doi:10.1093/brain/awx173 BRAIN 2017: 140; 2169–2182 | 2169

BRAIN
A JOURNAL OF NEUROLOGY

The circadian profile of epilepsy improves seizure forecasting

Philippa J. Karoly,[1,2,3] Hoameng Ung,[2] David B. Grayden,[3] Levin Kuhlmann,[1,4] Kent Leyde,[5] Mark J. Cook[1,6] and Dean R. Freestone[1]

It is now established that epilepsy is characterized by periodic dynamics that increase seizure likelihood at certain times of day, and which are highly patient-specific. However, these dynamics are not typically incorporated into seizure prediction algorithms due to the difficulty of estimating patient-specific rhythms from relatively short-term or unreliable data sources. This work outlines a novel framework to develop and assess seizure forecasts, and demonstrates that the predictive power of forecasting models is improved by circadian information. The analyses used long-term, continuous electrocorticography from nine subjects, recorded for an average of 320 days each. We used a large amount of out-of-sample data (a total of 900 days for algorithm training, and 2879 days for testing), enabling the most extensive *post hoc* investigation into seizure forecasting. We compared the results of an electrocorticography-based logistic regression model, a circadian probability, and a combined electrocorticography and circadian model. For all subjects, clinically relevant seizure prediction results were significant, and the addition of circadian information (combined model) maximized performance across a range of outcome measures. These results represent a proof-of-concept for implementing a circadian forecasting framework, and provide insight into new approaches for improving seizure prediction algorithms. The circadian framework adds very little computational complexity to existing prediction algorithms, and can be implemented using current-generation implant devices, or even non-invasively via surface electrodes using a wearable application. The ability to improve seizure prediction algorithms through straightforward, patient-specific modifications provides promise for increased quality of life and improved safety for patients with epilepsy.

1 Department of Medicine, The University of Melbourne, St. Vincent's Hospital, Fitzroy VIC 3065, Australia
2 Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA
3 NeuroEngineering Research Laboratory, Department of Biomedical Engineering, The University of Melbourne, Parkville VIC 3010, Australia
4 Brain Dynamics Lab, Centre for Human Psychopharmacology, Swinburne University of Technology, Hawthorne VIC 3122, Australia
5 Cascade Neuroscience, Seattle, WA 98109, USA
6 Graeme Clark Institute for Biomedical Engineering, The University of Melbourne

Correspondence to: Philippa Karoly
Department of Medicine, The University of Melbourne, St. Vincent's Hospital, Fitzroy VIC 3065, Australia
E-mail: pkaroly@student.unimelb.edu.au

Keywords: epilepsy; seizure prediction; circadian rhythms; forecasting

Abbreviations: AUC = area under the curve; ECoG = electrocorticography

Received January 23, 2017. Revised May 16, 2017. Accepted May 20, 2017

Introduction

The unpredictability of seizures often constitutes the most disabling component of epilepsy, and has a profound impact on safety. Accurate seizure prediction would greatly improve an individuals' quality of life, potentially enabling pre-emptive administration of therapies or allowing steps to ensure personal safety to be undertaken. It is well established that seizures in many patients are preceded by a measurable change in brain state (Litt and Echauz, 2002; Badawy *et al.*, 2009) providing a rationale for designing predictive algorithms. Previous attempts at seizure prediction have shown promise, yet many suffer from poor generalizability, chiefly due to the relatively short duration of available datasets (Mormann *et al.*, 2007). The first human trial for an implantable warning system demonstrated the viability of seizure forecasting in long-term recordings for patients with intractable epilepsy (Cook *et al.*, 2013). The increasing availability of long-term data has inspired a renewed focus on building predictive algorithms that are both patient- and seizure-specific (Freestone *et al.*, 2015; Gadhoumi *et al.*, 2016; Mormann and Andrzejak, 2016), and which incorporate sophisticated neural modelling (Kuhlmann *et al.*, 2015). More recently, engagement with the machine learning community through open-source Kaggle competitions has demonstrated alternative strategies are available to further improve seizure prediction algorithms (Brinkmann *et al.*, 2016; Kaggle.com, 2016). Given the shortcomings in previous studies attempting seizure prediction (Mormann *et al.*, 2007), it is clear that a crucial step towards translating predictive algorithms into clinical devices is a framework to evaluate the prospective online performance in addition to the classification of data segments.

In this work, we propose a probabilistic approach to seizure prediction that incorporates prior knowledge about underlying patterns in seizure occurrence with respect to time of day. There is overwhelming evidence that epilepsy adheres to cyclic patterns that modulate seizures and seizure susceptibility at certain times of day (Bercel, 1964; Shouse *et al.*, 1996; Carney *et al.*, 2011; Loddenkemper *et al.*, 2011; Fernandez *et al.*, 2013). The periods of highest seizure likelihood vary greatly between patients, but on an individual level remain consistent over many years (Karoly *et al.*, 2016). Tracking and utilizing circadian patterns of seizures presents an exciting opportunity to enhance patient management. This straightforward, patient-specific timing information can be used to titrate treatment (Thome-Souza *et al.*, 2016) and improve the performance of seizure advisory systems (Schelter *et al.*, 2006, 2011*a*; Sedigh-Sarvestani *et al.*, 2012; Sedigh-Sarvestani and Gluckman, 2013).

A forecasting approach that expresses the current degree of belief as a likelihood or probability (and incorporates prior information) is grounded in a probabilistic Bayesian epistemology. Traditional assessment metrics for seizure prediction are based on categorical statements—a seizure either will or will not happen—and are inappropriate for assessing probabilistic forecasts. However, in reality, the brain can enter a state of high seizure likelihood that does not necessarily terminate in a clinical seizure (Badawy *et al.*, 2012; Ly *et al.*, 2016), challenging the traditional definition of a false positive prediction. The challenge of assessing probabilistic forecasts was addressed in the meteorological community by Brier (1950), who introduced the Brier score to measure the probability error of weather forecasts. Since Brier's seminal work, meteorologists have spent decades refining attributes of forecasting 'goodness', and developing metrics to measure these different attributes (Murphy, 1973*b*, 1993).

Assessment of seizure prediction can benefit by applying additional probabilistic metrics using techniques from weather forecasting. There are several statistically robust approaches that are used to assess seizure prediction performance compared to chance outcomes (Winterhalder *et al.*, 2003; Kreuz *et al.*, 2004; Snyder *et al.*, 2008; Mader *et al.*, 2014); however, these methods are based on a categorical, rather than probabilistic prediction. Henceforth, the terms 'forecast' and 'prediction' will be used to differentiate between probabilistic and categorical statements about whether a seizure will occur within some future period. Ultimately, the final categorical prediction is the most clinically relevant outcome; however, it may also be possible, and even preferable, to improve prediction algorithms by first optimizing aspects of forecasting performance. The Brier score measures the difference between a continuous, probabilistic forecast and the observed rate of seizures, without requiring an explicit prediction to be made. The Brier score assesses an arbitrarily long sequence of consecutive forecasts over a continuous recording period, providing an excellent framework to assess and compare predictive models without additional tuneable parameters, such as true and false positive rates. Minimizing the number of tuneable parameters reduces the risk of in-sample optimization, thus increasing confidence that observed results will generalize to future data (Andrzejak *et al.*, 2003).

The Brier score has previously been used to evaluate seizure forecasting (Jachan *et al.*, 2009; Schelter *et al.*, 2011*b*); however, it is difficult to implement for rare event forecasting unless large amounts of observation data are available (Murphy, 1973*a*; Murphy and Winkler, 1987). Previous seizure prediction results have been based on short-term (typically 1 week) recordings from patients undergoing presurgical monitoring (Mormann *et al.*, 2007), and this limited time span is insufficient to build patient-specific models of seizure likelihood. Furthermore, there are acute and subacute effects of device implantation and hospitalization (Ghougassian *et al.*, 2004; Polikov *et al.*, 2005; Van Kuyck *et al.*, 2007); so short-term data may not provide a reliable test case scenario for building implantable prediction algorithms.

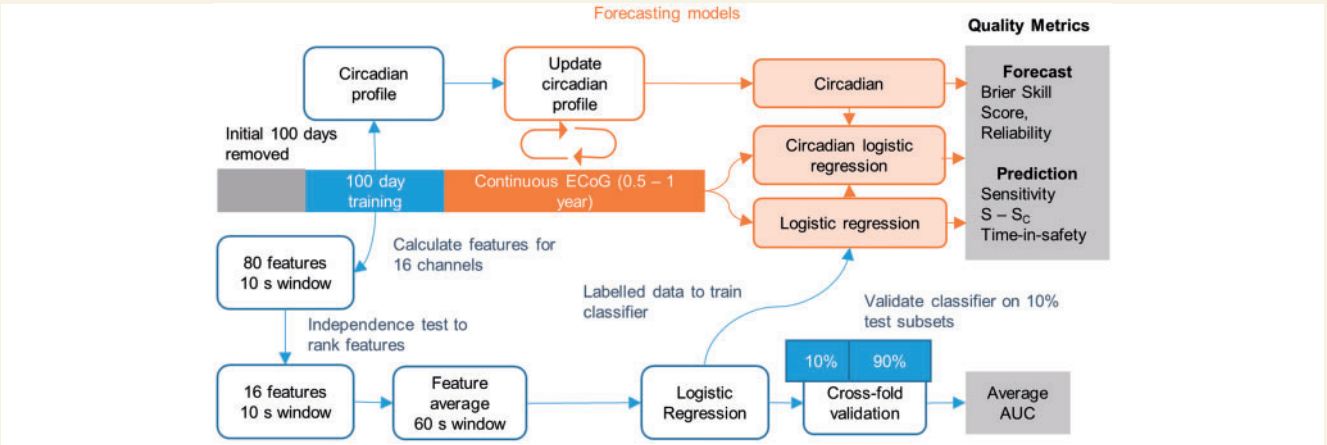

Figure 1 Schematic of the seizure prediction procedure. The first 100 days of data were discarded. Logistic regression classifiers were then trained and validated (10-fold cross-validation) on selected segments of data taken from the subsequent 100 days of recording using the average AUC. The circadian probability was initialized over the 100-day design phase, then used to update the weights of the logistic regression classifiers (combined circadian logistic regression). Continuous forecasts from all three models were evaluated on the remaining data (from Day 200 onwards). Forecast evaluation used the Brier skill score (BSS), reliability curves, sensitivity (S), sensitivity improvement-over-chance ($S-S_C$), and time-in-safety (t_S).

In this work, we used long-term data from a previous study (Cook *et al.*, 2013) (a total of 900 days for algorithm training, and 2879 days for testing) to evaluate patient-specific forecasting models, including the model originally reported in Cook *et al.* (2013). The recording duration provides an excellent opportunity to evaluate probabilistic performance and gain new insights into different aspects of seizure forecasting. The large amount of testing data also provides confidence that the predictive models were not simply optimized for a small number of seizures. The resulting analyses present a proof-of-concept for updating seizure forecasts based on patient-specific circadian information. Forecasts were compared using a range of probabilistic and traditional performance measures. As technology for long-term implantable recording devices becomes readily available, the use of individualized prior information and probabilistic metrics can benefit seizure prediction algorithms.

Materials and methods

The following sections describe the NeuroVista data (Cook *et al.*, 2013) and then outline the steps that were used to build and test forecasting models. To provide a straightforward proof-of-concept, we trained logistic regression classifiers to forecast seizure likelihood, and then evaluated their performance both with and without incorporating prior information based on time of day. All analysis was performed using MATLAB and Statistics Toolbox Release 2016 (The MathWorks, Inc., Massachusetts, USA) on computer clusters based at the Victorian Life Sciences Computation Initiative.

Data

Data for the study were collected from a clinical trial of an implantable seizure warning device (Cook *et al.*, 2013) and accessed from the International Epilepsy Electrophysiology Portal (ieeg.org). All subjects had focal onset seizures, with a seizure onset zone identified from pre-existing medical records and neuroimaging. Intracranial electrode arrays with a total of 16 platinum iridium contacts were implanted around the seizure onset zone. The electrocorticography (ECoG) was sampled at 400 Hz and wirelessly relayed to an external, portable, personal advisory device. Seizure detection was automated using a proprietary detection method. All detections were verified by expert investigators with the aid of audio recording from the handheld device and subjects' seizure diaries. Seizures were classified as being either clinical (type 1) or clinically equivalent (type 2). Type 1 events were associated with clinical symptoms; type 2 events had no verified clinical symptoms but were electroencephalographically indistinguishable from clinical seizures. Based on the similarity of the ECoG, type 2 seizures were considered relevant for developing methods of seizure prediction (Cook *et al.*, 2013), and types 1 and 2 seizures are treated equivalently in this work. Additionally, subclinical (type 3) seizures were detected in the Cook *et al.* (2013) study, but these were excluded from the current analysis. Type 3 events were not clinically manifest and had an electroencephalographic signature that differed from type 1 and 2 events. Prior to the following analyses, the ECoG was filtered between 1 Hz–140 Hz (zero-phase second-order Butterworth bandpass filter).

Figure 1 shows a schematic of the study design. The first 100 days of recording were discounted from the analysis due to disruption of the signal resulting from device implantation (Sillay *et al.*, 2013). The second 100 days of data were used for the algorithm design phase to compute and validate forecasting models. The remaining data (from Day 200 onwards), which ranged from 6 months to >1 year of continuous recording for each subject, were used to evaluate the pseudo-prospective forecasting performance of each model. Algorithm design used lead seizures only (seizures preceded by at least a 5-h seizure-free interval). Subjects with 10 or more lead seizures during the 100-day design phase were selected for the study, resulting in a total of nine subjects. Subjects had an

average of 38 lead seizures in the 100-day training phase, and an average of 116 lead seizures during the remaining evaluation period (all seizure numbers are given in Supplementary Table 1).

Forecasting models

Performance was evaluated for three models: combined circadian logistic regression, logistic regression only, and circadian probability only. Forecasting models output the probability that a seizure would occur within the next 30 min (the pre-ictal period), and made a forecast every 30 s (a 60-s sliding window with 50% overlap). The following sections provide further detail on each model.

Logistic regression

Logistic regression classifiers were trained on select segments of data taken from the 100-day algorithm design phase. Data were chosen from interictal and pre-ictal periods. Pre-ictal periods were defined as being a 30-min window ranging from 31 min to 1 min prior to a lead seizure. Interictal segments were chosen to be 30-min periods at least 6 h clear of a seizure (before or after). The number of interictal segments was matched to the number of pre-ictal segments, giving a balanced training dataset. Balanced training data can lead to the model being biased towards over-identifying pre-ictal segments. The bias can be adjusted by shifting the logistic regression intercept (Bishop, 2006).

A library of 80 signal features was computed for every data segment. Features were similar to the algorithm used in the original NeuroVista clinical trial (Cook *et al.*, 2013). The feature metrics were signal energy in four frequency bands (8–16 Hz, 16–32 Hz, 32–64 Hz, 64–128 Hz) and line length. These five metrics were calculated separately for the 16 electrode channels (5 metrics × 16 channels = 80 features). Features over a 10 s window (50% overlap) were computed from the pre-ictal and interictal segments. The features were then smoothed by taking an average over a 60 s moving window. For each subject the entire set of smoothed and labelled feature vectors were assigned into chronologically ordered training and test subsets (consisting of 90% and 10% of the data, respectively). Division was repeated 10 times (subsets chosen sequentially without replacement) to assess average performance (10-fold cross-validation). For each fold of the validation, a set of 16 features was selected (from original 80 features) by maximizing the relative entropy (also known as the Kullback-Leibler distance) for the training subset. The training feature vectors were then used to fit the weights of the logistic regression function. To validate the classification performance, we calculated the area under the curve (AUC) on the test set for each stage of cross-validation. If validation gave above-chance performance, the final logistic regression classifier was trained using all the pre-ictal and interictal feature vectors obtained from the 100-day training period. The final classifier was evaluated in a pseudo prospective manner using the remaining data (after Day 200).

Circadian probability

A circadian profile was created for each subject based on their seizure times (in 24-h UTC time) during the 100-day algorithm design phase. The probability density function was estimated from the histogram of seizure times using kernel density estimation. Circular Gaussian kernels (von Mises functions, see Supplementary material for further details) were used to represent the time as the variable. The probability density functions were initialized with a uniform (uninformative) prior to avoid any zero weights appearing in the distribution. The probability was then updated every time a seizure occurred, providing a progressive estimation of the circadian profile. The probability functions were created from histograms with a bin width of 1 h, which governed the time sensitivity of the model.

Circadian logistic regression

The logistic regression and circadian models were combined by iteratively updating the weights of the logistic regression classifier. The weight update was based on the subject-specific estimate of the probability of seizure occurrence given the time of day (the derivation of this weight update is provided in the Supplementary material).

Metrics

A range of metrics was used for performance assessment, each of which addressed distinct questions. During algorithm validation, performance was assessed using the area under the receiver-operating characteristic curve (AUC). The AUC addresses the ability of a classifier to discriminate between interictal and pre-ictal data (Brinkmann *et al.*, 2016). Additional measures were used to evaluate the pseudo prospective performance.

The following metrics were used to measure probabilistic forecast quality. (i) Reliability curve: how well do the predicted probabilities of an event correspond to their observed frequencies?; (ii) Brier score: what is the magnitude of the probability forecast errors?; and (iii) Brier skill score: what is the relative skill, or performance, of different probabilistic forecasts?

Investigating more traditional notions of prediction quality required an additional prediction rule for each forecasting model. We used a high and low probability threshold to trigger high and low risk warnings, as outlined by Snyder *et al.* (2008) for the prediction rule. We then applied the same performance metrics of the Cook *et al.* (2013) study, which were: (i) time in safety: what is the maximum amount of time a patient could be assured of low-risk status without a seizure occurring? (ii) prediction sensitivity: how many seizures were correctly identified to occur during high-risk periods? and (iii) sensitivity improvement-over-chance: how valuable was the high-risk warning light, considering the length of time spent in warning?

Brier score

To calculate the Brier score, the forecast of seizure likelihood is first quantized into probability bins (typically 0–10% and so on until 90–100%). The quantization step size reflects meaningful increments for a device and is limited by the number of seizures that occur. After quantization, the forecast is compared to observation data, which are coded into either 0 or 1 (a seizure does or does not occur in the next 30 min pre-ictal period). The quantization approach enables reliability curves to be constructed, as discussed below. A perfect Brier score is 0 (a forecast of 100% for every seizure), and the worst possible score is 1. Essentially the Brier score measures the mean squared error over every forecast; however, a more useful

decomposition is given by:

$$\text{BS} = \text{Reliability} - \text{Resolution} + \text{Uncertainty} \quad (1)$$

Uncertainty accounts for the baseline rate of seizures. The Resolution quantifies the average predictive power above the baseline rate. Reliability describes how close the forecast probability is to the observed data, i.e. the true rate of seizures, given a certain forecast is made. Each term is computed by:

$$\begin{aligned} \text{Reliability} &= \frac{1}{N}\sum_{i=1}^{N_f} n_i(f_i - b_i)^2, \\ \text{Resolution} &= \frac{1}{N}\sum_{i=1}^{N_f}(b_i - b)^2, \\ \text{Uncertainty} &= b(1-b) \end{aligned} \quad (2)$$

Where N is the total number of forecasts (number of 60-s windows), N_f is the number of forecast bins (10 was used), b is the baseline rate of seizures (Supplementary Table 1), b_i is the actual seizure occurrence rate when the forecast was in the i^{th} bin, and f_i is the average forecast for the i^{th} bin, and n_i is the number of forecasts made within each bin.

Brier skill score

It is difficult to use the Brier score to compare different forecasts if the data have a very low baseline rate of events (low uncertainty), because simply forecasting lower probabilities can greatly improve the Brier score. It has been shown that the naïve approach of always forecasting a constant, small probability (the baseline rate of seizures) gives impressive Brier scores due to the rarity of seizures (Schelter *et al.*, 2011*b*). For this reason, we used the Brier skill score to provide a relative measure for performance comparison. The Brier skill score is computed as

$$BSS = 1 - \frac{BS}{BS_{ref}}, \quad (3)$$

where BS and BS_{ref} are the Brier scores for a given forecast and some reference forecast. The Brier skill score measures improvement over a reference (where 1 is perfect, 0 shows no improvement, and negative values indicate worse performance than the reference).

Weather models typically use either the constant baseline, or the historical climatological forecast as a reference. However, here we want to explicitly evaluate the forecast based on an historical record of seizure times. Therefore, we used the Brier score derived from surrogate forecasts as a reference. Surrogate forecasts were constructed for each model by randomly drawing probabilities from the same distribution as the actual forecast made by that model. In this way, 1000 surrogate forecasts were generated to find the mean Brier skill score for each model (combined circadian logistic regression, logistic regression only, and circadian only). The use of forecast surrogates also handles the difficulty of comparing forecasts to the constant baseline model, as any constant forecast has a Brier skill score of zero.

Reliability curves

The reliability curve is a useful visualization tool for the Brier score components, showing a plot of the forecast seizure rate versus the actual seizure occurrence rate. Actual seizure rate was determined by how often a seizure occurred in the preictal period following every forecast. An ideal reliability curve is the diagonal line where forecast probabilities are equal to the actual outcomes.

Seizure prediction

To validate the utility of circadian-weighted forecast in a predictive setting, we also evaluated performance using the same metrics that were used for the NeuroVista device trial. To calculate these statistics, it was necessary to set an upper (lower) probability threshold to trigger a high (low) seizure likelihood advisory period. We used the same threshold trigger scheme described by Cook *et al.* (2013) and based on the process outlined by Snyder *et al.* (2008). For the range of possible high risk thresholds, we calculated the amount of time spent in warning, t_w, as well as the sensitivity, $S(t_w)$, and sensitivity improvement-over-chance, $S(t_w) - S_C(t_w)$, where $S_C(t_w)$ is the sensitivity of a time-matched chance indicator. These assessment metrics can be mathematically related to the seizure prediction characteristic outlined by Winterhalder *et al.* (2003). Sensitivity improvement-over-chance compares the accuracy of a predictor to the performance of a time-matched chance prediction (based on a Poisson process), thereby penalizing methods where the seizure warning light is on for a high percentage of time. The *P*-values for the hypothesis that 'the sensitivity is significantly better than chance performance', were also calculated for the sensitivity improvement-over-chance metric (see Supplementary material for further details).

For the low-risk threshold, we tuned the prediction horizon (between 30 and 60 min) and threshold so that no seizures would occur within the resultant safety advisory period (measured during the 100-day design phase). We then used the remaining data to evaluate the time spent in safety, t_s, and the number of seizures that occurred during the safety advisory.

Note that prediction sensitivity was calculated for lead seizures only (defined as seizures with a preceding seizure-free interval of at least 5 h). All seizures were used to evaluate the rate of false negatives.

Results

In the following sections, we include previous results from the Cook *et al.* (2013) study as a baseline. However, it was not possible to make a direct comparison. The previous trial reported on a 4-month prospective evaluation period following a training period, yet recording continued for months to years beyond the initial evaluation phase. Therefore, the current results are based on new data.

Circadian profiles

Figure 2 shows raster plots of every subject's seizures with respect to time of day, as well as the circadian probability

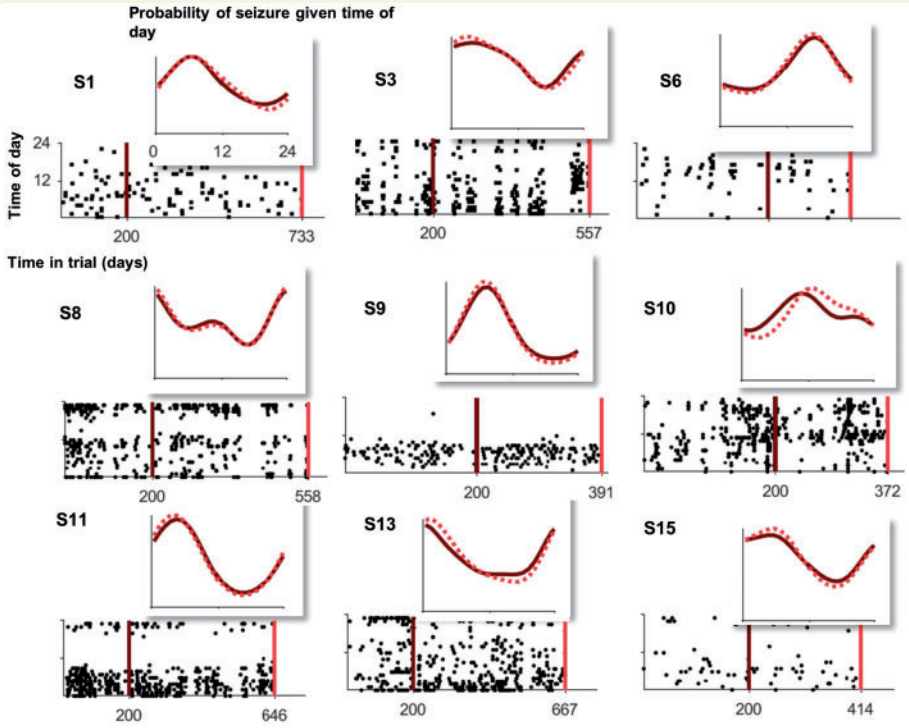

Figure 2 Circadian patterns of seizures. Each subplot contains a raster showing the hour (*y*-axis) of each seizure recorded over the entire trial (*x*-axis). The line graph (inset) shows the estimated probability density after training (200 days, solid line), compared with the end of the trial (dashed red line). Note that the trial period varied between 1 and 2 years for different subjects. Note that times here are reported in UTC, and do not reflect the correct time zone.

distributions constructed from the seizure times. The circadian distribution (probability of seizure with respect to time of day) was initialized after the 100-day training period and then updated every time a seizure was observed. Figure 2 shows the initial (after training) and final (at the completion of the trial) estimates of the circadian probability. Although we updated the probability with every seizure for most subjects, the training period was sufficient to obtain an excellent approximation of the circadian profile. The stability of the seizure probability curves demonstrates that reliable prior information can be obtained after a relatively short training period.

Classifier training and validation

The logistic regression classifier was trained over 100 days of recording. To ensure the model provided reasonable accuracy we measured the average AUC after 10-fold cross-validation (90% training, 10% test, segments chosen sequentially without replacement). The AUC provides a measure of sensitivity and specificity, and has been used to assess seizure prediction competitions (Brinkmann *et al.*, 2016; Kaggle.com, 2016). All subjects had AUC performance above chance (note that the AUC for chance performance is 0.5, and a perfect score is 1). The average AUC across the nine subjects was 0.79, and Subjects 11 and 15 had impressive results of 0.90 (all AUC results are shown in Supplementary Table 2). Reasonable baseline performance with logistic regression classifiers was necessary to demonstrate that improvements using time of day were not trivial. We also noted that the final features selected for the classifiers were relatively stable across the 10 folds of cross-validation, although the most important features varied between subjects (Supplementary Figs 1 and 2).

Seizure forecasting

In this section, we extend our analysis to evaluating the forecasts and predictions in the pseudo prospective case using the entire remaining day after Day 200. Figure 3 shows an example output of the two forecasting models over a 72-h period. Forecasts were made every 30 s (black dots), and an example prediction threshold is shown at the 95th percentile for each model (red dots); although during later prediction analyses all thresholds were evaluated. Seizures are marked by red vertical lines. Note that the combined circadian logistic regression model (Fig. 3B) can reduce the number of erroneous threshold crossings.

Forecasting quality

The Brier skill score (mean and standard error) calculated from every forecast made by the four different models is given in Fig. 4. The maximum possible score is 1. For all

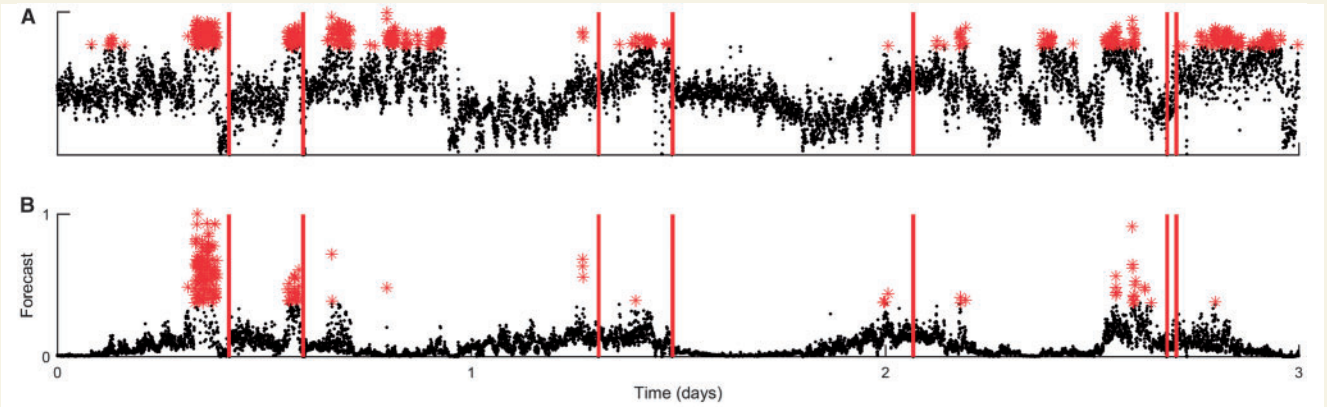

Figure 3 Example forecast with high-alert threshold crossings. The predicted seizure probability from two forecasting models for Subject 3. Forecasts marked in red were above a prediction threshold (set to the 95th percentile of the forecasts) that initiates the high-risk advisory light. Red vertical line indicates seizure onset. (**A**) Logistic regression model only and (**B**) combined circadian logistic regression model.

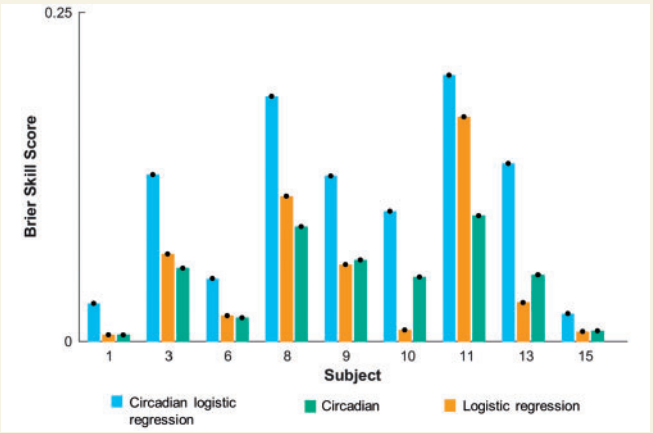

Figure 4 Brier skill score for different forecasts. Each forecast was a vector of probabilities that were made in 30-s intervals (60 s forecasting window with 50% overlap). The skill score shows the average improvement compared to 1000 surrogate forecasts. The standard error of the mean with a 95% confidence interval is shown as a black dot above each bar. Exact values for the mean and standard deviation of the scores are given in the Supplementary material.

subjects the combined circadian logistic regression showed the most improvement compared to surrogate data, and significantly outperformed all other models. All forecasting models performed better than the naïve constant baseline forecast, which would give a Brier skill score of 0.

Figure 5 shows the reliability curves of each forecast. A perfectly calibrated forecast is a diagonal line (where forecast probability is equal to actual probability). Also shown in the figure is the line of no-reliability (equivalent to the calibration levels of a constant baseline forecast), and the line of no-skill (the point where the Brier skill score is not higher than the constant baseline forecast). Forecasts above the no-reliability (no-skill) line show improved calibration (skill) compared to the constant baseline forecast. Figure 5 also shows forecast histograms, which are used as a measure of sharpness (how many forecasts are made at different levels of seizure likelihood). Note that the circadian only forecast makes no predictions for high probability values. This reflects the fact that a cyclic forecast is not capable of providing good calibration for rare events, and the intuition that forecasts need to outperform a repetitive model to be clinically useful.

We begin by addressing the potential concern that forecasts were predominately below the no-skill line. This result does not demonstrate that the forecasting models are poor. In fact, the no-skill line highlights the difficulty of evaluating forecasts for very rare events, since the baseline constant forecast (a model that always predicts a very low chance of seizure) is a close match to reality, and provides reasonable forecasting skill despite having no practical utility for patients (Schelter *et al.*, 2011*b*). Note that the Brier skill score addressed this challenge, showing improvements above the constant forecast for all subjects, especially for the combined circadian logistic regression model (Fig. 4).

Figure 5 shows that most subjects outperformed a constant forecast in terms of reliability (exceptions were Subjects 1, 6 and 15). Subject 3 shows high forecasting skill, and Subjects 9, 10, and 13 show some skill within the high likelihood of seizure regime. Furthermore, the combined circadian logistic regression model (Fig. 5, blue line) has superior calibration (closer to diagonal) for all subjects, except Subjects 1, 6, and 15.

It can be seen from the inset histograms that the combined circadian logistic regression forecasts were more heavily skewed towards low probability of seizure. We also note that the very poor performers (Subjects 1, 6 and 15) all had logistic regression models that were skewed towards higher probability of seizures (inset histograms, Fig. 5), indicating that performance was compromised by false positives. It is important to make note of several points about these subjects. Subject 1 had

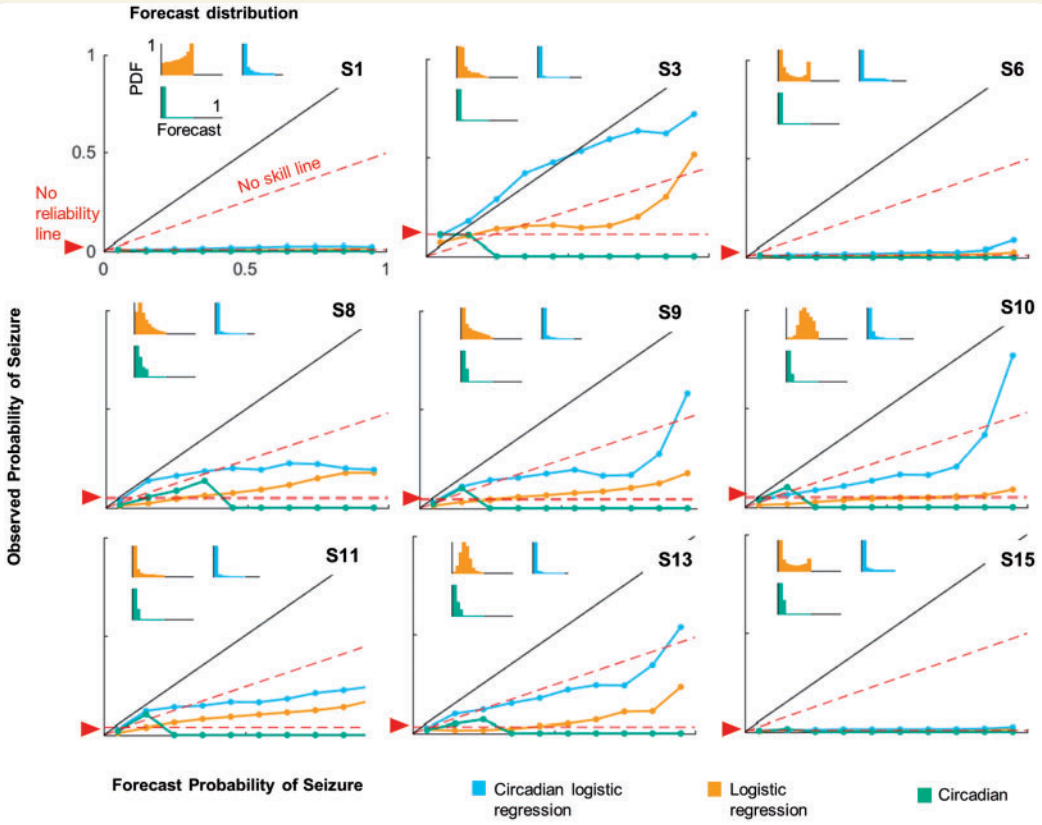

Figure 5 Reliability curves for three different forecasts. Subpanels show the forecast versus actual seizure probability for each subject, based on three different forecasting models. *Insets* show the quantized forecast histograms for each model (the probability that forecasts were between a given interval of seizure likelihood, from 0–10% to 90–100%). The line of no reliability is the same as the baseline rate of seizures, and reflects the fact that a constant forecast has no calibration (slope). The line of no skill reflects the point where a forecast has a Brier skill score of zero compared to the constant baseline forecast. A perfect forecast would fall on the diagonal line.

a highly unusual medication regime. During the initial 4-month algorithm design phase (during which no warning lights were enabled), the subject experienced an average of 10 seizures per month. Following this period, the algorithm and warning lights were enabled. During the subsequent 20-month period, in accordance with the investigators directions and in addition to their normal medications, the subject took medications in response to activation of the warning light. During this recording period, the subject experienced significantly less seizures, with an average of 5.2 seizures per month ($P < 0.01$ using a Wilcoxon rank sum test). Therefore, for Subject 1, results may be confounded by false positives (seizures that were forestalled by medication). Subject 6 was not included in the Cook *et al.* (2013) trial due to insufficient performance during algorithm cross-validation. Subject 15 had low seizure numbers and spent a relatively high proportion of time in warning, in both the current and previous study.

Prediction quality

It is worth noting that the following prediction results can be interpreted in a way that is analogous to a traditional confusion matrix. We report on true positive rates (prediction sensitivity), and false negative rates (number of seizures that occur during safety advisory periods). The times in warning and safety advisory periods are reported *in lieu* of the false positive rate and true negative rate, as rates are considered less useful than time in warning for evaluating seizure prediction devices (Mormann *et al.*, 2007; Gadhoumi *et al.*, 2016; Freestone *et al.*, 2017).

The Cook *et al.* (2013) study measured the amount of time subjects spent in a high alert phase and corresponding prediction sensitivity In the current work, we matched the time in warning to the original Cook *et al.* (2013), then measured the corresponding sensitivity (see Supplementary material for exact corresponding time in warning of each forecasting model). Table 1 shows both the results from the Cook *et al.* (2013) study, along with the sensitivity for the combined circadian logistic regression, logistic regression only, and circadian only forecasts.

It is important to point out that it is not possible to make a direct comparison to the Cook *et al.* (2013) results, as current results were based on a different test period, incorporating more seizures for each subject (Table 1). However, the sensitivity of the new models was equal to or greater

Table 1 Prediction performance and results

Subject	Time in high risk (%)	Cook *et al.* (2013) results		New results (sensitivity)			
		Total seizures	Sensitivity	Lead seizures	Circadian	Logistic regression	Circadian logistic regression
1	27	13	**0.77**	56	0.34	0.54	0.61
3	29	106	0.45	129	0.36	0.53	**0.55**
6	NA	NA	NA	21	0.52	0.61	**0.65**
8	28	86	0.62	177	0.58	0.71	**0.76**
9	11	52	0.17	102	0.28	0.29	**0.45**
10	17	164	0.51	96	0.36	0.38	**0.52**
11	15	39	0.39	186	0.43	0.57	**0.58**
13	28	113	0.50	242	0.61	0.78	**0.76**
15	41	24	**0.71**	36	**0.71**	0.51	0.60

The number of seizures and trial results during the 4-month assessment phase of the original clinical trial. The number of seizures used for assessment in the current study (lead seizures defined as having a preceding seizure free interval of at least 5 h). The new results for the three forecasting models are reported based on a matched proportion of time in warning (see Supplementary material for exact time in warning data). Sensitivity was calculated according to Cook *et al.* (2013), using lead seizures only. The highest sensitivity for each subject is highlighted in bold.

Table 2 Time in low risk (%) for each forecasting model

Patient	Circadian logistic regression	Circadian	Logistic regression	Cook *et al.* (2013)
1	0	0	0	7
6	24	0	0	NA
8	11	19	0	NA
9	30	31	19	48
10[a]	17	11	10	NA
11	34	30	0	26
15	38	30	30	NA

Low risk activation was modelled using the triggering scheme described in Cook *et al.* (2013). Zero indicates a threshold could not be found based on the design phase data (Day 100 – Day 200). NA indicates that the low risk advisory state was not enabled in the NeuroVista trial. Note that no threshold could be found for Subjects 3 and 13 in either the current or previous study.

[a]Subject 10 had one seizure during their low-risk advisory test period (after Day 200). All other subjects had no seizures during their low-risk advisory test period.

than the original trial for all subjects, except Subject 1. For eight of nine subjects, the combined circadian logistic regression model showed the highest sensitivity, indicating that prior information based on time of day improved performance.

Table 2 shows the time spent in the low-risk advisory phase (note that no seizures occurred during low-risk advisory phase unless indicated by an asterisk). For all subjects, the circadian only or combined circadian logistic regression model provided the most time in the low-risk phase. In the Cook *et al.* (2013) study, only three of the subjects had this feature enabled. Therefore, it can be concluded that using seizure timing provides new insight into times of safety.

To establish improvement using circadian weighting for a range of operating points, we plotted the true positive rate against the time spent in warning, t_w (Fig. 6). The range of time in warning was obtained by exploring all possible thresholds. In this way, the plotted curves are similar to a receiver operating characteristic curve; however, the time in warning is plotted rather than the false positive rate

We tested the prediction sensitivity at the maximum sensitivity above chance ($S - S_C$ for each subject labelled in Fig. 6). These true positive rates ranged from 49% to 91%, demonstrating good prospective performance outcomes for most subjects. The maximum sensitivity was obtained by the combined circadian logistic regression models for all subjects except Subjects 3 and 6 (where the logistic regression model had the best performance). Furthermore, for most subjects, the combined circadian logistic regression model has superior performance across the entire operating spectrum of a prospective device (highest curves in Fig. 6). Exceptions are Subject 13, where the combined circadian logistic regression has the best performance only at the maxima rather than over the entire range.

Discussion

The aim of this work was to test the predictive benefits of including patient-specific circadian information in a forecasting model for seizures. For most subjects, circadian patterns were pronounced and consistent, although there was significant variation between individuals in the shape of the circadian distribution (as shown in Fig. 2). We presented a framework that enables arbitrary (patient-specific) circadian patterns to be combined with any form of probabilistic prediction, and robustly demonstrated predictive improvement for the case of a logistic regression classifier. Including circadian information resulted in superior forecast and prediction quality compared to purely EEG-based logistic regression, and when compared to the results from the Cook *et al.* (2013) study. All subjects demonstrated clinical prediction performance significantly better than a chance Poisson prediction (Fig. 6). Furthermore, the time-matched sensitivity results (Table 1) demonstrate that the signal features used in the original trial can provide comparable

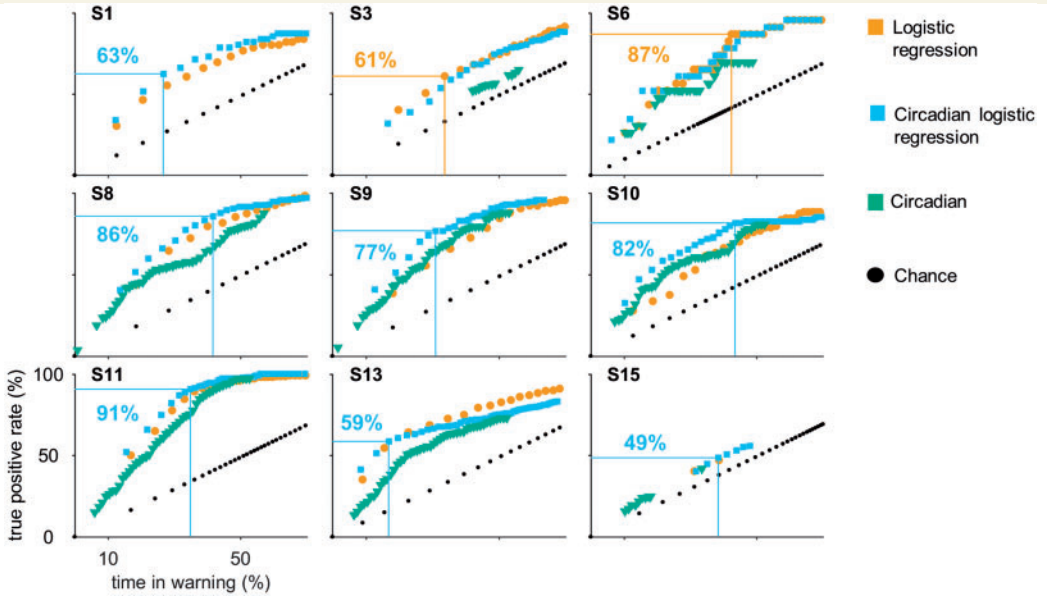

Figure 6 True positive rate of seizure prediction for a range of times spent in warning. Panels show the plots for nine subjects across a range of possible time in warning, t_W (*x*-axis). The time in warning was calculated by exploring all possible prediction thresholds. The prediction sensitivity (true positive rate) at the maximum sensitivity above chance ($S - S_C$) is labelled for each subject (and colour-coded according to the model that achieved maximal improvement above chance). Data are only plotted where the sensitivity improvement-over-chance obtained significance ($P < 0.05$).

prediction performance over many years (whereas previously reported results were based on a 4-month period).

For all patients, the combined circadian logistic regression showed superior performance to the logistic regression only and circadian only models in terms of probabilistic measures (Brier skill score, Fig. 4; calibration, Fig. 5). The combined circadian logistic regression model also provided the best predictive performance for a majority of subjects in a realistic clinical use scenario (Table 1 and Fig. 6). Overall, we conclude that using circadian information improves forecasting accuracy regardless of the precise shape of the circadian distribution, which is highly patient-specific (Fig. 2). Furthermore, circadian information provides additional benefits in terms of informing patients of low-risk periods (Table 2). To fully investigate the benefits of using circadian information for low-risk periods it is necessary to further investigate the interaction between the high-risk and low-risk alert systems, and the relationship between a low-risk advisory and times of sleep.

The sensitivity performance at the greatest improvement above chance ranged from 49% to 91% (Fig. 6), demonstrating excellent performance for some subjects. At the original operating points (based on time in warning), the prediction sensitivity ranged from 45% to 76%, higher than compared to previous prediction algorithms on the same dataset (Cook *et al.*, 2013). The low-risk advisory phase was enabled for more patients, and for longer periods with the inclusion of circadian patterns (Table 2). Crucially, only one subject had one seizure during their pseudo prospective evaluation of the low-risk advisory phase. This result is highly meaningful to patients, as knowing times of safety may sometimes be even more valuable than knowing times of seizure risk. For a majority of subjects, the combined circadian logistic regression models had the best prediction sensitivity across all operating thresholds of a prospective implant device (Fig. 6). This is an important result, as patients have different requirements for device specificity, related to the amount of time that they are prepared to spend in warning. These results show that regardless of how the forecasting model was implemented, time-of-day information improved prediction sensitivity.

Validation

Improvement using time-of-day information was shown here for a predictive model that used relatively simple ECoG features. However, we hypothesize that more sophisticated classification algorithms will also reap the benefits of using an informative prior. In support of this hypothesis, we note that the logistic regression classifiers showed excellent classification performance based on cross-validation results (average AUC, see Supplementary Table 2). The only performance benchmark for classification of long-term seizure data is from recent Kaggle competitions using canine data (Brinkmann *et al.*, 2016) and human data (Kaggle.com, 2016), where the winning algorithms achieved AUC results of 0.84 and 0.81, respectively. The AUC performance of the current logistic regression classifiers ranged from 0.69 to 0.9. While results are not directly comparable, reaching similar values as state-of-the-art machine learning algorithms provides some assurance that

circadian information was not merely beneficial because the original classifiers were weak.

Despite impressive validation results, AUC was not a reliable indication of clinical performance. For instance, Subjects 9 and 15 had excellent AUC results (0.75 and 0.90, respectively), yet had the worst sensitivity following pseudo prospective evaluation (Table 1). A potential pitfall of the AUC as a performance metric is that it is typically derived from balanced data (50% interictal and 50% preictal). However, during continuous forecasting the data are highly skewed towards interictal segments. Therefore, forecast specificity may be overestimated, as seen with the reliability curves (Fig. 5), which all demonstrated overconfidence. We also noted from the reliability curves that different attributes of poor prediction performance (affecting Subjects 1, 6 and 15) were all unambiguously detected by the reliability curves, but not necessarily by AUC results, or sensitivity analysis (Table 1). These mismatched performance scores highlight the necessity of evaluating predictive algorithms on continuous data, and underscore the relevance of using probabilistic methods of assessment. Furthermore, to simplify the comparison using the Brier skill score, we formulated the problem as a binary classification task; however, it may be more appropriate to develop a model that can be regressed on the time since last seizure. An avenue of future investigation is to devise more suitable training data or classifiers for rare events to address the skewed probability of seizures. The trade-off in constructing training data this way is that it becomes necessary to address other design questions, such as the relative importance of false positives and false negatives (Bishop, 2006). With more tuneable design parameters, in-sample optimization can become problematic (Andrzejak *et al.*, 2003; Kreuz *et al.*, 2004).

Forecasting

In all cases where seizure forecasts did provide a reasonable level of accuracy, they were consistently overconfident. We speculate that, in some cases, this overconfidence is due to homeostatic mechanisms or exogenous environmental adjustments that could forestall an imminent seizure. For instance, during data collection, subjects were using medication that may have affected seizure rates. Subject 1 (who had poor performance), selectively used medication during the Cook *et al.* (2013) trial, increasing their usage at times when the high-risk advisory light was turned on. It is highly likely that these behavioural modifications contributed to a high rate of false positives, and poor performance following the initial 4-month assessment phase. Another possibility is that false positives arise because of subclinical events or other epileptic activity, such as spike-wave discharges (Cook *et al.*, 2016; Karoly *et al.*, 2016). In a system where homeostatic mechanisms may correct abnormal activity, it is possible to enter a state of high seizure likelihood without a corresponding seizure (Badawy *et al.*, 2012; Ly *et al.*, 2016). Therefore, it is important to take a more nuanced approach to reducing false positives in seizure prediction, for instance by using algorithms that can distinguish between different classes of epileptic activity. Forecasting models that are trained with the entire spectrum of epileptiform activity may extract additional information relevant to detecting periods of high seizure likelihood. This approach would provide a large volume of training data, while also being intuitively reasonable for a homeostatic system like the brain, where there are numerous 'necessary but not sufficient' conditions for seizure.

The use of forecasting metrics can provide useful diagnostic information before designing clinical prediction rules (i.e. implementing seizure warning lights). For instance, a clear message from the reliability plots (Fig. 6) was that predictive power could be primarily attributed to forecast calibration rather than skill. Therefore, we conclude that improving skill is a more promising avenue to advancing the field of seizure prediction. The use of an informative prior based on time of day increased the forecast calibration. However, improving the skill requires a more precise input than coarse-grained probabilistic information. Increased skill should be derived from improved extraction of information from the ECoG (or brain) itself, rather than lower resolution, typically cyclic meta-data. Therefore, to improve forecasting skill, we speculate that the search for better pre-ictal features from the EEG signal is important; and, given the variability in patient skill levels, this search is likely to require patient-specific features. The reported prediction results were based on commonly used line-length and energy features for all patients; however, the performance and stability of these features was patient-specific (Supplementary Figs 1 and 2). In future, more complex features could be investigated to increase forecasting skill. We were also interested to note from Fig. 3 that seizures appeared to be preceded by a peak in seizure likelihood that was followed by a steady decrease. It is possible that the brain goes through a series of state changes prior to seizures, which were not adequately captured by features within a 60-s window. Including circadian information in our algorithm provided some ability to use patterns over longer time scales. However, it may also be beneficial to calculate signal features over a longer time scale.

Reliability curves can also be used to design warning light thresholds. The key features of these curves are their range and monotonic increase. Monotonicity enables variation in the range of forecasts to be mapped to some patient-specific level of seizure likelihood (i.e. for a device that provides a graded warning of low, medium, and high risk of seizure). Better calibration suggests that a probabilistic model has a greater chance of being converted into a useful prediction rule, with patient-specific thresholds for high and low risk levels of seizure. Therefore, based on their superior calibration, the combined circadian logistic regression model should result in a more clinically useful seizure warning system. Probabilistic evaluation results were consistent with pseudo prospective prediction outcomes.

Future work

Machine learning techniques, such as deep convolutional neural networks, have made impressive inroads with previously intractable classification problems (LeCun *et al.*, 2015), such as image recognition (Krizhevsky *et al.*, 2012), speech processing (Graves *et al.*, 2013), and decoding motor signals from EEG (Nurse *et al.*, 2016). However, these techniques are still in their infancy with respect to integrating temporal information across many orders of magnitude. For instance, in seizure prediction, signal features from 24 h ago may be as relevant as those from 1 min ago. Neural networks are not yet well equipped to deal with this problem. In contrast, weather forecasters have spent decades developing techniques to incorporate multiscale spatiotemporal information into models (Richardson, 2007). An appealing approach for seizure prediction is to unite these fields, for instance using deep learning for feature extraction, while conditioning a neural network with temporal information.

While advanced machine learning techniques have shown promise (Brinkmann *et al.*, 2016), algorithms based on simpler features have the distinct advantages of being easy to iteratively train, incorporate up-to-date prior information, and interpret. In this work, we have demonstrated that time-of-day information can be used to improve forecast calibration and performance. However, there are no limitations on the sources of prior information that can be combined to improve performance (Bishop, 2006; Satopaa *et al.*, 2014). In future, predictive inputs should include additional biometrics and statistics that may be relevant to seizure prediction, such as heart rate (Valderrama *et al.*, 2010; Fujiwara *et al.*, 2016), interictal spike rate (Li *et al.*, 2013), and other temporal information (day of week, month, etc.) (Cook *et al.*, 2014; Karoly *et al.*, 2016). Another key factor in improving forecasting accuracy is the ability to make regular predictions to update the model based on previous performance (Kalman, 1960; Schiff, 2012; Mellers *et al.*, 2015). A promising avenue for seizure prediction is to begin implementing rudimentary warning systems for minimally invasive devices, where algorithms are reliant on simple, linear features and measurable environmental factors, such as time of day.

Conclusion

We have outlined a probabilistic framework for developing and improving patient-specific seizure forecasting models using circadian patterns of seizures. The most promising step towards making probabilistic seizure forecasting a clinically relevant process is the development of implantable devices that continuously record and store neural data. These devices will enable prediction algorithms to be rapidly developed, tested, and calibrated on an individualized basis. We have made many analogies between weather forecasting and seizure prediction, though the dynamics of brain activity and epileptic processes are not directly comparable to the earth and meteorology. Nevertheless, the brain, like the earth, is an immensely complex nonlinear system; and, like the weather, epileptic dynamics exhibit consistent cyclic patterns. We have presented a method for exploiting these dynamics to improve the forecasting accuracy using a weighting based on each patient's seizure history. The simplicity of this method is its greatest advantage; as it requires almost no additional computation (simply recording seizure times), and can be implemented for an arbitrary prior distribution with any other probabilistic measure of seizure likelihood. This framework shows promise for improving the quality of life and safety for patients with epilepsy.

Acknowledgements

The authors acknowledge the support and contribution of colleagues at the Brain Dynamics and Neural Engineering Group at The University of Melbourne, and the Center for NeuroEngineering and Therapeutics at the University of Pennsylvania. PJ.K also acknowledges the support of the Kenneth Myer foundation, Australia and Pro Medicus Ltd.

Funding

This project was funded by the National Health and Medical Research Council, Australia (NHMRC Project APP1065638). Research was supported by the Victorian Life Sciences Computation Initiative (VLSCI), an initiative of the Victorian Government, Australia, on its facility hosted at the University of Melbourne, grant number VR0003. The International Epilepsy Electrophysiology Portal is funded by the National Institute of Heath, United States (NIH U24NS063930).

Supplementary material

Supplementary material is available at *Brain* online.

References

Andrzejak RG, Mormann F, Kreuz T, Rieke C, Kraskov A, Elger CE, et al. Testing the null hypothesis of the nonexistence of a preseizure state. Phys Rev E 2003; 67: 010901.

Badawy R, Macdonell R, Jackson G, Berkovic S. The peri-ictal state: cortical excitability changes within 24 h of a seizure. Brain 2009; 132: 1013–21.

Badawy RAB, Freestone DR, Lai A, Cook MJ. Epilepsy: ever-changing states of cortical excitability. Neuroscience 2012; 222: 89–99.

Bercel NA. The periodic features of some seizure states. Ann N Y Acad Sci 1964; 117: 555–62.

Bishop CM. Pattern recognition. Mach Learn 2006; 128.

Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev 1950; 78: 1–3.

Brinkmann BH, Wagenaar J, Abbot D, Adkins P, Bosshard SC, Chen M, et al. Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. Brain 2016; 139: 1713–22.

Carney PR, Talathi SS, Hwang DU, Ditto W. Circadian Regulation of Neural Excitability in Temporal Lobe Epilepsy. In: Osorio I, Zaveri HP, Frei MG, Arthurs S, editors. Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics. Boca Raton, FL, USA: CRC Press; 2011. p. 289–306.

Cook MJ, Karoly PJ, Freestone DR, Himes D, Leyde K, Berkovic S, et al. Human focal seizures are characterized by populations of fixed duration and interval. Epilepsia 2016; 57: 359–68.

Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. Lancet Neurol 2013; 12: 563–71.

Cook MJ, Varsavsky A, Himes D, Leyde K, Berkovic SF, O'Brien T, et al. Long memory processes are revealed in the dynamics of the epileptic brain. Epilepsy 2014; 5: 168.

Fernandez IS, Ramgopal S, Powell C, Gregas M, Zarowski M, Shah A, et al. Clinical evolution of seizures: distribution across time of day and sleep/wakefulness cycle. J Neurol 2013; 260: 549–57.

Freestone DR, Karoly PJ, Cook MJ. A forward-looking review of seizure prediction. Curr Opin Neurol 2017; 30: 167–73.

Freestone DR, Karoly PJ, Peterson ADH, Kuhlmann L, Lai A, Goodarzy F, et al. Seizure prediction: science fiction or soon to become reality? Curr Neurol Neurosci Rep 2015; 15: 73.

Fujiwara K, Miyajima M, Yamakawa T, Abe E, Suzuki Y, Sawada Y, et al. Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features. IEEE Trans Biomed Eng 2016; 63: 1321–32.

Gadhoumi K, Lina JM, Mormann F, Gotman J. Seizure prediction for therapeutic devices: a review. J Neurosci Methods 2016; 260: 270–82.

Ghougassian DF, d'Souza W, Cook MJ, O'Brien TJ. Evaluating the utility of inpatient video-EEG monitoring. Epilepsia 2004; 45: 928–32.

Graves A, Mohamed A-r, Hinton G. Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on acoustics, speech and signal processing, 2013. Vancouver, BC, Canada: IEEE. pp. 6645–9.

Jachan M, Feldwisch-Drentrup H, Posdziech F, Brandt A, Altenmüller DM, Schulze-Bonhage A, et al. Probabilistic forecasts of epileptic seizures and evaluation by the Brier score. In: 4th European Conference of the International Federation for Medical and Biological Engineering, 2009. pp. 1701–5.

Jachan M, et al. (2009) Probabilistic Forecasts of Epileptic Seizures and Evaluation by the Brier Score. In: Vander Sloten J, Verdonck P, Nyssen M, Haueisen J, editors. 4th European Conference of the International Federation for Medical and Biological Engineering. IFMBE Proceedings, vol 22. Berlin, Heidelberg: Springer.

Kaggle.com. Melbourne University/AES/MathWorks/NIH Seizure Prediction Kaggle Competition. 2016. https://www.kaggle.com/c/melbourne-university-seizure-prediction.

Kalman RE. A new approach to linear filtering and prediction problems. J Basic Eng 1960; 82: 35–45.

Karoly PJ, Freestone DR, Boston R, Grayden DB, Himes D, Leyde K, et al. Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. Brain 2016; 139 (Pt 4): 1066–78.

Kreuz T, Andrzejak RG, Mormann F, Kraskov A, Stögbauer H, Elger CE, et al. Measure profile surrogates: a method to validate the performance of epileptic seizure prediction algorithms. Phys Rev E 2004; 69: 061915.

Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proceedings of the Advances in Neural Information Processing Systems, 2012. pp. 1097–105.

Kuhlmann L, Grayden DB, Wendling F, Schiff SJ. Role of multiple-scale modeling of epilepsy in seizure forecasting. J Clin Neurophysiol 2015; 32: 220–6.

LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436–44.

Li S, Zhou W, Yuan Q, Liu Y. Seizure prediction using spike rate of intracranial EEG. IEEE Trans Neural Syst Rehabil Eng 2013; 21: 880–6.

Litt B, Echauz J. Prediction of epileptic seizures. Lancet Neurol 2002; 1: 22–30.

Loddenkemper T, Lockley SW, Kaleyias J, Kothare SV. Chronobiology of epilepsy: diagnostic and therapeutic implications of chrono-epileptology. J Clin Neurophysiol 2011; 28: 146–53.

Ly JQM, Gaggioni G, Chellappa SL, Papachilleos S, Brzozowski A, Borsu C, et al. Circadian regulation of human cortical excitability. Nature Commun 2016; 7: 11828.

Mader M, Mader W, Gluckman BJ, Timmer J, Schelter B. Statistical evaluation of forecasts. Phys Rev E 2014; 90: 022133.

Mellers B, Stone E, Murray T, Minster A, Rohrbaugh N, Bishop M, et al. Identifying and cultivating superforecasters as a method of improving probabilistic predictions. Perspect Psychol Sci 2015; 10: 267–81.

Mormann F, Andrzejak RG. Seizure prediction: making mileage on the long and winding road. Brain 2016; 139: 1625–7.

Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. Brain 2007; 130: 314–33.

Murphy AH. Hedging and skill scores for probability forecasts. J Appl Meteorol 1973a; 12: 215–23.

Murphy AH. A new vector partition of the probability score. J Appl Meteorol 1973b; 12: 595–600.

Murphy AH. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather Forecast 1993; 8: 281–93.

Murphy AH, Winkler RL. A general framework for forecast verification. Mon Weather Rev 1987; 115: 1330–8.

Nurse E, Mashford BS, Yepes AJ, Kiral-Kornek I, Karoly PJ, Harrer S, et al. Decoding EEG and LFP signals using deep learning: heading TrueNorth. In: Proceedings of the ACM International Conference on Computing Frontiers, 2016. pp. 259–66.

Polikov VS, Tresco PA, Reichert WM. Response of brain tissue to chronically implanted neural electrodes. J Neurosci Methods 2005; 148: 1–18.

Richardson LF. Weather prediction by numerical process. Cambridge, UK: Cambridge University Press; 2007.

Satopaa VA, Baron J, Foster DP, Mellers BA, Tetlock PE, Ungar LH. Combining multiple probability predictions using a simple logit model. Int J Forecast 2014; 30: 344–56.

Schelter B, Feldwisch-Drentrup H, Ihle M, Schulze-Bonhage A, Timmer J. Seizure prediction in epilepsy: from circadian concepts via probabilistic forecasting to statistical evaluation. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference; 2011a. pp. 1624.

Schelter B, Feldwisch-Drentrup H, Schulze-Bonhage A, Timmer J. Epilepsy: the intersection of neurosciences, biology, mathematics, engineering, and physics. In: Osorio I, Zaveri HP, Frei MG, Arthurs S, editors. Seizure Prediction: An Approach Using Probabilistic Forecasting. Boca Raton, FL, USA: CRC Press; 2011b. pp. 249–56.

Schelter B, Winterhalder M, Maiwald T, Brandt A, Schad A, Timmer J, et al. Do False Predictions of Seizures Depend on the State of Vigilance? A Report from Two Seizure-Prediction Methods and Proposed Remedies. Epilepsia 2006; 47: 2058–70.

Schiff SJ. Neural control engineering: the emerging intersection between control theory and neuroscience. Cambridge, MA, USA: MIT Press; 2012.

Sedigh-Sarvestani M, Gluckman BJ. Observing the sleep-wake regulatory system to improve prediction of seizures. In: Tetzlaff R, Elger CE, Lehnertz K, editors. Recent Advances in Predicting and Preventing Epileptic Seizures. Dresden, Germany: World Scientific; 2013. pp. 264–77.

Sedigh-Sarvestani M, Schiff SJ, Gluckman BJ. Reconstructing mammalian sleep dynamics with data assimilation. PLoS Comput Biol 2012; 8: e1002788.

Shouse MN, da Silva AM, Sammaritano M. Circadian rhythm, sleep, and epilepsy. J Clin Neurophysiol 1996; 13: 32–50.

Sillay KA, Rutecki P, Cicora K, Worrell G, Drazkowski J, Shih JJ, et al. Long-term measurement of impedance in chronically implanted depth and subdural electrodes during responsive neurostimulation in humans. Brain Stimul 2013; 6: 718–26.

Snyder DE, Echauz J, Grimes DB, Litt B. The statistics of a practical seizure warning system. J Neural Eng 2008; 5: 392.

Thome-Souza S, Klehm J, Jackson M, Kadish N, Manganaro S, Fernandez IS, et al. Clobazam higher-evening differential dosing as an add-on therapy in refractory epilepsy. Seizure 2016; 40: 1–6.

Valderrama M, Nikolopoulos S, Adam C, Navarro V, Le Van Quyen M. Patient-specific seizure prediction using a multi-feature and multi-modal EEG-ECG classification. In: XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010, 2010. pp. 77–80.

Van Kuyck K, Welkenhuysen M, Arckens L, Sciot R, Nuttin B. Histological alterations induced by electrode implantation and electrical stimulation in the human brain: a review. Neuromodulation 2007; 10: 244–61.

Winterhalder M, Maiwald T, Voss HU, Aschenbrenner-Scheibe R, Timmer J, Schulze-Bonhage A. The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. Epilepsy Behav 2003; 4: 318–25.