

Quantitative Electroencephalogram Trends Predict Recovery in Hypoxic-Ischemic Encephalopathy*

Mohammad M. Ghassemi, PhD¹; Edilberto Amorim, MD^{2,3}; Tuka Alhanai, MSc¹;
Jong W. Lee, MD, PhD⁴; Susan T. Herman, MD⁵; Adithya Sivaraju, MD⁶; Nicolas Gaspard, MD, PhD⁷;
Lawrence J. Hirsch, MD⁶; Benjamin M. Scirica, MD⁸; Siddharth Biswal, MSc⁹;
Valdery Moura Junior, MSc²; Sydney S. Cash, MD, PhD²; Emery N. Brown, MD, PhD^{10,11};
Roger G. Mark, MD, PhD^{1,12}; M. Brandon Westover, MD, PhD²; for the Critical Care
Electroencephalogram Monitoring Research Consortium

*See also p.1474.

¹Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

²Department of Neurology, Massachusetts General Hospital, Boston, MA.

³Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

⁴Department of Neurology, Brigham and Women's Hospital, Boston, MA.

⁵Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA.

⁶Department of Neurology, Yale School of Medicine, New Haven, CT.

⁷Department of Neurology, Universite Libre de Bruxelles, Brussels, Belgium.

⁸Division of Cardiology, Department of Medicine, Brigham and Women's Hospital, Boston, MA.

⁹School of Computer Science, Georgia Institute of Technology, Atlanta, GA.

¹⁰Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA.

¹¹Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA.

¹²Information Systems, Beth Israel Deaconess Medical Center, Boston, MA.

Drs. Ghassemi and Amorim contributed equally as co-first authors of this work.

The Critical Care Electroencephalogram Monitoring Research Consortium Board consists of: Chair: Brandon M. Westover, MD, PhD; Vice-Chair: Emily Gilmore, MD; Secretary: Aaron Struck, MD; Member-at-Large: Nicholas Gaspard, MD, PhD; Immediate Past Chair: Jong Woo Lee, MD, PhD; and Past Chair: Nicholas S. Abend, MD, MSCE. Drs. Ghassemi, Amorim, Lee, Cash, Brown, Mark, and Westover contributed to conception and design of the study. Drs. Ghassemi, Amorim, and Westover contributed to analysis of data. Drs. Ghassemi, Amorim, and Westover contributed to preparing the figures. Drs. Ghassemi and Amorim, Mr. Al Hanai, Drs. Lee, Herman, Sivaraju, and Gaspard, Mr. Biswal, Mr. Moura Junior, and Dr. Westover contributed to data acquisition. Drs. Ghassemi and Amorim, Mr. Al Hanai, Drs. Lee, Herman, Sivaraju, and Gaspard, Mr. Biswal, Mr. Moura Junior, and Drs. Cash, Brown, Mark, and Westover contributed to drafting the text.

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/ccmjjournal>).

Supported, in part, by grants from National Institutes of Health (NIH) 1R01NS102190, 1R01NS102574, and 1R01NS107291 (to Dr. Westover); R01GM104987 (to Dr. Mark); T32HL007901, T90DA22759, and Copyright © 2019 by the Society of Critical Care Medicine and Wolters Kluwer Health, Inc. All Rights Reserved.

DOI: 10.1097/CCM.0000000000003840

T32EB001680 (to Dr. Ghassemi); National Institute of Neurological Disorders and Stroke 1K23NS090900 (to Dr. Westover); Salerno foundation (M.G.M.); Neurocritical Care Society research training fellowship and American Heart Association postdoctoral fellowship (to Dr. Amorim); and Andrew David Heitman Neuroendovascular Research Fund and the Rappaport Foundation (to Dr. Westover).

Preliminary findings of this study were presented at the 14th Annual Neurocritical Care Society Meeting, National Harbor, MD, September 15–18, 2016.

Dr. Amorim's institution received funding from the National Institutes of Health (NIH), Neurocritical Care Society, and American Heart Association. Drs. Amorim, Mark, and Westover received support for article research from the NIH. Dr. Lee received funding from SleepMed/DigiTrace, Advance Medical, and United Diagnostics. Drs. Lee's and Mark's institutions received funding from the NIH. Dr. Herman's institution received funding from UCB Pharma, Sage Therapeutics, Neurospace, Epilepsy Therapy Development Project, Acorda Therapeutics, Pfizer, and Philips. Dr. Hirsch's institution received funding from Upsher-Smith and Monteris. He received funding from Adamas; consultation fees for advising from Aquestive, Ceribell, Eisai, and Medtronic; honoraria for speaking from Neuropace; and royalties for authoring chapters for UpToDate-Neurology and from Wiley for coauthoring a book on electroencephalograms in critical care. Dr. Scirica's institution received funding from Merck, Eisai, and Novartis, and he received consulting fees from AbbVie, Allergan, AstraZeneca, Boehringer Ingelheim, Covance, Eisai, Elsevier Practice Update Cardiology, GlaxoSmithKline, Lexicon, Merck, NovoNordisk, Sanofi, and equity in Health [at] Scale. Dr. Brown's institution received funding from Massachusetts General Hospital and Massachusetts Institute of Technology. The remaining authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: mwestover@mgh.harvard.edu; edilbertoamorim@gmail.com.

Objectives: Electroencephalogram features predict neurologic recovery following cardiac arrest. Recent work has shown that prognostic implications of some key electroencephalogram features change over time. We explore whether time dependence exists for an expanded selection of quantitative electroencephalogram features and whether accounting for this time dependence enables better prognostic predictions.

Design: Retrospective.

Setting: ICUs at four academic medical centers in the United States.

Patients: Comatose patients with acute hypoxic-ischemic encephalopathy.

Interventions: None.

Measurements and Main Results: We analyzed 12,397 hours of electroencephalogram from 438 subjects. From the electroencephalogram, we extracted 52 features that quantify signal complexity, category, and connectivity. We modeled associations between dichotomized neurologic outcome (good vs poor) and quantitative electroencephalogram features in 12-hour intervals using sequential logistic regression with Elastic Net regularization. We compared a predictive model using time-varying features to a model using time-invariant features and to models based on two prior published approaches. Models were evaluated for their ability to predict binary outcomes using area under the receiver operator curve, model calibration (how closely the predicted probability of good outcomes matches the observed proportion of good outcomes), and sensitivity at several common specificity thresholds of interest. A model using time-dependent features outperformed (area under the receiver operator curve, 0.83 ± 0.08) one trained with time-invariant features (0.79 ± 0.07 ; $p < 0.05$) and a random forest approach (0.74 ± 0.13 ; $p < 0.05$). The time-sensitive model was also the best-calibrated.

Conclusions: The statistical association between quantitative electroencephalogram features and neurologic outcome changed over time, and accounting for these changes improved prognostication performance. (*Crit Care Med* 2019; 47:1416–1423)

Key Words: cardiac arrest; electroencephalogram; hypoxic-ischemic encephalopathy; machine learning; quantitative electroencephalogram

Over 500,000 cardiac arrests occur every year in the United States (1). Most patients sustain severe brain injury resulting in coma (2). Despite recommendations to postpone prognostication at least 72 hours postarrest, the most common proximal cause of mortality is early withdrawal of life-sustaining therapies (3, 4). There is a need for accurate methods to assign probability of good and poor neurologic outcomes early after cardiac arrest, to reduce subjectivity, and to avoid poor outcomes as a result of self-fulfilling prophecies.

Specific electroencephalogram patterns are associated with eventual recovery from coma due to hypoxic-ischemic encephalopathy (HIE) after cardiac arrest (5, 6). However, existing electroencephalogram review practices rely on visual analysis, which does not translate into reproducible quantitative predictions of neurologic outcome (7–15). Several studies have shown that quantitative electroencephalogram (QEEG) features may carry prognostic information that is useful in outcome prediction on HIE (6, 13, 16–25). More recent models have achieved promising results by combining many QEEG features (26) and leveraging local temporal trends of multiple QEEG features (5).

In this study, we hypothesize that a machine-learning method can predict long-term neurologic outcome in HIE by integrating clinical data to existing and novel QEEG features. Our work goes beyond prior results by using a more comprehensive set of QEEG features and more completely leveraging temporal

trends, and by rigorously characterizing statistical performance of the model.

MATERIAL AND METHODS

Subjects and Clinical Management

Adult subjects diagnosed with in- or out-of-hospital cardiac arrest from October 2009 to April 2016 in four university-affiliated hospitals in the United States (Massachusetts General Hospital; Brigham and Women's Hospital; Yale New Haven Hospital; and Beth Israel Deaconess Medical Center) were retrospectively reviewed. Subjects who were comatose after return of spontaneous circulation (ROSC) underwent continuous electroencephalogram monitoring and received targeted temperature management (TTM) with goal temperature of 32–34°C were screened (TTM 36°C was not used at the time of data abstraction). Retrospective data collection and analysis was performed under independent Institutional Review Board approvals at participating hospitals and informed consent was waived.

Data Collection and Functional Outcome Assessment

Clinical and outcome information were collected retrospectively. Subjects were stratified by age, gender, time to ROSC, and initial cardiac rhythm, dichotomized as shockable (ventricular fibrillation or ventricular tachycardia) or nonshockable (asystole, pulseless electrical activity, and unknown).

The primary outcome was defined as the “best” neurologic function achieved up to 6 months after initial cardiac arrest, according to the Glasgow-Pittsburgh Cerebral Performance Categories (CPCs) scale, ascertained by medical record review (M.M.G., E.A., J.W.L., S.T.H., A.S., M.B.W.). “Good” outcome was defined as a CPC score of 1 or 2 and “poor” functional outcome as a CPC of 3–5 (27). All patients had CPCs assessed at discharge. Patients with CPC 3 and 4 were further assessed by chart review to determine whether they achieved a better CPC (1 or 2) in the following 6 months. CPC was chosen both because it is widely used in neurologic outcomes studies and is simple enough to be easily ascertained via chart review. Prior studies have shown that CPC can be extracted reliably via chart review (28). Reliability is further enhanced in our study by the lumping together of CPC scores into dichotomous “good” and “poor” outcomes.

Electroencephalogram Data Acquisition

Continuous electroencephalogram monitoring was started at each institution as early as possible during TTM and maintained for 24–72 hours, unless the subject regained consciousness, had life-sustaining therapies withdrawn, or died. Digital electroencephalogram was recorded using the international 10–20 system. Electroencephalogram was analyzed up to 72 hours of monitoring. Electroencephalogram preprocessing and artifact detection were performed before feature acquisition (**Supplementary materials**, Supplemental Digital Content 12, <http://links.lww.com/CCM/E728>).

To aid in reproducibility of this study, we have made our feature extraction code public: <https://github.com/deskool/ComaPrognosticanUsingEEG>.

QEEG Features Extraction

We extracted 52 QEEG features: 45 single channel and seven multichannel (Table s1, Supplemental Digital Content 1, <http://links.lww.com/CCM/E665>) (26, 29–39). Single-channel features were computed on the following frontotemporal electrodes: Fp1, Fp2, F3, F4, F7, F8, and Fz. Multichannel features were computed on the 21 unique pairs of the selected frontotemporal electrodes. To reduce feature dimensionality, the mean value of the extracted features across all channels was used. All electroencephalogram features were z scored (i.e., scaled to zero mean and unit variance). A more detailed description of the QEEG features extracted, corresponding equations, and QEEG processing pipeline is available in Table s1 (Supplemental Digital Content 1, <http://links.lww.com/CCM/E665>).

The 52 QEEG features fell into three electroencephalogram signal property domains: a) complexity features (21 total) quantify the degree of randomness or irregularity in the electroencephalogram signal (29, 30, 35–39); b) category features (24 total) quantify the degree to which brain states fall into certain key electroencephalogram patterns likely to carry prognostic significance (26); and c) connectivity features (seven total) quantify interactions across electrodes (26, 31–34). All analyses were performed using MATLAB 2016a (MathWorks, Natick, MA).

Statistical Analysis

Univariate Analysis. We began our study exploring Spearman correlations between our 56 features (52 QEEG and four clinical) and CPC scores at consecutive 12-hour time blocks, up to 72 hours.

Sequential Logistic Regression With Elastic Net Regularization. The prognostic importance of electroencephalogram features changes over time (5, 26). We accounted for this evolution in the feature-outcome relationship by training a contiguous sequence of logistic regression models, one for every 12 hours of electroencephalogram. All subjects with electroencephalogram discontinued “before” the end of a given time interval (e.g., before 48 hr in the 36- to 48-hr

interval) were excluded for the model trained “within” that interval. Therefore, each model was trained to predict 6-month neurologic outcome for subjects still monitored to the end of a given interval. Each model in the sequence included feature information from the current time interval, and feature information from preceding time intervals (e.g., the model at 36 hr contains feature information from hours 1–12, 13–24, and 25–36). This approach allowed models later in the sequence to consider both past and present feature information when making predictions. The number of subjects available for modeling purposes in each of the 12-hour time intervals, and their outcomes, are displayed in Table 1. In Figure 1, we graphically illustrate our approach. Features are selected using an Elastic Net model (see supplementary materials, Supplemental Digital Content 12, <http://links.lww.com/CCM/E728>) (40).

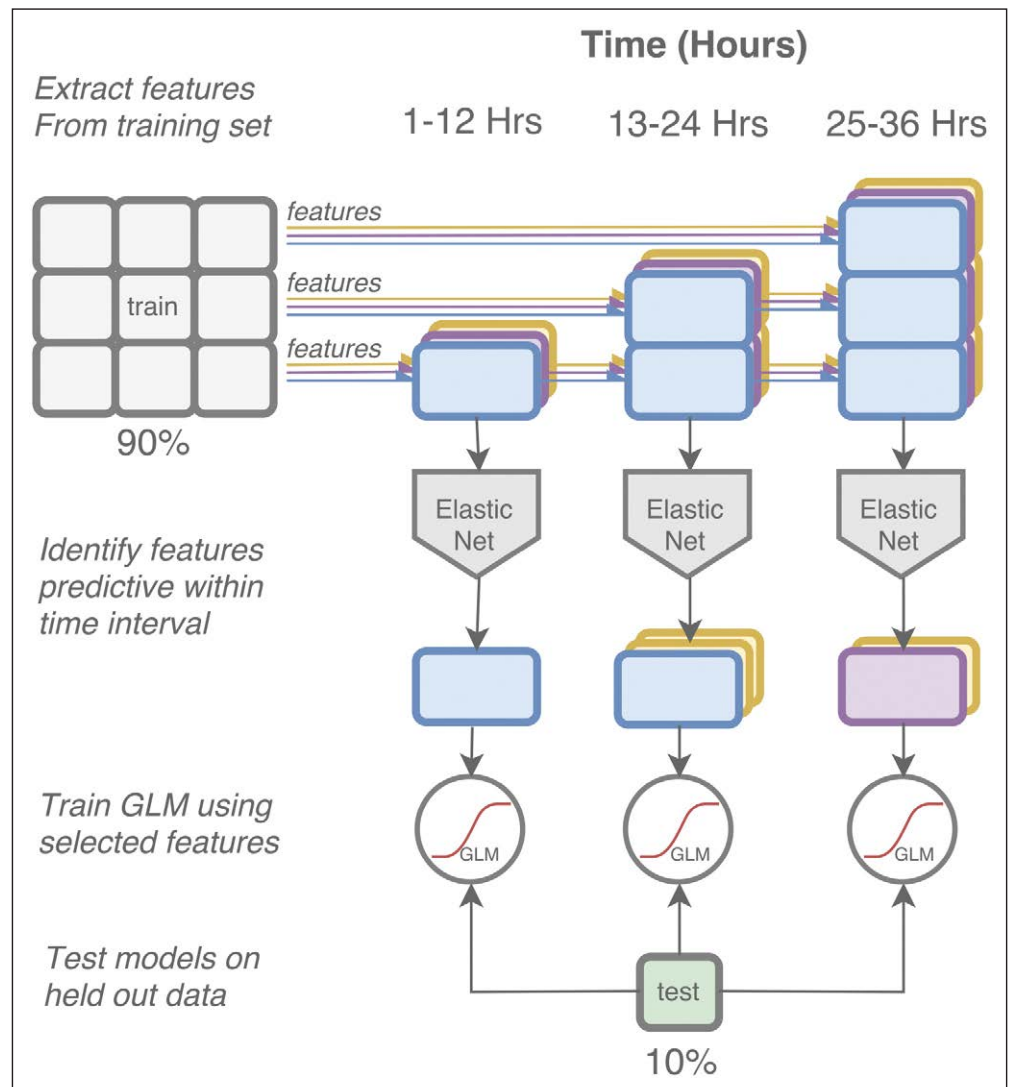


Figure 1. Modeling approach. An overview of our modeling approach. Features (depicted as colored rectangles) are extracted from contiguous temporal partitions of our training data. All features within and preceding a given temporal partition are provided to an Elastic Net regularized logistic regression model, which identifies the subset of features within the temporal partition most predictive of patient outcome. The selected features and training data are then used to generate a final general linear model (GLM)—the performance of which is evaluated on a held-out test set (green box).

TABLE 1. Collected Clinical Information on Patients, Partitioned by Cerebral Performance Category Outcome

Variable	CPC 1 (n = 111)	CPC 2 (n = 23)	CPC 3 (n = 9)	CPC 4 (n = 14)	CPC 5 (n = 281)
Age, yr (mean)	53 (16)	60 (14)	68 (13)	58 (20)	62 (17)
Male gender, %	59	70	30	36	67
Return of spontaneous circulation, min, (mean, SD)	20 (19)	14 (12)	13 (10)	13 (8)	28 (61)
Rhythm at arrest, %					
Asystole	5	13	10	7	18
Pulseless electrical activity	27	26	50	36	34
VFib	41	17	10	21	14
VT	7	4	0	7	3
Other (non-VFib or VT), %	20	40	30	29	31
Cause of arrest, %					
Cardiac	23	4	10	29	18
Pulmonary	5	13	10	7	12
Other/unknown	72	83	80	64	70
Arrest location, %					
In-hospital	18	30	50	0	10
Out-of-hospital	35	17	0	43	40
Unknown	47	53	50	57	50

CPC = Cerebral Performance Category, VFib = ventricular fibrillation, VT = ventricular tachycardia.

Henceforth, we refer to the Elastic Net regularized sequential logistic regression model as the time-sensitive model.

Baseline Models. We compared the time-sensitive model to four baseline approaches: 1) a clinical baseline: a logistic regression using age, gender, ROSC, and initial rhythm; 2) a time-insensitive model: a logistic regression model using features selected by Elastic Net across all time intervals, but retrained within each time interval; 3) the 2013 Cerebral Recovery Index (CRI): a heuristic approach using α/δ band power, SD, coherence in δ band, Shannon entropy, and regularity; and 4) a random Forest classifier with features and settings inspired by the 2017 CRI and the CRI methods were selected as literature baselines given their impressive reported performance. The 2017 CRI used the five features from the 2013 study, in addition to four features that characterized burst-suppression activity. The authors reported these additional features provided modest improvements in model performance compared with a random forest classifier using the original five features. For this reason, we used the five CRI features from the 2013 model when implementing the random forest classifier.

Performance Characterization Metrics. We evaluated model performance using electroencephalogram data for each consecutive 12-hour interval, up to 72 hours. Performance metrics included area under the receiver operating curve (AUC),

used to evaluate the model's ability to make binary predictions, and the sensitivity and specificity for specific operating points on the receiver operating curve; and statistical calibration, which measures how well the observed proportions of good outcomes match predicted probabilities. Calibration is particularly important for our model when used as a risk score.

Model Validation Approach. All models were validated using 10-fold cross-validation. That is, we partitioned available data into 10-folds. In each fold, 90% of the data were used to identify model parameters and the remaining 10% were used to evaluate model performance on subjects never seen in the corresponding training sets. Subjects within each testing fold were unique. We used the average performance of models across the unseen testing sets in the 10-folds when comparing performance. Feature selection was performed using an Elastic Net with 10 inner validation folds (supplementary materials, Supplemental Digital Content 12, <http://links.lww.com/CCM/E728>).

RESULTS

Data Characteristics

We identified 536 subjects across the four contributing hospitals. Of these, 76 subjects were excluded from final analysis due to missing admission clinical information and 22 were excluded

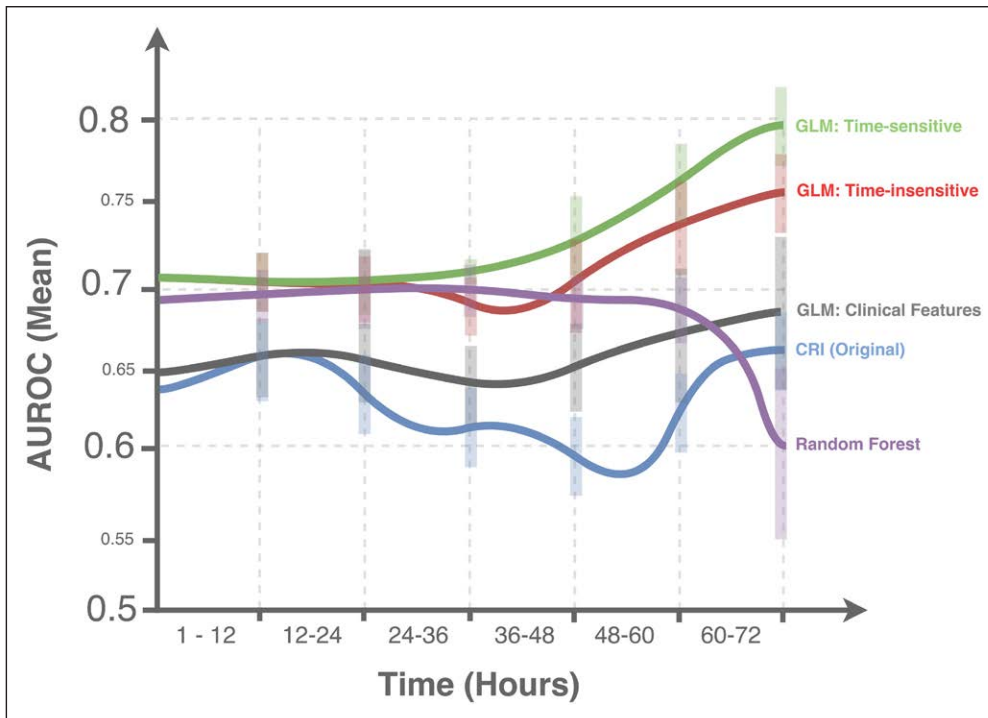


Figure 2. Model performance comparison as a function of time. A comparison of our time-sensitive (green) model's 10-fold area under the receiver operator curve (AUROC) performance on the held-out testing sets over time and four baseline approaches: time-insensitive (red), the 2013 Cerebral Recovery Index (CRI) (blue), random forest model inspired by the 2017 CRI (purple), and a model using four clinical features (age, gender, return of spontaneous circulation, and arrest rhythm) (gray). GLM = general linear model.

due to insufficient electroencephalogram recording quality or missing outcome assessments. The 438 remaining subjects had 12,397 hours of electroencephalogram data available.

By hospital discharge, 120 subjects had CPC of 1 and 2 and 281 subjects had a CPC of 5 (401/438, 91.6% of subjects). The remaining 8.4% (37/438) had a CPC of 3 or 4 at discharge. Among these, 14 ultimately improved to a CPC score of 1 or 2 by 6 months, five remained with CPC of 3 and 4, and 10 died. For the remaining eight cases (1.8% of 438), 6-month

outcomes could not be determined by chart review. For these, the discharge CPC score was carried forward and taken as the final outcome. In Table 1, Table s1 (Supplemental Digital Content 1, <http://links.lww.com/CCM/E665>), and Table s2 (Supplemental Digital Content 2, <http://links.lww.com/CCM/E666>), we summarize subject characteristics.

Univariate Analysis

Table s3 (Supplemental Digital Content 3, <http://links.lww.com/CCM/E667>) displays univariate Spearman correlations between features, and the CPC scores in contiguous 12-hour intervals. Several QEEG features were predictive exclusively in specific time windows (time sensitive) with correlation weights varying over time. Figure s1 (Supplemental Digital Content 4, <http://links.lww.com/CCM/E668>; legend, Supplemental Digital Content 12, <http://links.lww.com/CCM/E676>) illustrates three examples of features whose relationship with the most probable outcome changes over time, and three features whose prognostic value over time is stable.

Performance Characterization

We evaluate the ability of the models in two ways: first, to make accurate binary predictions (good vs poor outcome) and

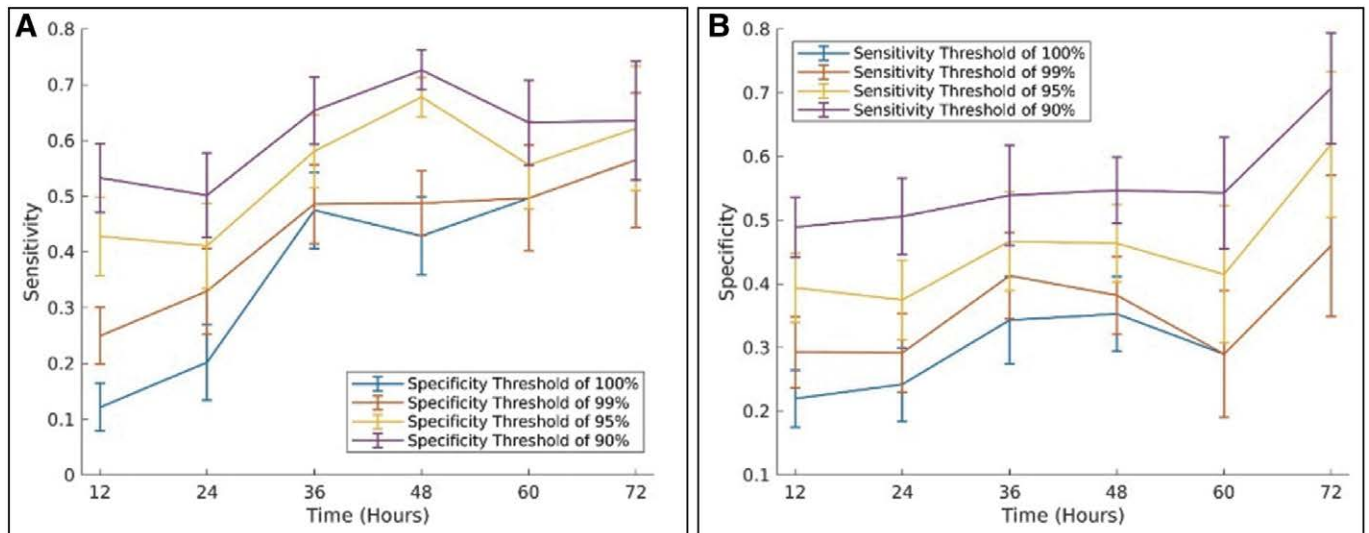


Figure 3. Sensitivity and specificity of the time-sensitive model at different sensitivity and specificity thresholds. **A.** The mean sensitivity (and \pm SE) of the proposed approach at a variety of sensitivity thresholds of interest. **B.** The mean specificity (and \pm SE) of the proposed approach at a variety of specificity thresholds of interest.

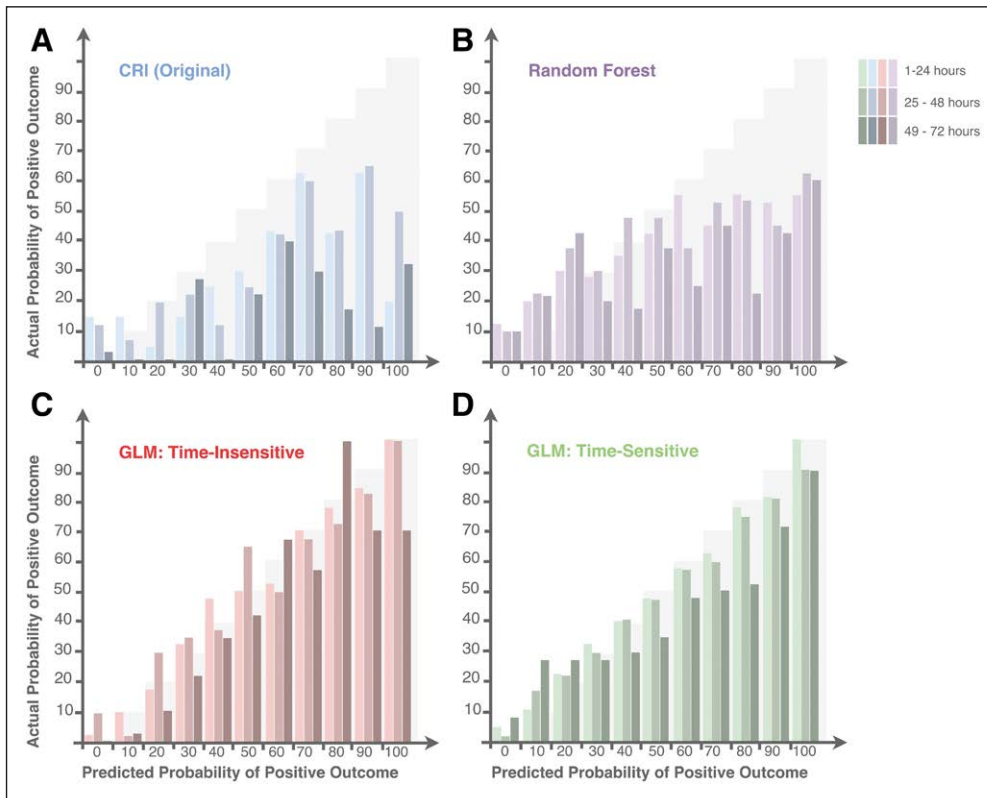


Figure 4. Model calibration performance comparison for each 24-hr block. A comparison of our model's calibration approach to three baseline approaches over 3 d postcardiac arrest. Ideal calibration is achieved when *color bars* perfectly overlap the *gray shading* in the background of each image. The calibration of a particular prediction level is shown as a *bar*. The *shading of the bar* reflects time (*lighter* is earlier and *darker* is later). **A**, The calibration of the 2013 Coma Recovery Index (CRI). **B**, The calibration of the random forest model based on the 2017 CRI. **C**, The calibration of a time-insensitive logistic regression model. **D**, The time-sensitive logistic regression model. GLM = general linear model.

second, to function as a risk scoring system by predicting the probability of a good (or, complementarily, a bad) outcome.

In **Figure 2**, we compare the performance (i.e. AUC) of the time-sensitive model to other baseline approaches predicting binary outcomes. Note that, the model for each time point is trained only electroencephalogram data from subjects who survived and remained on electroencephalogram up to that time; model predictions likewise pertain only to those who remain on monitoring. Here, the time-sensitive approach performed best, exhibiting consistent improvements in performance ($p < 0.04$, according to two-sample t test) with increased observation time (from AUC of 0.71 ± 0.05 at 12 hr to 0.79 ± 0.08 at 72 hr). The leave-one-institution-out cross-validation results are available in **Figure s2** (Supplemental Digital Content 5, <http://links.lww.com/CCM/E669>; legend, Supplemental Digital Content 12, <http://links.lww.com/CCM/E676>).

With regard to overall area under the receiver operating curve, AUC_0 (Fig. 2; and **Table s4**, Supplemental Digital Content 6, <http://links.lww.com/CCM/E670>), the time-sensitive model accurately predicted 6-month functional outcome (AUC_0 , 0.83 ± 0.08 across the 10 testing folds) compared with the time-insensitive model (AUC_0 , 0.79 ± 0.07), random forest (AUC_0 , 0.74 ± 0.13), original CRI (AUC_0 , 0.69 ± 0.07), and clinical baseline model (AUC_0 , 0.68 ± 0.05). The improvement

in prediction performance was statistically significant according to a two-sample t test ($p < 0.05$). In **Figure 3**, we report the overall sensitivity and specificity of the time-sensitive model over time. **Tables s5** (Supplemental Digital Content 7, <http://links.lww.com/CCM/E671>) and **Tables s6** (Supplemental Digital Content 8, <http://links.lww.com/CCM/E672>) contain sensitivity and specificity data for time-sensitive and time-insensitive models.

Final Model Using All Data

Table s7 (Supplemental Digital Content 9, <http://links.lww.com/CCM/E673>) shows coefficients for features selected by the time-sensitive model trained using all available data. The association of QEEG features with functional outcome varied across time. Not all types of electroencephalogram complexity positively correlated with outcomes. Fractal dimension and false nearest neighbor were associated with poor out-

comes, whereas entropy and cepstrum coefficients were associated with good outcomes. As expected, the association between outcome and electroencephalogram category depended on the particular category: epileptiform activity was associated with poor outcomes, whereas regularity (a measure of continuity) was associated with good outcomes. Increased connectivity between channels was consistently associated with poor outcomes.

Certain features were more predictive early on (Tsallis entropy, autoregressive-moving average, and coherence in the δ band), whereas other features were more predictive later (number of sharp waves, fractal dimension, and cross-correlation magnitude).

DISCUSSION

The technical contribution of this study lies in the methodologic approach we developed. Our modeling framework reflects the decision process of care providers, considering information across multiple points in time when predicting outcomes. We showed that the QEEG model which retains “memories” of previously encountered features outperformed state-of-the-art approaches which used only present features. The time-sensitive model had the best classification (overall AUC_0 , 0.83 ± 0.08) and statistical calibration of all tested approaches (**Fig. 4**; and **Table s5**, Supplemental Digital Content 7, <http://links.lww.com/CCM/E671>). These test characteristics show that our model

makes well-calibrated predictions. That is, the observed proportions of good outcomes match the predicted probabilities of good outcomes quite well. The superior calibration of our time-sensitive model provides strong motivation for its use as a risk score. In **Figure s3** (Supplemental Digital Content 10, <http://links.lww.com/CCM/E674>; legend, Supplemental Digital Content 12, <http://links.lww.com/CCM/E676>), we illustrate how outcome probabilities of the model could be used to score patient risk in a continuously updated fashion. An important strength of this study was its multicenter nature, incorporating data from four institutions and more than 400 subjects.

Several features selected by our final model reflect prior findings in the literature. We found that information-theoretic measures of electroencephalogram complexity and regularity were predictive of good functional outcomes, whereas features measuring epileptiform discharges were associated with poor outcomes (5, 17, 19, 22, 26, 41, 42). Two electroencephalogram complexity features (cepstrum and Tsallis entropy) and electroencephalogram regularity contributed to predictions in the first 24 hours of monitoring. These findings substantiate prior reports that specific electroencephalogram signatures observed during the first 24 hours after cardiac arrest have strong predictive value despite hypothermia and sedative use (15). Other QEEG features available to our model, but not selected, have been previously reported as useful in HIE prognostication. These include measures of spectral content and burst suppression (5, 13, 16–18, 20, 22–26, 41). This apparent difference from previous literature is likely due to our feature selection method, which when faced with multiple informative but correlated features chooses among them. Thus, features that have significant predictive value individually are not necessarily retained in the final multivariate prognostication model.

Our model demonstrates that electroencephalogram provides valuable prognostic information early after cardiac arrest, and that temporal trends can be used to further improve predictions. The time-sensitive model performance continued to improve as more data became available, highlighting the incremental prognostic value of continuous electroencephalogram beyond 24 hours (5, 15, 18, 19).

In our analysis, the time-sensitive method outperformed a random forest classifier and the original CRI model using five QEEG features (5, 26). We note that the performance of our random forest implementation was inferior to that reported by Tjepkema-Cloostermans et al (5) (AUC, 0.90 at 24 hr). The discrepancy may result from: 1) the heterogeneity of our data, which came from four different centers compared with the authors' two-center study; 2) our decision to align data with respect to the start of electroencephalogram recording rather than the time of cardiac arrest; and 3) use of more extensive validation (10-fold vs one-fold).

This study has several limitations. We have attempted to address each of them at least in part, but further work remains.

1) Our method focused on electroencephalogram and initial clinical features. A multimodal prognostication strategy integrating data from serial neurologic examinations, notably the pupillary reflex and Glasgow Coma Score, somatosensory-evoked potentials, neuroimaging, serum biomarkers,

and visual electroencephalogram review could likely provide more-accurate prognostic predictions. Future studies need to investigate whether QEEG analysis improves on current multimodal prognostication methods.

- 2) Outcome data beyond hospital discharge was missing for some subjects with CPC 3 or 4 at discharge; however, this problem was limited to 1.8% (eight of 438) of subjects. A sensitivity analysis excluding these cases showed no significant change in prediction performance; therefore, we do not believe that loss to follow-up in this small subpopulation significantly affected our results.
- 3) The duration of electroencephalogram monitoring was not uniform across all subjects. This is typical of retrospective studies in this population and is reflected in the decline in prediction calibration after 48 hours. For pragmatic reasons (to facilitate training machine-learning models), we aligned electroencephalogram data based on time of initiation of electroencephalogram rather than time of ROSC. This caveat may decrease the interpretability of our findings regarding how QEEG features change across time from initial injury. However, this data alignment facilitated deployment of a sequential model that requires feeding features forward from one time interval to another, which would be compromised if large sections of data from the first 12 hours were unavailable. A visual representation of the time to ROSC compared with the time of electroencephalogram initiation is shown in **Figure s4** (Supplemental Digital Content 11, <http://links.lww.com/CCM/E675>; and legend, Supplemental Digital Content 12, <http://links.lww.com/CCM/E676>).
- 4) Healthcare providers participating in decision-making regarding life support were not blinded to electroencephalogram results; therefore, we cannot exclude that self-fulfilling prophecies affected the outcomes observed and which features were most predictive in our QEEG model.
- 5) Some drugs administered to cardiac arrest patients are known to modulate the electroencephalogram (e.g., propofol, midazolam). Limited data are available on the specific effects of sedatives on QEEG trends in the cardiac arrest population, and unfortunately hourly sedation information for each patient is not available for the cohort in our study. If we had been able to account for effects of sedatives in training the model, this might have further improved prediction performance. Future studies should analyze and incorporate specific effects of sedation in QEEG trends.
- 6) The performance of our time-sensitive model is modest (AUC, 0.83), and our current modeling approach is not ready for deployment at the bedside. A model that can integrate initial clinical information and continuous electroencephalogram to other data streams such as serial clinical examinations, brain imaging tests, sedation, serum biomarkers, and somatosensory-evoked potentials has the potential to improve prediction performance beyond our current method.

CONCLUSIONS

We used a large, multicenter cohort with HIE to train a sequential prognostication model for good and poor functional outcomes. The QEEG model we developed is time-sensitive,

selecting specific feature values at specific points in time that are most predictive of outcome. The time-sensitive model had better classification and statistical calibration compared with several state-of-the-art baseline approaches. These results demonstrate that the statistical association between quantitative electroencephalogram features and neurologic outcome in HIE changes over time, and accounting for these changes improves performance of predictive models.

REFERENCES

- Go AS, Mozaffarian D, Roger VL, et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee: Heart disease and stroke statistics—2013 update: A report from the American Heart Association. *Circulation* 2013; 127:e6–e245
- Dragancea I, Wise MP, Al-Subaie N, et al; TTM Trial Investigators: Protocol-driven neurological prognostication and withdrawal of life-sustaining therapy after cardiac arrest and targeted temperature management. *Resuscitation* 2017; 117:50–57
- Nielsen N, Wetterslev J, Cronberg T, et al; TTM Trial Investigators: Targeted temperature management at 33°C versus 36°C after cardiac arrest. *N Engl J Med* 2013; 369:2197–2206
- Lemiale V, Dumas F, Mongardon N, et al: Intensive care unit mortality after cardiac arrest: The relative contribution of shock and brain injury in a large cohort. *Intensive Care Med* 2013; 39:1972–1980
- Tjepkema-Cloostermans MC, Hofmeijer J, Beishuizen A, et al: Cerebral recovery index: Reliable help for prediction of neurologic outcome after cardiac arrest. *Crit Care Med* 2017; 45:e789–e797
- van Putten MJ, Tjepkema-Cloostermans MC, Hofmeijer J: Infraslow EEG activity modulates cortical excitability in postanoxic encephalopathy. *J Neurophysiol* 2015; 113:3256–3267
- Amorim E, Rittenberger JC, Zheng JJ, et al; Post Cardiac Arrest Service: Continuous EEG monitoring enhances multimodal outcome prediction in hypoxic-ischemic brain injury. *Resuscitation* 2016; 109:121–126
- Amorim E, Rittenberger JC, Baldwin ME, et al; Post Cardiac Arrest Service: Malignant EEG patterns in cardiac arrest patients treated with targeted temperature management who survive to hospital discharge. *Resuscitation* 2015; 90:127–132
- Rossetti AO, Oddo M, Logroscino G, et al: Prognostication after cardiac arrest and hypothermia: A prospective study. *Ann Neurol* 2010; 67:301–307
- Thenayan EA, Savard M, Sharpe MD, et al: Electroencephalogram for prognosis after cardiac arrest. *J Crit Care* 2010; 25:300–304
- Crepeau AZ, Rabinstein AA, Fugate JE, et al: Continuous EEG in therapeutic hypothermia after cardiac arrest: Prognostic and clinical value. *Neurology* 2013; 80:339–344
- Bagnato S, Boccagni C, Sant'Angelo A, et al: EEG predictors of outcome in patients with disorders of consciousness admitted for intensive rehabilitation. *Clin Neurophysiol* 2015; 126:959–966
- Noirhomme Q, Lehembre R, Lugo Zdel R, et al: Automated analysis of background EEG and reactivity during therapeutic hypothermia in comatose patients after cardiac arrest. *Clin EEG Neurosci* 2014; 45:6–13
- Sivaraju A, Gilmore EJ, Wira CR, et al: Prognostication of post-cardiac arrest coma: Early clinical and electroencephalographic predictors of outcome. *Intensive Care Med* 2015; 41:1264–1272
- Hofmeijer J, Beernink TM, Bosch FH, et al: Early EEG contributes to multimodal outcome prediction of postanoxic coma. *Neurology* 2015; 85:137–143
- Rundgren M, Rosén I, Friberg H: Amplitude-integrated EEG (aEEG) predicts outcome after cardiac arrest and induced hypothermia. *Intensive Care Med* 2006; 32:836–842
- Wennervirta JE, Ermes MJ, Tainien SM, et al: Hypothermia-treated cardiac arrest patients with good neurological outcome differ early in quantitative variables of EEG suppression and epileptiform activity. *Crit Care Med* 2009; 37:2427–2435
- Wiley SL, Razavi B, Krishnamohan P, et al: Quantitative EEG metrics differ between outcome groups and change over the first 72h in comatose cardiac arrest patients. *Neurocrit Care* 2018; 28:51–59
- Zubler F, Steimer A, Kurmann R, et al: EEG synchronization measures are early outcome predictors in comatose patients after cardiac arrest. *Clin Neurophysiol* 2017; 128:635–642
- Oh SH, Park KN, Kim YM, et al: The prognostic value of continuous amplitude-integrated electroencephalogram applied immediately after return of spontaneous circulation in therapeutic hypothermia-treated cardiac arrest patients. *Resuscitation* 2013; 84:200–205
- Beudel M, Tjepkema-Cloostermans MC, Boersma JH, et al: Small-world characteristics of EEG patterns in post-anoxic encephalopathy. *Front Neurol* 2014; 5:97
- Ruijter BJ, van Putten MJ, Hofmeijer J: Generalized epileptiform discharges in postanoxic encephalopathy: Quantitative characterization in relation to outcome. *Epilepsia* 2015; 56:1845–1854
- Elmer J, Gianakas JJ, Rittenberger JC, et al; Pittsburgh Post-Cardiac Arrest Service: Group-based trajectory modeling of suppression ratio after cardiac arrest. *Neurocrit Care* 2016; 25:415–423
- Ghassemi M, Amorim E, Pati SM, et al: An enhanced cerebral recovery index for coma prognostication following cardiac arrest. *Conf Proc IEEE Eng Med Biol Soc* 2015; 2015:534–537
- Forgacs PB, Frey HP, Velazquez A, et al: Dynamic regimes of neocortical activity linked to corticothalamic integrity correlate with outcomes in acute anoxic brain injury after cardiac arrest. *Ann Clin Transl Neurol* 2017; 4:119–129
- Tjepkema-Cloostermans MC, van Meulen FB, Meisma G, et al: A Cerebral Recovery Index (CRI) for early prognosis in patients after cardiac arrest. *Crit Care* 2013; 17:R252
- Safar P: Resuscitation after brain ischemia. In: Brain Failure and Resuscitation. Grenvik A, Safar P (Eds). New York, NY, Churchill Livingstone, 1981, pp 155–184
- Grossestreuer AV, Abella BS, Sheak KR, et al: Inter-rater reliability of post-arrest cerebral performance category (CPC) scores. *Resuscitation* 2016; 109:21–24
- Geocadin RG, Muthuswamy J, Sherman DL, et al: Early electrophysiological and histologic changes after global cerebral ischemia in rats. *Mov Disord* 2000; 15(Suppl 1):14–21
- Shin HC, Tong S, Yamashita S, et al: Quantitative EEG and effect of hypothermia on brain recovery after cardiac arrest. *IEEE Trans Biomed Eng* 2006; 53:1016–1023
- Schoffelen JM, Gross J: Source connectivity analysis with MEG and EEG. *Hum Brain Mapp* 2009; 30:1857–1865
- Gow DW Jr, Segawa JA, Ahlfors SP, et al: Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *Neuroimage* 2008; 43:614–623
- Ioannides AA, Liu LC, Kwapien J, et al: Coupling of regional activations in a human brain during an object and face affect recognition task. *Hum Brain Mapp* 2000; 11:77–92
- Stam CJ, Nolte G, Daffertshofer A: Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp* 2007; 28:1178–1193
- Pritchard WS, Duke DW, Kriebel KK: Dimensional analysis of resting human EEG. II: Surrogate-data testing indicates nonlinearity but not low-dimensional chaos. *Psychophysiology* 1995; 32:486–491
- Pereda E, Gamundi A, Rial R, et al: Non-linear behaviour of human EEG: Fractal exponent versus correlation dimension in awake and sleep stages. *Neurosci Lett* 1998; 250:91–94
- Hjorth B: EEG analysis based on time domain properties. *Electroencephalogr Clin Neurophysiol* 1970; 29:306–310
- Fatma Guler N, Derya Ubeyli E, Guler I: Combined neural network model employing Lyapunov exponents: Internal carotid arterial disorders detection case. *Conf Proc IEEE Eng Med Biol Soc* 2005; 5:4564–4567
- Thakor NV, Tong S: Advances in quantitative electroencephalogram analysis methods. *Annu Rev Biomed Eng* 2004; 6:453–495
- Tibshirani R: Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 1996; 58:267–288
- Ermes M, Sarkela M, van Gils M, et al: Prediction of poor outcome using detector of epileptiform EEG in ICU patients resuscitated after cardiac arrest. Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference 2007, Lyon, France, 2007, pp 3056–3059
- Moshirvaziri H, Ramezan-Arab N, Asgari S: Prediction of the outcome in cardiac arrest patients undergoing hypothermia using EEG wavelet entropy. Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference 2016, Orlando, FL, 2016, pp 3777–3780