity (GRDA), localized rhythmic delta activity (LRDA), generalized periodic discharges (GPDs), localized periodic discharges (LPDs), and bilateral independent periodic discharges (BIPDs); primary neurological diagnosis (altered mental status, infection, inflammatory disease, cerebral neoplasm, hypoxic/ischemic encephalopathy, intracerebral hemorrhage, metabolic encephalopathy, stroke, subarachnoid hemorrhage, subdural hemorrhage, traumatic brain injury, and hydrocephalus); frequency of rhythmic or periodic patterns; presence of a stimulus-induced pattern; and presence of a “plus factor” (ie, superimposed rhythmic, fast, or sharp activity). Candidate variables were selected based on prevalence within the database and previous associations with seizures.

Variables were combined into single factors to simplify the prediction model and increase the effect size for each factor. This was performed for variables that are associated with a similar risk of seizures and rarely co-occur. To create a frequency binary variable, frequency was divided into binary variables at each 0.5-Hz interval from 0.5 to 3 Hz. Each potential dividing point was analyzed to find the cut point with maximal predictive value.

Descriptive statistics are reported with 95% CIs. Odds ratios and Fisher exact test results are reported for candidate variables with a set to .05.

#### Risk Score Methods

Our goal was to create a risk score similar to CHADS2 (congestive heart failure, hypertension, age greater than 75, diabetic, and history of stroke [doubled]),9 that is, a simple additive model with a limited number of factors and small integer weights for quick calculations. There is no standard method to create such models. Existing tools were built manually (eg, CHADS2 a point system for stroke risk with atrial fibrillation)10 or by combining logistic regression with ad hoc feature selection and rounding (eg, simplified acute physiology score [SAPS II], a point system for mortality in the intensive care unit).10

Existing approaches may fail to produce risk-calibrated models. Therefore, we built our risk score using a new method known as Risk-Calibrated Supersparse Linear Integer Model (RiskSLIM).11 This RiskSLIM method uses optimization techniques to find the best logistic regression model with bounded integer coefficients (integers between -10 and 10), and a limited number of risk factors (at most 6). In such settings, RiskSLIM can output an optimized risk score with superior risk-calibration and/or area under the curve (AUC). Because RiskSLIM is a new method, we compared RiskSLIM models with baseline models built using state-of-the-art methods: penalized logistic regression (PLR) with a combined L1/L2 penalty using the same constraints.

### Key Points

#### Question

Can the risk of seizures in critically ill patients be accurately determined with a simple clinical tool?

#### Findings

In this study, a point system using 6 variables (brief ictal rhythmic discharges [2 points]; presence of lateralized periodic discharges, lateralized rhythmic delta activity, or bilateral independent periodic discharges [1 point]; prior seizure [1 point]; sporadic epileptiform discharges [1 point]; frequency greater than 2.0 Hz of periodic/rhythmic pattern [1 point]; and presence of “plus” features [1 point]) was associated with seizure risk of 5% with a score of 0, 12% with a score of 1, 27% with the score of 2, 50% with the score of 3, 73% with a score of 4, 88% with a score of 5, and greater than 95% with a score of 6 or 7.

#### Meaning

The 2HELPS2B score may provide accurate seizure risk stratification from patient history and initial electroencephalography.
Risk Score Evaluation
We evaluated all models for accuracy and risk calibration (ie, how well the predicted probability of a seizure matches the true prevalence). To assess accuracy, we computed the area under the receiver operating characteristic curve (ROC). To assess risk calibration, we constructed reliability diagrams plotting the observed prevalence of seizures vs the predicted probability (eg, Figure, A). In addition, we examined the average calibration (CAL) error, the mean squared error between the predicted probability and the observed prevalence. When a model has perfect risk calibration, the reliability curve should lie on the 45° line, and CAL should be 0% (Figure, A). The average CAL error is a measure of how close the probable risk of seizures and the actual risk of seizures are. It is minimized to find the best risk model.

Risk Score Validation
We validated the performance of all models using standard 5-fold cross-validation (5-CV). That is, we randomly split the data into 5 parts, fit the model using 4 of 5 folds, and validated this model on last fold (that the model had not seen). This procedure was repeated 5 times, each time using a different fold for validation, to obtain 5 independent estimates of CAL and AUC. We report the mean of these estimates as 5-CV CAL and 5-CV AUC.

Because fitting models with PLR requires us to specify free parameters, we fit models for more than 1100 combinations of free parameters and picked the combination that maximized the 5-CV test AUC. This required us to validate the performance using a nested 5-CV procedure. All results for model performance are reported with respect to the left-out data (the fold used for testing) only; testing data were held out and were not used for either choosing the values of free parameters nor for training the model. This rigorous separation of training and testing data provides protection against overfitting and minimizes bias in the reported model performance.

Results
Patients
Among 5427 cEEG sessions, 719 (12.52%) had a seizure during cEEG; 2315 (40.03%) had GRDA, LRDA, BIPDs, LPDs, or GPDs. A total of 340 (5.92%) had sporadic epileptiform discharges.

Seizure Prediction Risk Score
After fitting several models using RiskSLIM and PLR for model size constraints ranging between 4 and 27, we selected a RiskSLIM model with 6 variables shown in Table 2.

In contrast to the baseline PLR model, the RiskSLIM model was simpler, had superior risk calibration (mean 5-CV CAL of 2.7% [95% CI, 2.0%-3.6%] vs 8.9% [95% CI, 7.9%-9.8%] for PLR), and had comparable AUC (mean 5-CV AUC of 0.819 [95% CI, 0.799-0.849] vs 0.821 [95% CI, 0.801-0.855] for PLR). We also compared 2HELPS2B with a PLR

Table 1. Univariate Risk Factor Analysis*  

<table>
<thead>
<tr>
<th>Variable</th>
<th>Proportion With S (95% CI)</th>
<th>No. (%) of Patients With Finding</th>
<th>OR (95% CI)</th>
<th>P Valueb</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDR</td>
<td>0.10 (0.08-0.12)</td>
<td>1166 (21.5)</td>
<td>0.711 (0.50-0.92)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BRDs</td>
<td>0.69 (0.62-0.76)</td>
<td>176 (3.2)</td>
<td>18.805 (18.47-19.14)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Unreactive background</td>
<td>0.18 (0.15-0.22)</td>
<td>454 (8.4)</td>
<td>1.610 (1.36-1.86)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Prior Sz</td>
<td>0.23 (0.20-0.25)</td>
<td>1168 (21.5)</td>
<td>2.663 (2.50-2.83)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>GRDA</td>
<td>0.13 (0.11-0.15)</td>
<td>927 (17.1)</td>
<td>1.071 (0.862-1.28)</td>
<td>.51</td>
</tr>
<tr>
<td>LRDA</td>
<td>0.28 (0.23-0.32)</td>
<td>410 (7.6)</td>
<td>2.967 (2.73-3.20)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>GPDs</td>
<td>0.16 (0.13-0.19)</td>
<td>696 (12.8)</td>
<td>1.368 (1.15-1.59)</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>LPDs</td>
<td>0.44 (0.41-0.48)</td>
<td>802 (14.8)</td>
<td>9.825 (9.65-10.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BIPDs</td>
<td>0.28 (0.21-0.36)</td>
<td>122 (2.2)</td>
<td>2.784 (2.38-3.19)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Infection</td>
<td>0.16 (0.11-0.24)</td>
<td>118 (2.2)</td>
<td>1.350 (0.853-1.85)</td>
<td>.26</td>
</tr>
<tr>
<td>Inflammation</td>
<td>0.13 (0.06-0.24)</td>
<td>56 (1.0)</td>
<td>0.998 (0.202-1.80)</td>
<td>.99</td>
</tr>
<tr>
<td>Neoplasm</td>
<td>0.19 (0.16-0.22)</td>
<td>577 (10.6)</td>
<td>1.694 (1.47-1.92)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>ICH</td>
<td>0.12 (0.10-0.15)</td>
<td>543 (10.0)</td>
<td>0.963 (0.693-1.23)</td>
<td>.84</td>
</tr>
<tr>
<td>Metabolic encephalopathy</td>
<td>0.06 (0.04-0.10)</td>
<td>325 (6.0)</td>
<td>0.467 (0.0177-0.916)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Stroke</td>
<td>0.10 (0.07-0.13)</td>
<td>452 (8.3)</td>
<td>0.737 (0.416-1.059)</td>
<td>.06</td>
</tr>
<tr>
<td>SAH</td>
<td>0.07 (0.05-0.09)</td>
<td>508 (9.4)</td>
<td>0.499 (0.140-0.845)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SDH</td>
<td>0.13 (0.10-0.17)</td>
<td>353 (6.5)</td>
<td>1.078 (0.760-1.40)</td>
<td>.62</td>
</tr>
<tr>
<td>TBI</td>
<td>0.07 (0.04-0.12)</td>
<td>142 (2.6)</td>
<td>0.523 (0.01-1.70)</td>
<td>.05</td>
</tr>
<tr>
<td>Hypoxic/ischemic</td>
<td>0.13 (0.10-0.16)</td>
<td>390 (7.2)</td>
<td>1.004 (0.684-1.311)</td>
<td>.99</td>
</tr>
<tr>
<td>I VH</td>
<td>0.10 (0.06-0.15)</td>
<td>146 (2.7)</td>
<td>0.736 (0.180-1.29)</td>
<td>.31</td>
</tr>
<tr>
<td>Hydrocephalus</td>
<td>0.07 (0.03-0.14)</td>
<td>86 (1.6)</td>
<td>0.520 (0.1-3.15)</td>
<td>.14</td>
</tr>
<tr>
<td>Discharges (&gt;2Hz)</td>
<td>0.29 (0.26-0.33)</td>
<td>763 (14.1)</td>
<td>3.733 (3.55-3.91)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Frequency (&gt;2Hz)</td>
<td>0.42 (0.32-0.53)</td>
<td>77 (1.4)</td>
<td>2.570 (1.62-4.09)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Abbreviations: BIPD, bilateral periodic discharges; BRDs, brief rhythmic discharges; GPDs, generalized periodic discharges; GRDA, generalized rhythmic delta activity; ICH, intracranial hemorrhage; IVH, intraventricular hemorrhage; LPDs, lateralized periodic discharges; LRDA, lateralized rhythmic delta activity; OR, odds ratio; PDR, posterior dominant rhythm; Prop, proportion with seizures; SAH, subarachnoid hemorrhage; SDH, subdural hemorrhage. Sz, seizure; TBI, traumatic brain injury.

* A total of 2868 were men (49.9%), median age: 61 years.

b P value calculated with Fisher exact test.

c Variable only evaluated for ictal-intercal continuum patterns.
model where we did not round the coefficients or constrain the number of variables. In this case, we did obtain a model with slightly better risk calibration (mean 5-CV CAL 2.0% [95% CI, 1.5%-3.0%]) and improved AUC (mean 5-CV AUC of 0.837 [95% CI, 0.815-0.868]), but this model is no longer simple enough to use for quick predictions as the points are not integers and it used 21 of the 29 variables.

As a mnemonic, we call this RiskSLIM model the 2HELPSB2B score, which represents GRDA, LRDA, BIPDs, LPDs, or GPDs with a frequency greater than 2 Hz and discharges. Note, no patients had 7 points in the cohort (all possible risk factors); hence, its noninclusion.

The risks of seizures for each possible 2HELPSB2B score are 5% for a score of 0, 12% for a score of 1, 27% for a score of 2, 50% for a score of 3, 73% for a score of 4, 88% for a score of 5, and greater than 95% for a score of 6 or 7. Table 2 provides a reference with the probabilities for each score from 1 to 6. The area under the ROC for this model applied to all patients was 0.819 and for the 5 folds ranged between 0.776 and 0.849.

Discussion

The 2HELPSB2B score is an accurate, simple, and clinically practical risk score for seizure occurrence in hospitalized patients undergoing cEEG. The large sample size of data collected at multiple institutions with a systematic application of standardized EEG nomenclature fostered development of a robust risk scoring system. The large sample size provides statistical power; the multiple institutions and uniform data collection ensure broad applicability.

The 2HELPSB2B system combines 5 readily observable EEG features with a single factor from the patient history (any known history of seizure, remote or acute) to assign a score between 0 and 7. The score has good face validity, being based on established clinical and EEG risk factors. Moreover, it shows excellent CAL: the probabilities it assigns for each level of risk look for bias. The number associated with each dot is the number of patients in the Critical Care EEG Research Consortium database with the associated number of points. B. Receiver operating characteristics curve for the RiskSLIM model with 95% CIs developed from bootstrapping from the full training set is represented by the dashed lines. The solid black line represents the null classifier. Area under the curve = 0.819.

Table 2. Optimized Risk Score for Seizure Probabilitya

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total Score</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>&gt;6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probable risk of Sz, %b</td>
<td>5</td>
<td>12</td>
<td>27</td>
<td>50</td>
<td>73</td>
<td>88</td>
<td>&gt;95</td>
<td></td>
</tr>
<tr>
<td>Actual prevalence of Sz, % (95% CI)c</td>
<td>3 (2-3)</td>
<td>12 (10-13)</td>
<td>34 (31-37)</td>
<td>52 (46-57)</td>
<td>71 (63-78)</td>
<td>84 (71-99)</td>
<td>92 (77-100)</td>
<td></td>
</tr>
</tbody>
</table>

a Scoring for each risk factor was 2 points for brief rhythmic discharge and 1 point each for lateralized periodic discharges/bilateral independent periodic discharges/lateralized rhythmic delta activity; plus features; prior seizure; frequency greater than 2 Hz; and discharges. Note, no patients had 7 points in the cohort (all possible risk factors); hence, its noninclusion.

b Probability of seizure presented as the mean; probable risk is the probability of seizure based on RiskSLIM.

c The numbers in parentheses are 95% CIs obtained using bootstrap resampling.
closely match those observed in our cohort. The association of higher frequency (>1.5-Hz) discharges and increased risk of seizures seen in the study by Rodriguez Ruiz et al was confirmed to have independent association value in the 2HELPS2B investigation.

The rigorous cross-validation method that we used and the large cohort size of 5427 ensures our results are widely applicable. Supporting the generalizability of our study, the incidence of seizures in our cohort is within the 8% to 34% range.1,4–22 Subgroups also have an incidence similar to prior studies, such as stroke at 10% (range, 6%-26%) and subarachnoid hemorrhage at 7% (range, 4%-19%).3,17

Limitations
There are some limitations of the study. The duration of cEEG was not included in the database; thus, this study does not address the change in probability of seizures with increased observation duration. This issue has been partially addressed in prior studies. Risk of a seizure within 72 hours was found to be less than 5% if a seizure was not detected within 16 hours of monitoring.2,4 Future studies should explore the association between the time-dependent risk for seizures under continued observation in relation to the 2HELPS2B score. No cEEG sessions of less than 6 hours were included in this study; hence, these criteria should be applied with caution to studies of less than 6 hours. However, a reasonable approach for use of the 2HELPS2B score would be to calculate the score at the initial reading of the cEEG, typically within the first half hour of recording (>68% of EEG abnormalities are evident by this time).2 If new EEG findings emerge, the 2HELPS2B score should be modified at the second reporting, typically on the order of 6 to 8 hours. By this time, 95% of epileptiform abnormalities have been detected.2 Initially, 2HELPS2B can serve as a tool to augment clinical judgment regarding duration of monitoring and need for antiseizure medications. We anticipate future clinical studies using 2HELPS2B as a risk-stratifying metric to define rigorous cut points guiding clinical management, similar to the way the CHADS2 score guides anticoagulation in atrial fibrillation.9

Conclusions
The 2HELPS2B score is an easy-to-use tool to augment clinical judgment of the risk for seizures in individual patients. The simplicity of the system allows for easy integration into clinical workflow. With increasing familiarity, 2HELPS2B will improve communication between EEG interpreters and clinicians through the use of a quickly comprehensible single number to describe seizure risk for patients on cEEG.