Published in final edited form as:
J Neural Eng. 2008 December ; 5(4): 392–401. doi:10.1088/1741-2560/5/4/004.

The Statistics of a Practical Seizure Warning System

David E. Snyder[1], **Javier Echauz**[2], **David B. Grimes**[1], and **Brian Litt**[3]

[1]NeuroVista Corporation, 100 4th Avenue North, Suite 600, Seattle, WA 98109

[2]JE Research, Inc., 170 Wentworth Terrace, Alpharetta, GA 30022

[3]University of Pennsylvania, Department of Neurology, 3400 Spruce St., Philadelphia, PA 19104

Abstract

Statistical methods for evaluating seizure prediction algorithms are controversial and a primary barrier to realizing clinical applications. Experts agree that these algorithms must, at a minimum, perform better than chance, but the proper method for comparing to chance is in debate. We derive a statistical framework for this comparison, the expected performance of a chance predictor according to a predefined scoring rule, which is in turn used as the control in a hypothesis test. We verify the expected performance of chance prediction using Monte Carlo simulations that generate random, simulated seizure warnings of variable duration. We propose a new test metric, the difference between algorithm and chance sensitivities given a constraint on proportion of time spent in warning, and use a simple spectral power-based measure to demonstrate the utility of the metric in four patients undergoing intracranial EEG monitoring during evaluation for epilepsy surgery. The methods are broadly applicable to other scoring rules. We present them as an advance in the statistical evaluation of a practical seizure advisory system.

Keywords

seizure prediction; EEG; epilepsy; intracranial; statistics; brain

1. Introduction

Methods for evaluating seizure prediction algorithms represent a statistical challenge and have generated a great deal of controversy (Litt and Lehnertz 2002, Lehnertz and Litt 2005). When Viglione's group formally assessed the first seizure warning system over 35 years ago, they built a pocket-sized scalp-EEG audiovisual warning device using spectral features, and a trainable analog decision structure (Viglione *et al* 1970a, 1970b, 1973a, 1973b, 1975a). Their 15-month study concluded that the device was “reasonably efficient” (Viglione 1975b). Although the resulting system exhibited a high true prediction rate, it also produced a high rate of false positive warnings (Gevins 2001). The same malady—sensitive but nonspecific prediction—may afflict many of the seizure advisory systems proposed to date, though formal validation is lacking (see reviews Litt and Echauz 2002, Iasemidis 2003, Mormann *et al* 2007). Wide recognition of this problem and some consensus on data and statistical methods required to evaluate the performance of seizure prediction algorithms began to emerge at the most recent International Seizure Prediction Workshop (Schulze-Bonhage *et al* 2007). A recurring theme is that algorithms must, at a minimum, perform better than chance. Investigators address this challenge through a variety of approaches: benchmarks against

dsnyder@neurovista.com .

Disclosure: Drs. Echauz and Litt are paid consultants to NeuroVista.

chance predictors, scientific debates on what electrophysiological events constitute a seizure, and an international collaborative effort to create an archive of continuous, prolonged, gold-standard data sets (Schulze-Bonhage *et al* 2007). Some of these formulations consider seizure generation as an intrinsically probabilistic process, where pro-seizure “permissive” states might come and go, not always leading to ictal events (Wong *et al* 2007). This conclusion is supported by clinical observation (Litt *et al* 2001, Haut 2006, Haut *et al* 2007) but remains difficult to confirm. Other formulations in the category of benchmarks against chance predictors include surrogate data sets, periodic predictors, and random or pseudo-random processes (Andrzejak *et al* 2003, Chaovalitwongse *et al* 2005).

Despite these efforts, investigators in the field of seizure prediction do not yet agree on: (1) how to account for the imprecise and enormously variable temporal relationship between initial algorithm warnings and seizure onset, and (2) a mathematical framework for chance prediction that enables them to translate statistical evaluation of research into clinical utility for epilepsy patients. In our experience, experimental seizure warning signals may vary from brief and sporadic to lengthy or repeated events lasting many hours. One proposal for resolving this dilemma is to trigger a *persistent* warning light in response to a seizure warning. The duration of the light is designed to accommodate the temporal uncertainty of the prediction. The light stays on for a predetermined period of time, so that it does not flash off and on as the next data window is processed, which might confuse the patient. A persistent warning scheme has been described by Winterhalder *et al* (2003), who also propose a methodology for evaluating algorithm performance.

In this paper, we introduce a statistical approach for evaluating practical seizure advisory systems, specifically, systems in which interactions with patients are carefully considered. First we propose to allow the illumination period of a seizure warning light to be extended in response to additional positive predictions by an algorithm. A detailed Poisson-process-based model for a corresponding persistence *chance predictor* is developed, with appropriate performance metrics and methods for hypothesis testing. The theory is corroborated via Monte Carlo simulation (detailed in the Appendix), and then applied to the seizure prediction performance of a simple illustrative algorithm reproducible from published literature, observed over long-term intracranial EEG recordings from four patients. We further demonstrate via Monte Carlo simulation that the derived expressions for this statistical validation method are useful even in non-Poisson stochastic settings, in particular ones that produce clustered warnings with gamma distributed inter-warning intervals. Finally, the candidate and chance predictors are constrained to have a matched percentage of time spent in warning. This stipulation indirectly controls for specificity, without explicitly pitting sensitivity against specificity, as is classically done, which forces data segments to be scored into *a priori* mutually exclusive preictal vs. non-preictal classes.

2. Methods

2.1 Phenomenology of the Seizure Advisory Algorithm

Consider a seizure advisory algorithm that analyzes real-time EEG and produces a uniformly sampled time series of binary data classifications: 1 if preictal, 0 if interictal. These binary classifications are translated into a suitable user interface by interpreting preictal classifications as seizure warnings, and illuminating a persistent warning light according to the following rules: whenever a preictal classification is observed, it triggers a timer of duration τ_w; when another preictal classification occurs, an additional and independent timer is triggered. The warning light is illuminated for as long as any timer is activated, i.e., the logical disjunction (OR) of the timer outputs. In the parlance of electrical engineering, this process is referred to as a retriggerable monostable multivibrator. The basic warning duration τ_w is referred to as *persistence* parameter here. Uninterrupted illumination of the warning light is considered a single warning, regardless of its duration.

2.2 Sensitivity

Algorithm sensitivity S_n, typically defined as the probability of correctly anticipating a seizure within a time horizon, is vital for judging the performance of seizure prediction algorithms as it directly affects patients. In a microtemporal sense, the predictor output at each time point can be compared against preseizure-class labels, with each seizure contributing multiple points to count. In a macrotemporal sense, each seizure is taken as one single, missed or predicted, event to count. We adopt this latter framework, and further define sensitivity as the probability of illuminating a warning light at a time at least τ_{w0} prior to the electrographic onset of a seizure, with the light remaining illuminated at least until the seizure begins. The period τ_{w0} is provided to accommodate uncertainty regarding the precise moment of seizure onset, and to distinguish seizure detection from seizure prediction. This parameter, τ_{w0}, is referred to in this paper as the *detection interval*. Alternatively, it can be used to represent a minimum desired prediction interval, e.g., to accommodate a particular intervention strategy. While these definitions differ in detail and perspective from the scoring intervals defined by Winterhalder *et al* (2003) and Schelter *et al* (2006), they are nonetheless mathematically equivalent: the detection interval τ_{w0} corresponds to the seizure prediction horizon (SPH), while τ_w corresponds to the sum of SPH and the seizure occurrence period of the earlier work.

It is important to note that the definition of seizure, particularly in the setting of assessing algorithm performance in a clinical setting, is not straightforward. One working definition for clinical use might focus on seizures that have clinical impact. In this way, subclinical electrographic events or events that might have minimal quality-of-life impact, such as mild sensory “auras,” might not be counted. Thus, an algorithm meant to predict both clinical seizures and subclinical bursts on the intracranial EEG might have a very different seizure definition than a system meant to prevent patient injury from significant clinical events. The examples given in this paper are scored against electrographic seizures, regardless of clinical manifestation, but the method is applicable to any *a priori* definition of “seizure.”

2.3 Specificity-Related Metrics

A classic definition of specificity in the seizure prediction context is the probability of correctly indicating a nonpreseizure state at any given time. In our experience, the temporal relationship between predictor outputs and seizure onset is highly variable, likely as a function of currently hidden properties of the system, in contrast to the classic treatment where preictal periods are fixed *a priori*. Additionally, any warning (true or false) represents an interruption to the patient that must be dealt with. Thus we propose the *proportion of time spent in warning* ρ_w, and the *warning rate* r_w, as specificity-related metrics. These metrics are compelling from a patient perspective, since they provide information on the frequency and duration of inconveniences caused by warnings. Some combination of proportion of time in warning and warning rate appears to provide a good balance of universality and unambiguous information while disallowing degenerate solutions. In particular, a chance predictor can be constructed so as to provide the same proportion of time under warning as the algorithm under evaluation. The corresponding sensitivity and warning rate may be calculated for the chance predictor and compared against observed performance of the algorithm under evaluation. A successful algorithm will provide higher sensitivity with the same or fewer warning light illuminations per seizure. Alternatively, the chance predictor can be constructed to match the warning rate, and the proportion of time under warning calculated for comparison with the algorithm under test.

2.4 The Chance Predictor

To test whether an algorithm is different from chance, it is necessary to define and characterize a sensorless chance predictor, i.e., without knowledge of EEG signal. One sensible choice is to use a Poisson process, whereby the probability of generating a preictal classification is uniformly distributed. More precisely, the probability of raising a preictal flag in a short interval of duration Δt is approximately equal to $\lambda_{\mathrm{w}}\Delta t$, independent of t, where λ_{w} is referred to as the *Poisson rate parameter*. For any time interval $[0, t]$, the count of trigger events follows a Poisson probability distribution given in the Appendix (along with all derivations of our method). The continuous-time idealization in this Poisson framework leads to formulas that are significantly more compact than those using a purely discrete-time and binomial framework.

2.4.1 Sensitivity of the Chance Predictor—To determine the sensitivity of the chance predictor, S_{nc}, we must first identify the conditions that result in a correct seizure warning. Examples of both successful and unsuccessful warnings are shown in figure 1. Six observations follow: (a) All successful warnings require one or more trigger events to occur in period T_{B} and/or T_{C}, otherwise the warning light is extinguished prematurely. (b) One or more trigger events occurring in period T_{B} is *sufficient* for a correct warning. (c) One or more trigger events occurring in period T_{C} is *not sufficient* for a correct warning, since the light may not have been illuminated early enough. (d) Trigger events occurring in period T_{C} can only contribute to a correct warning if they serve to retrigger an already illuminated warning light. (e) Correct warnings can occur without a trigger event in period T_{B}, but only if there are one or more trigger events in period T_{A} that are successfully retriggered in T_{C}. (f) The only trigger events for which a retrigger event in period T_{C} is *necessary*, are those that occur in period T_{A}.

With these observations in mind, an expression for S_{nc} is shown to be

$$S_{\mathrm{nc}}=1-\exp\left(-\lambda_{\mathrm{w}}\tau_{\mathrm{w}}+\left(1-e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w0}}}\right)\right). \tag{1}$$

2.4.2 Specificity-Related Metrics of the Chance Predictor—In order to compare algorithm performance versus chance, it is also necessary to characterize our proposed specificity-related metrics—proportion of time spent in warning and warning rate—for a corresponding chance predictor. To accomplish this, one constructs a mathematical model for the process of warning light illuminations as originally described, i.e., the logical OR of a collection of independently triggered warning light timers. The proportion of time that the chance predictor spends in warning is shown to be

$$\rho_{\mathrm{w}}=1-e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}}}. \tag{2}$$

Comparison of (2) to (1) reveals that for $\lambda_{\mathrm{w}}\tau_{\mathrm{w0}} << 1$ (as will be the case for any clinically useful seizure advisory algorithm), the sensitivity of the chance predictor is approximately equal to the proportion of time spent in warning—an intuitive result. The corresponding warning rate is shown to be

$$r_{\mathrm{w}}=\lambda_{\mathrm{w}}e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}}}. \tag{3}$$

2.5 Performance Metric for Hypothesis Testing

Comparing the performance of different algorithms presents an interesting dilemma, since both sensitivity and the proportion of time in warning will typically differ for alternative algorithms applied to the same patient. If one algorithm offers better sensitivity, but a greater percentage of time in warning, should it be considered superior or inferior? A solution is suggested by the earlier observation that the sensitivity of a chance predictor is approximately equal to ρ_{w}, the proportion of time in warning. Therefore, the difference between observed and chance sensitivity, subject to matched ρ_{w}, is a powerful metric of predictive ability. It also offers appropriate behavior for limiting cases, being equal to zero if the warning light is either permanently illuminated or extinguished. For an algorithm with observed sensitivity S_{n} and proportion of time in warning ρ_{w}, it follows from (1) and (2) that the *sensitivity improvement-over-chance* metric is calculated as:

$$S_{\mathrm{n}} - S_{\mathrm{nc}} = S_{\mathrm{n}} - 1 + \exp\left(-\lambda_{\mathrm{w}}\tau_{\mathrm{w}} + \left(1 - e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w0}}}\right)\right), \tag{4}$$

where

$$\lambda_{\mathrm{w}} = -\frac{1}{\tau_{\mathrm{w}}}\ln\left(1-\rho_{\mathrm{w}}\right). \tag{5}$$

2.5.1 P-Value of the Performance Test Result—

To assess the significance of an improvement over chance, suppose a candidate algorithm identifies *n* of *N* seizures (i.e., observed sensitivity $S_{\mathrm{n}} = n/N$) for an individual patient. The two-sided p-value is the probability of observing a difference $|n/N - S_{\mathrm{nc}}|$ or greater if the algorithm under evaluation is not different from chance. This is shown to be, in terms of the binomial cumulative distribution function, equal to

$$p = \begin{cases} \left[1 - F_{\mathrm{B}}\left(n-1; N, S_{\mathrm{nc}}\right)\right] + F_{\mathrm{B}}\left(k_{\mathrm{f}}; N, S_{\mathrm{nc}}\right), & \text{for } \frac{n}{N} \geq S_{\mathrm{nc}} \\ \left[1 - F_{\mathrm{B}}\left(k_{\mathrm{c}}-1; N, S_{\mathrm{nc}}\right)\right] + F_{\mathrm{B}}\left(n; N, S_{\mathrm{nc}}\right), & \text{for } \frac{n}{N} < S_{\mathrm{nc}}, \end{cases} \tag{6}$$

where $k_{\mathrm{f}} = \lfloor 2NS_{\mathrm{nc}} - n \rfloor$ and $k_{\mathrm{c}} = \lceil 2NS_{\mathrm{nc}} - n \rceil$. The square bracketed terms of (6) comprise the one-sided p-value for superiority of the algorithm compared to chance. Of particular note is the observation that the remaining terms in (6) will contribute to the p-value only when expected sensitivity of the chance predictor is at least half that of the algorithm under test.

When algorithms are designed and applied prospectively to a population of patients, the values of ρ_{w} will vary from patient to patient. In this case, the overall significance involves the statistics of, e.g., the median sensitivity improvement, with corresponding hypotheses

$$\begin{array}{l} \mathrm{H}_0\text{:median} \quad \left(\left(S_{\mathrm{n}} - S_{\mathrm{nc}}\right) \quad \text{for algorithm}\right) = 0 \\ \mathrm{H}_1\text{:median} \quad \left(\left(S_{\mathrm{n}} - S_{\mathrm{nc}}\right) \quad \text{for algorithm}\right) \neq 0. \end{array}$$

Similarly, when comparing two different algorithms that have been applied to the same population, a paired test is indicated. In general, exact permutation tests may be applied to the Studentized and/or rank-ordered observations involved in the sensitivity improvement metric calculation to account for unknown distributions and multiple-comparison testing in a patient population.

2.6 Monte Carlo Corroboration of the Algorithm-vs.-Chance Prediction Theory

The essential formulas of the retriggerable Poisson chance predictor presented thus far, (1)-(3) and (6), were corroborated using discrete-time Monte Carlo simulations. Additionally, gamma distributed interarrival times were evaluated to examine whether chance prediction might be improved by clustering the seizures and/or the generated warnings. The purpose of these simulations was not only to corroborate the analytical formulas, but to confirm our suspicion that using a temporal distribution of seizures and/or chance predictions which can model temporal clustering does not change the result. The latter could only be done by simulation because no closed-form analytical expressions could be derived from gamma distributions. Results of these Poisson and gamma simulations are shown in Appendix D, and lend support to the Poisson-based formulas and their applicability in a wide sense of algorithm-vs.-chance prediction.

2.7 Application to Intracranial EEG Recordings

The method of comparing candidate seizure predictors to chance was applied to continuous EEG obtained from four patients undergoing evaluation for epilepsy surgery with intracranial electrodes. These data sets were chosen such that half of them display significant prediction performance and half of them do not, in order to illustrate application of the statistical methods detailed in this paper to real clinical data.

2.7.1 Patient and Data Characteristics—Multichannel digital intracranial EEG recordings were obtained from Epilepsy Monitoring Units (EMUs) in Europe and the United States, with approval of the Institutional Review Boards of universities contributing data to NeuroVista Corporation under appropriate material transfer agreements, using approved (NBs/CE Mark in Europe, FDA in the US) clinical EMU equipment. Recordings were meticulously checked for misplugged or mislabeled electrodes, accurate seizure annotations, and electrode placement (validated by MRI). Only complete, continuous recordings were used for the present application, covering each patient's entire EMU stay. Data were normalized to have a common 16-bit dynamic range rescaled to microvolts, and sampling rate of 400 Hz. Table 1 shows redacted characteristics of the four patients whose records were used to demonstrate the statistical methods.

2.7.2 Algorithm—A simple candidate seizure predictor involving spectral power feature extraction and patient-specific classification was employed for all patients. The algorithm analyzed all available channels of cortical potentials including electrodes placed on and surrounding the seizure focus, as well as a reference electrode well-separated from the focus. Multichannel referential EEG was first digitally remontaged to average reference. Each channel was pre-whitened by taking the first forward difference, bandpass filtered to beta band (implemented as 16-32 Hz corner frequencies in a Kaiser window FIR filter design), then subjected to feature extraction in a 5-second sliding window scheme displaced in 1-second increments. For each channel, the output feature every 1 second was the beta power

$$\beta[m]=\frac{1}{2000}\sum_{k=m-1999}^{m}x_{\beta}^{2}[k], \tag{7}$$

where $x_\beta[m]$ is the beta-filtered signal at time index m. A feature vector at each time was formed by collecting the beta powers of all channels.

The feature vectors collected over decimated time samples covering both preictal and interictal periods formed training inputs with which kNN classifiers (k=15 neighbors) were induced. The training outputs stored by the kNNs were integers {1,2} representing “interictal” vs. “90-minute preictal” classes. This latter label was used for immediate training, keeping in mind that final “preictal” labels are revealed only during actual scoring/testing in our framework (e.g., a warning could end up being 5 hours long thus scored as “preictal,” with specificity-related metrics controlling for this asymmetry). When confronted with a new feature vector, kNN looks up the nearest neighbor in its training table, however, the simple estimate of preictal posterior probability (fraction of nearest neighbors belonging to preictal class) was employed instead. No adjustment was used for prior probabilities of the classes. When run as time-serial probability estimator, the kNN output was 60-point-Chebyshev filtered in preparation for smooth predictions. A prediction alert was issued whenever the smooth probability output met/exceeded a threshold. The threshold was algorithmically set such that, in-sample, percentage of time in warning tracked to 25%, however, the relation to actual measured ρ_{w} over a test set is inexact. This internal threshold can be fixed arbitrarily (recall our comparison to chance requires only some "final" S_{n} and ρ_{w}). The output of all predictors were subjected to persistence processing with parameters τ_{w} = 90 minutes and τ_{w0} = 1 minute.

If seizures clustered within a 4-hour period, only the leading (first) seizure in that cluster was enforced for prediction scoring purposes, for reasons similar to those of earthquake prediction —the goal is prediction of main events rather than detection of aftershocks. The scoring of follow-up clustered seizures was treated as deleted or never-seen data. For example, if original data had seizure onsets indicated as [0 0 0 0 1 0 0 0 0 1 0 1 1 0 0 0 0 1], the removal of clustered seizures looked like [0 0 0 0 1 0 0 0 0 1 0 0 0 0 1]. Once these labels were fixed, the treatment of candidate predictors and the theoretical chance predictor against which we compare remains equitable and consistent with the theory. Without loss of generality, the estimation of expected S_{n}, ρ_{w}, and r_{w} for each patient was based on N-fold cross validation, where N is the number of leading seizures, with data assigned to folds in approximately 3-hour epochs.

2.7.3 Resetting the Warning System After Seizures

It is important to note that the occurrence of seizures has been reported to "reset" the process of seizure generation, and that performance of a seizure warning system, particularly its specificity, may be enhanced by including this feature. We have chosen not to expand our analysis to take this into account for the following reasons: First, when a clinical seizure warning system is deployed, patient-system interactions will inevitably form a closed loop. For example, patients may take fast-acting medications when receiving indication of a high probability of seizure onset, which will likely prevent, delay or otherwise alter the process of seizure generation. At present these is insufficient data on such interactions to generate models to incorporate into the above paradigm. In such a system, even without intervention, a patient may learn to alter their seizures when informed of oncoming events. Second, we wanted to keep the chance predictor "sensorless," without mixing it with EEG information.

With regard to seizure localization during epilepsy monitoring, clustered seizures within 4 hours or less which we delete for scoring in our example are often treated as a single event, because of their tendency to arise from the same location (Wieser *et al* 1979). The formulas in our model applied to the cluster-removed data are still the same. Formulas would change with warning cancellation after seizure detection, but as explained above this falls outside the desired scope of comparing against pure chance.

3. Results

Figure 2 shows one week's worth of the candidate warning outputs for each of the patients in relation to their leading seizures (patients A and C had data exceeding 1 week, not fitting in the plots but represented in the statistical results).

Table 2 lists the observed prediction metrics and significance levels for the four exemplary patients. We now use Patient C in order to clarify the calculation of p-value in the framework presented thus far. The seizure advisory algorithm with persistence of warning lights was applied to intracranial EEG over the entire 187.4 hour observation period. We have:

$$\begin{aligned}\tau_{\mathrm{w0}} &= 1/60 \quad \text{of an hour}\\ \tau_{\mathrm{w}} &= 1.5 \quad \text{hours}\\ N &= 5 \quad \text{seizures}\\ n &= 3 \quad \text{seizures predicted successfully}\\ \rho_{\mathrm{w}} &= 26.5\% \quad \text{of time spent in warning.}\end{aligned}$$

The Poisson rate for the chance predictor is

$$\lambda_{\mathrm{w}} = -\frac{1}{\tau_{\mathrm{w}}}\ln\left(1-\rho_{\mathrm{w}}\right) = 0.205/\mathrm{hour}$$

and sensitivity is

$$S_{\mathrm{nc}} = 1-\exp\left(-\lambda_{\mathrm{w}}\tau_{\mathrm{w}}+\left(1-e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w0}}}\right)\right) = 26.3\%.$$

Hence, the p-value is

$$p = 1-F_{\mathrm{B}}\left(3-1;5,0.263\right) = 0.118.$$

In the above result, the upper right term of (23) does not contribute, since S_{nc} is less than half of the observed sensitivity (k_{f} evaluates to a negative number).

3.3. Discussion

The utility of our method is demonstrated by the example calculations given above. Although the use of actual patient data is almost inconsequential in illustrating the comparison method —completely synthetic streams of predictor outputs could have done the same job—it does give a glimpse into what capabilities may lie in store for some patients.

For patient A, the clustering of warnings in the proximity of seizures is evident (both pre- and postictal), combined with day-long interictal intervals without warning. The p-value indicates that such a result is highly unlikely by chance. Patient B individually also achieves statistical significance, but in spite of higher sensitivity and lower proportion of time in warning compared to A, the p-value is less compelling. This is partly a consequence of the smaller number of seizures, and therefore decreased statistical power. Table 2 suggests that the time in warning averaged about a quarter of the day, and the warning rates averaged about 2 disruptions per day in the lives of these two patients, in exchange for 87.5-100% prediction sensitivity. Patient C fails to achieve significance because of the relatively low sensitivity and modest event count, while patient D fails because of the high proportion of time in warning, in spite of 100% sensitivity. The sensitivity improvement-over-chance metric, equation (4), was implicit in the calculation of p-value in Table 2, but explicit evaluation of $S_{\mathrm{n}} - S_{\mathrm{nc}}$ would be useful, e.g., for ranking several candidate predictors during algorithm development.

These results are presented not to suggest that they are adequate for a clinical seizure warning device, but rather to illustrate the translation of the methods presented in this paper into a practical warning system for patient use. A more accurate prediction algorithm would be required to obtain clinically useful results, along with careful consideration of performance targets required for clinical utility, which might vary according to individual patients.

4. Conclusions

We present a framework for comparing the performance of EEG-based candidate seizure predictors to sensorless chance predictors in a way that facilitates development of a clinically useful system. A key departure of this method from previous approaches is allowing the brain to inform the algorithm, dynamically, how far in advance of seizures the epileptic network becomes “proictal” (i.e., enters a state with a high probability of seizure onset), as opposed to the algorithm evaluator predefining rigid preictal vs. non-preictal time windows. This variable-length warning requirement is modestly accomplished through our proposed: (1) retriggerable persistence indicator light, and (2) new performance metric: difference between algorithm and chance sensitivities, given a constraint on proportion of time spent in warning, which indirectly controls for specificity. For evaluating the statistical significance of candidate predictors, we present an explicit p-value formula. Both the theory and the Monte Carlo simulations suggest that our chance prediction formulation holds independent of seizure clustering and for a wide class of sensorless predictors.

A practical example of the method is offered in the analysis of four patient records selected so that half of them show predictive value and half of them do not. A simple algorithm is used to show how the evaluation scheme works to identify, accept, and reject predictive results, and how a candidate predictor can show statistically honest superiority over chance without necessarily involving extreme complexity. We do not present this algorithm or its performance as sufficient for a clinically useful device. For those patients whose prediction performance shows statistical significance, the sensitivity, proportion of time in warning, and warning rates of the advisory systems may be within the range of clinical useful values, though they have not been controlled for possible confounding effects (e.g., states of awareness). Based upon the above methods and results, however, we believe that an appropriately powered, controlled study showing prospective prediction from EMU data is warranted to guide the development of a clinically useful, patient-oriented, seizure advisory system.

Acknowledgments

The authors would like to gratefully acknowledge the collaboration of Gregory Worrell, M.D., Ph.D., in providing clinical data and insight into this research. Dr. Litt's research is funded by the National Institutes of Health (RO1NS048598, 1NS041811-02, 2RO1CA084438-05A2), the Klingenstein Foundation, the Dana Foundation, The Epilepsy Research Foundation, the Pennsylvania Tobacco Fund, and the Whitaker Foundation.

Appendix A. Derivation of Chance Predictor Sensitivity

To count the trigger events, the counting random process $\{N_t, 0 \leq t < \infty\}$ is introduced, which has the Poisson probability distribution

$$P\,[N_t{=}k] = \frac{e^{-\lambda_{\mathrm{w}} t}(\lambda_{\mathrm{w}} t)^k}{k!}. \tag{A.1}$$

From the six figure 1 observations,

$$\begin{aligned} S_{\mathrm{nc}} = \; & P\,[\text{successful seizure prediction} \quad | \quad \text{one or more trigger events in} \quad \mathrm{B}] \times \\ & P\,[\text{one or more trigger events in} \quad \mathrm{B}] + \\ & P\,[\text{successful seizure prediction} \quad | \quad \text{no trigger events in} \quad \mathrm{B}] \times \\ & P\,[\text{no trigger events in} \quad \mathrm{B}]\,. \end{aligned} \tag{A.2}$$

As already observed, *P*[successful seizure advisory | one or more trigger events in B] is equal to unity, so (A.2) reduces to

$$\begin{aligned} S_{\mathrm{nc}} = \; & P[\text{one or more trigger events in} \quad \mathrm{B}] + \\ & P[\text{successful seizure prediction} \quad | \quad \text{no trigger events in} \quad \mathrm{B}] \times \\ & (1 - P[\text{one or more trigger events in} \quad \mathrm{B}]) . \end{aligned} \tag{A.3}$$

P[one or more trigger events in B] is easily calculated using (A.1), and represents the sensitivity due to preictal classifications that do not require retriggering. The result is in agreement with the formula of Winterhalder *et al* (2003):

$$\begin{aligned} P[\text{one or more trigger events in} \quad \mathrm{B}] & = P[N_{\tau_{\mathrm{w}} - \tau_{\mathrm{w0}}} > 0] \\ & = 1 - P[N_{\tau_{\mathrm{w}} - \tau_{\mathrm{w0}}} = 0] \\ & = 1 - e^{-\lambda_{\mathrm{w}}(\tau_{\mathrm{w}} - \tau_{\mathrm{w0}})} . \end{aligned} \tag{A.4}$$

The conditional term in (A.3) is more subtle. To have a successful warning without a trigger event in period T_{B} requires that a trigger event first occur in period T_{A} and be successfully retriggered in period T_{C}. The rate of these rarer events is found by multiplying the underlying trigger event rate, λ_{w}, by the probability of being retriggered successfully in T_{C}:

$$\lambda_{\mathrm{A}}(t) = \lambda_{\mathrm{w}} P[\text{successful retrigger in} \quad \mathrm{C}] . \tag{A.5}$$

To illustrate, the example of figure 1c is reproduced in figure A.1 with the time origin shifted for notational convenience. A trigger event occurring at time *t* in period T_{A} is successfully retriggered in T_{C} if another trigger event occurs within the period τ_{w} to $t+\tau_{\mathrm{w}}$. As a consequence, the probability of a successful retrigger in T_{C} is simply the probability of one or more trigger events occurring in an interval of duration *t* (the period from τ_{w} to $t+\tau_{\mathrm{w}}$) which according to (A.1) is

$$\begin{aligned} P[\text{successful retrigger in} \quad \mathrm{C}] & = P[N_t > 0] \\ & = 1 - P[N_t = 0] \\ & = 1 - e^{-\lambda_{\mathrm{w}} t} . \end{aligned} \tag{A.6}$$

Substitution of (A.6) into (A.5) yields

$$\lambda_{\mathrm{A}}(t) = \lambda_{\mathrm{w}} \left(1 - e^{-\lambda_{\mathrm{w}} t}\right), \qquad t < \tau_{\mathrm{w0}} . \tag{A.7}$$

An important observation is that the rate of successful trigger events in A is not constant, but a function of *t*, so that the distribution of successful trigger events in period T_{A} is characterized by a non-homogeneous Poisson process. Accordingly, we define the counting random variable $\{M_t, 0 \le t < \tau_{\mathrm{w0}}\}$ for the non-homogeneous Poisson process characterized by the rate parameter λ_{A}. The probability distribution of the non-homogeneous Poisson process is given by

$$\begin{aligned} P[M_t = k] & = \frac{1}{k!} \left(\int_0^t \lambda_{\mathrm{A}}(u)\, du \right)^k \exp\left(- \int_0^t \lambda_{\mathrm{A}}(u)\, du \right) \\ & = \frac{1}{k!} \left[\lambda_{\mathrm{w}} t - \left(1 - e^{-\lambda_{\mathrm{w}} t}\right) \right]^k \exp\left(-\lambda_{\mathrm{w}} t + \left(1 - e^{-\lambda_{\mathrm{w}} t}\right) \right) . \end{aligned} \tag{A.8}$$

Hence,

$$\begin{aligned} P\left[\text{successful seizure prediction} \quad | \quad \text{no trigger events in} \quad \mathrm{B}\right] &= P\left[M_{\tau_{\mathrm{w}0}} > 0\right] \\ &= 1 - P\left[M_{\tau_{\mathrm{w}0}} = 0\right] \\ &= 1 - \exp\left(-\lambda_{\mathrm{w}}\tau_{\mathrm{w}0} + \left(1 - e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}0}}\right)\right). \end{aligned} \tag{A.9}$$

Substitution of (A.4) and (A.9) into (A.3) yields the sensitivity of the chance predictor

$$\begin{aligned} S_{\mathrm{nc}} &= \left(1 - e^{-\lambda_{\mathrm{w}}(\tau_{\mathrm{w}} - \tau_{\mathrm{w}0})}\right) + \left[1 - \exp\left(1 - \lambda_{\mathrm{w}}\tau_{\mathrm{w}0} - e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}0}}\right)\right] \times \left[1 - \left(1 - e^{-\lambda_{\mathrm{w}} - (\tau_{\mathrm{w}} - \tau_{\mathrm{w}0})}\right)\right] \\ &= 1 - e^{-\lambda_{\mathrm{w}}(\tau_{\mathrm{w}} - \tau_{\mathrm{w}0})} \exp\left(1 - \lambda_{\mathrm{w}}\tau_{\mathrm{w}0} - e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}0}}\right) \\ &= 1 - \exp\left(-\lambda_{\mathrm{w}}\tau_{\mathrm{w}} + \left(1 - e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}0}}\right)\right). \end{aligned} \tag{A.10}$$

Appendix B. Derivation of Chance Predictor Specificity-Related Metrics

Consider a set of warning trigger events occurring at times $\{T_j, j = 1, 2, 3, \ldots, N_t\}$ where N_t is the previously defined Poisson counting variable and the T_j are referred to as arrival times. Define a new random process I_t to count the number of active timers:

$$I_t \equiv \sum_{j=1}^{N_t} h\left(t - T_j\right), \tag{B.1}$$

where

$$h(t) = \begin{cases} 1 & \text{for} \quad 0 \le t < \tau_{\mathrm{w}} \\ 0 & \text{otherwise.} \end{cases} \tag{B.2}$$

The function $h(t)$ represents a warning light timer of duration τ_{w}. Since the warning light is extinguished only when I_t is identically zero, the proportion of time in alert is equal to $1 - P[I_t = 0]$.

Equation (B.1) represents a filtered (or compound) Poisson process. Since I_t is comprised of a sum of functions of random variables, evaluation of $P[I_t = k]$ is most easily performed via Fourier transform to obtain the characteristic function of the filtered Poisson process (Davenport 1970).

$$\begin{aligned} \varphi_{I_t}(\nu) &= E\left[e^{i\nu I_t}\right] \\ &= \exp\left[\lambda_{\mathrm{w}} \int_0^t \left(e^{i\nu h(u)} - 1\right) du\right]. \end{aligned} \tag{B.3}$$

Substitution of (B.2) into (B.3) leads to

$$\varphi_{I_t}(\nu) = \exp\left[\lambda_{\mathrm{w}}\tau_{\mathrm{w}}\left(e^{i\nu} - 1\right)\right]. \tag{B.4}$$

Fourier inversion of (B.4) provides

$$P\left[I_t=k\right]=\frac{e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}}}\left(\lambda_{\mathrm{w}}\tau_{\mathrm{w}}\right)^k}{k!}, \tag{B.5}$$

and the proportion of time in warning is

$$\rho_{\mathrm{w}}=1-P\left[I_t=0\right]=1-e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}}}. \tag{B.6}$$

To calculate the rate of warnings, r_{w}, we observe that a new warning will occur whenever the interval between successive preictal trigger events exceeds τ_{w}. To examine the properties of such intervals, we define a corresponding random process:

$$Z_k \equiv T_k - T_{k-1}. \tag{B.7}$$

For T_k determined by a Poisson process, such “interarrival times” are themselves exponentially distributed, independent random variables (Davenport 1970).

$$F_{Z_k}(z)=\begin{cases} 1-e^{-\lambda_{\mathrm{w}}z} & \text{for} \quad z \geq 0 \\ 0 & \text{for} \quad z<0. \end{cases} \tag{B.8}$$

With this information, the rate of warnings can be computed by multiplying the rate of preictal classifications by the proportion of interarrival times greater than τ_{w}:

$$r_{\mathrm{w}}=\lambda_{\mathrm{w}}P\left[Z_k>\tau_{\mathrm{w}}\right]=\lambda_{\mathrm{w}}\left(1-F_{Z_k}\left(\tau_{\mathrm{w}}\right)\right)=\lambda_{\mathrm{w}}e^{-\lambda_{\mathrm{w}}\tau_{\mathrm{w}}}. \tag{B.9}$$

Appendix C. Derivation of P-Value for the Performance Test Result

Given a candidate algorithm's observed sensitivity $S_{\mathrm{n}} = n/N$, the two-sided p-value is the probability of the chance predictor also displaying a difference $|n/N - S_{\mathrm{nc}}|$ or greater:

$$\begin{aligned} p &= P\left[\left(\frac{n'}{N}-S_{\mathrm{nc}}\right) \geq \left|\frac{n}{N}-S_{\mathrm{nc}}\right|\right]+P\left[\left(\frac{n'}{N}-S_{\mathrm{nc}}\right) \leq -\left|\frac{n}{N}-S_{\mathrm{nc}}\right|\right] \\ &= \begin{cases} P\left[n' \geq n\right]+P\left[n' \leq (2N \cdot S_{\mathrm{nc}}-n)\right], & \text{for}\,\frac{n}{N} \geq S_{\mathrm{nc}} \\ P\left[n' \geq (2N \cdot S_{\mathrm{nc}}-n)\right]+P\left[n' \leq n\right], & \text{for}\,\frac{n}{N}<S_{\mathrm{nc}}. \end{cases} \end{aligned} \tag{C.1}$$

where n' is the random variable corresponding to the number of predicted seizures under the null hypothesis. Each of the N seizures can be considered a Bernoulli trial with the probability of advisory equal to the expected sensitivity of the chance predictor. Accordingly, (C.1) can be rewritten in terms of the binomial cumulative distribution function

$$p=\begin{cases} \left[1-F_{\mathrm{B}}\left(n-1;N,S_{\mathrm{nc}}\right)\right]+F_{\mathrm{B}}\left(k_{\mathrm{f}};N,S_{\mathrm{nc}}\right), & \text{for}\,\frac{n}{N} \geq S_{\mathrm{nc}} \\ \left[1-F_{\mathrm{B}}\left(k_{\mathrm{c}}-1;N,S_{\mathrm{nc}}\right)\right]+F_{\mathrm{B}}\left(n;N,S_{\mathrm{nc}}\right), & \text{for}\,\frac{n}{N}<S_{\mathrm{nc}}, \end{cases} \tag{C.2}$$

where

$$F_{\mathrm{B}}(k;n,p) \equiv \sum_{j=0}^{k} f_{\mathrm{B}}(j;n,p), \tag{C.3}$$

$$f_{\mathrm{B}}(k;n,p) \equiv \binom{n}{k} p^{k}(1-p)^{n-k}, \tag{C.4}$$

and

$$\begin{aligned} k_{\mathrm{f}} &= \mathrm{floor}\,(2N \cdot S_{\mathrm{nc}} - n) \\ k_{\mathrm{c}} &= \mathrm{ceiling}\,(2N \cdot S_{\mathrm{nc}} - n). \end{aligned} \tag{C.5}$$

Appendix D. Monte Carlo Corroboration of the Theory

The performance of a hypothetical EEG-based predictor was used as a reference for defining chance prediction experiments to corroborate formulas (1)-(3) and (6). The reference predictor had sensitivity $S_{\mathrm{n}} = 80\%$ (e.g., predicts 4 out of 5 seizures), with a proportion of time in warning $\rho_{\mathrm{w}} = 27.5\%$, persistence $\tau_{\mathrm{w}} = 90$ minutes, and detection interval $\tau_{\mathrm{w0}} = 1$ minute. To test the chance predictor, two types of seizure sequences were generated: the first with uniformly distributed seizures (5 seizures over a 140-hour period), and the second based on gamma distributions to simulate the seizure clustering phenomena often seen in epilepsy (Haut 2006).

To generate chance predictions, the time axis was discretized into 1-second increments. Trigger events for the Poisson-based discrete-time chance predictor were obtained as a sequence of Bernoulli trials with probability p_{B} equal to the rate parameter given by formula (5): $\lambda_{\mathrm{w}} = 5.955\times10^{-5}$ chance assertions/second ($p_{\mathrm{B}} \approx \lambda_{\mathrm{w}}\,\Delta t$, with $\Delta t = 1$ s). The trigger events were used to generate persistent warning indicators with persistence sample count = 5400 (90 minutes). Ten thousand Monte Carlo simulation trials were conducted for each type of seizure sequence (uniform and clustered). The following quantities were tallied: (a) number of true positives, (b) cumulatively revised sensitivity, proportion of time spent in warning and rate of warnings (number of rising edges per hour), and (c) fraction of trials where sensitivity of the chance predictor was equal to or better than the 80% sensitivity of the reference predictor (for p-value corroboration). At the end of a simulation, each Monte Carlo estimate is an average across all trials. Results of these trials are summarized in the first two rows of Table D.1.

An experiment was additionally performed to explore whether formulas (1)-(3) and (6) might generalize to non-Poisson chance predictors. A gamma-based chance predictor that issued clustered warnings was constructed and scored against clustered seizure sequences. This was accomplished by sampling inter-trigger intervals from a gamma distribution

$$f_{\Gamma}(x;\alpha,\beta) = \frac{x^{\alpha-1}\beta^{\alpha}e^{-\beta x}}{\Gamma(\alpha)} \tag{D.1}$$

with shape parameter $\alpha = 0.2$, and scale parameter β set to obtain the desired $\rho_{\mathrm{w}} = 27.5\%$, and then applying the persistent warning indicator rules. The results of this experiment are shown in the final row of Table D.1. For three combinations of seizure distribution and type of chance predictor {(uniform, Poisson), (gamma, Poisson), and (gamma, gamma)}, Table D.1 shows the agreement found between expected values of the variables according to the Poisson formulas and the corresponding Monte Carlo estimates.

References

Andrzejak RG, Mormann F, Kreuz T, Rieke C, Kraskov A, Elger CE, Lehnertz K. Testing the null hypothesis of the nonexistence of a preseizure state. Phys. Rev. E Stat. Nonlin. Soft Matter Phys 2003;67:010901. [PubMed: 12636484]

Chaovalitwongse W, Iasemidis LD, Pardalos PM, Carney PR, Shiau DS, Sackellares JC. Performance of a seizure warning algorithm based on the dynamics of intracranial EEG. Epilepsy Res 2005;64:93–113. [PubMed: 15961284]

Davenport, WB. Probability and Random Processes. McGraw-Hill Book Company; New York: 1970.

Gevins, A. Pers. comm. 2001.

Haut SR. Seizure clustering. Epilepsy Behav 2006;8:50–5. [PubMed: 16246629]

Haut SR, Hall CB, Masur J, Lipton RB. Seizure occurrence: precipitants and prediction. Neurology 2007;69:1905–10. [PubMed: 17998482]

Iasemidis LD. Epileptic seizure prediction and control. IEEE Trans. Biomed. Eng 2003;50:549–58. [PubMed: 12769431]

Lehnertz K, Litt B. The First International Collaborative Workshop on Seizure Prediction: summary and data description. Clin. Neurophysiol 2005;116:493–505. [PubMed: 15721063]

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

Litt B, Esteller R, Echauz J, D'Alessandro M, Shor R, Henry T, Pennell P, Epstein C, Bakay R, Dichter M, et al. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. Neuron 2001;30:51–64. [PubMed: 11343644]

Litt B, Lehnertz K. Seizure prediction and the preseizure period. Curr. Opin. Neurol 2002;15:173–7. [PubMed: 11923631]

Marsaglia G, Tsang W. A simple method for generating gamma variables. ACM Trans. Math. Software 2000;26:363–72.

Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. Brain 2007;130:314–33. [PubMed: 17008335]

Schelter B, Winterhalder M, Maiwald T, Brandt A, Schad A, Schulze-Bonhage A, Timmer J. Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. Chaos 2006;16:013108. [PubMed: 16599739]

Schulze-Bonhage, A.; Timmer, J.; Schelter, B. 3rd International Workshop on Epileptic Seizure Prediction; Freiburg, Germany. 2007.

Viglione, S. Applications of pattern recognition technology. In: Mendel, J.; Fu, K., editors. Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications. Academic Press; New York: 1970. p. 115-61.

Viglione, S.; Ordon, V.; Martin, W.; Kesler, C. Epileptic seizure warning system (U.S.). 1973. Patent 3,863,625

Viglione, S.; Ordon, V.; Risch, F. Proc. 21st Western Institute on Epilepsy. McDonnell Douglas Astronautics Co.; West Huntington Beach, CA: 1970. A methodology for detecting ongoing changes in the EEG prior to clinical seizures. p WD1399(A)

Viglione, SS. Detection and prediction of epileptic seizures validation program--Proposal for testing and evaluation of epileptic seizures alerting systems. McDonnell Douglas Astronautics Co.; West Huntington Beach, CA: 1973.

Viglione, SS. Validation of epilepsy seizure warning system. McDonnell Douglas Astronautics Co.; West Huntington Beach, CA: 1974. APA 74133

Viglione, SS. Validation of the epilepsy seizure warning system (special report). Social and Rehabilitation Service; Washington, D.C.: 1974. Grant SRS-23-57680

Viglione, SS. Validation of the epilepsy seizure warning system (final report). Social and Rehabilitation Service; Washington, D.C.: 1975. Grant SRS-23-57680

Viglione SS, Walsh GO. Proceedings: Epileptic seizure prediction. Electroencephalogr. Clin. Neurophysiol 1975;39:435–6. [PubMed: 51767]

Wieser HG, Bancaud J, Talairach J, Bonis A, Szikla G. Comparative value of spontaneous and chemically and electrically induced seizures in establishing the lateralization of temporal lobe seizures. Epilepsia 1979;20:47–59. [PubMed: 421676]

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. [PubMed: 12791335]

Wong S, Gardner AB, Krieger AM, Litt B. A stochastic framework for evaluating seizure prediction algorithms using hidden Markov models. J. Neurophysiol 2007;97:2525–32. [PubMed: 17021032]

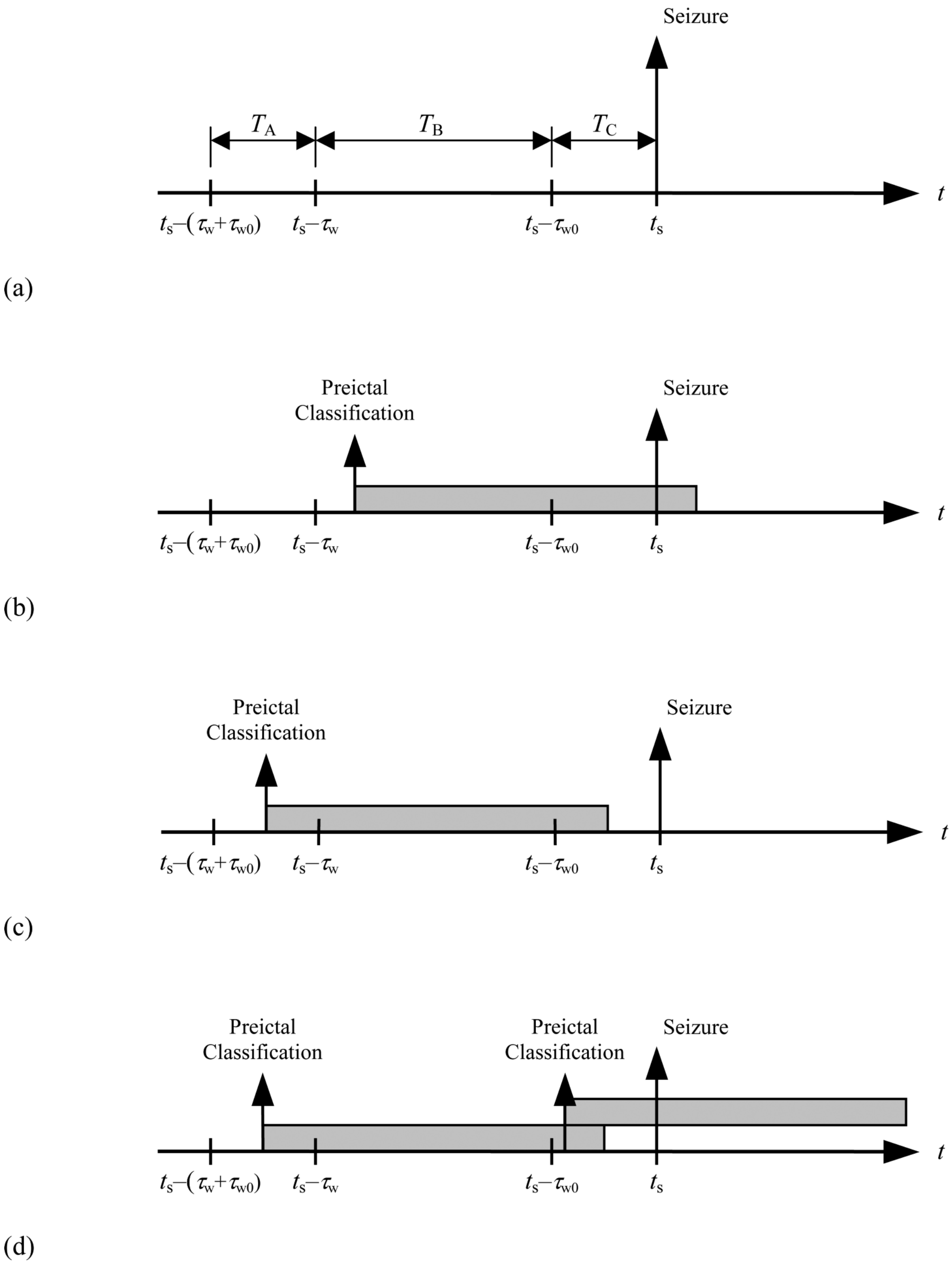

Figure 1.

Examples of successful and unsuccessful warnings in relation to a seizure onset at time t_s: (a) Definition of T_A, T_B, T_C time periods. (b) A successful seizure warning. (c) A failed seizure warning. (d) A seizure warning made successful by retriggering.

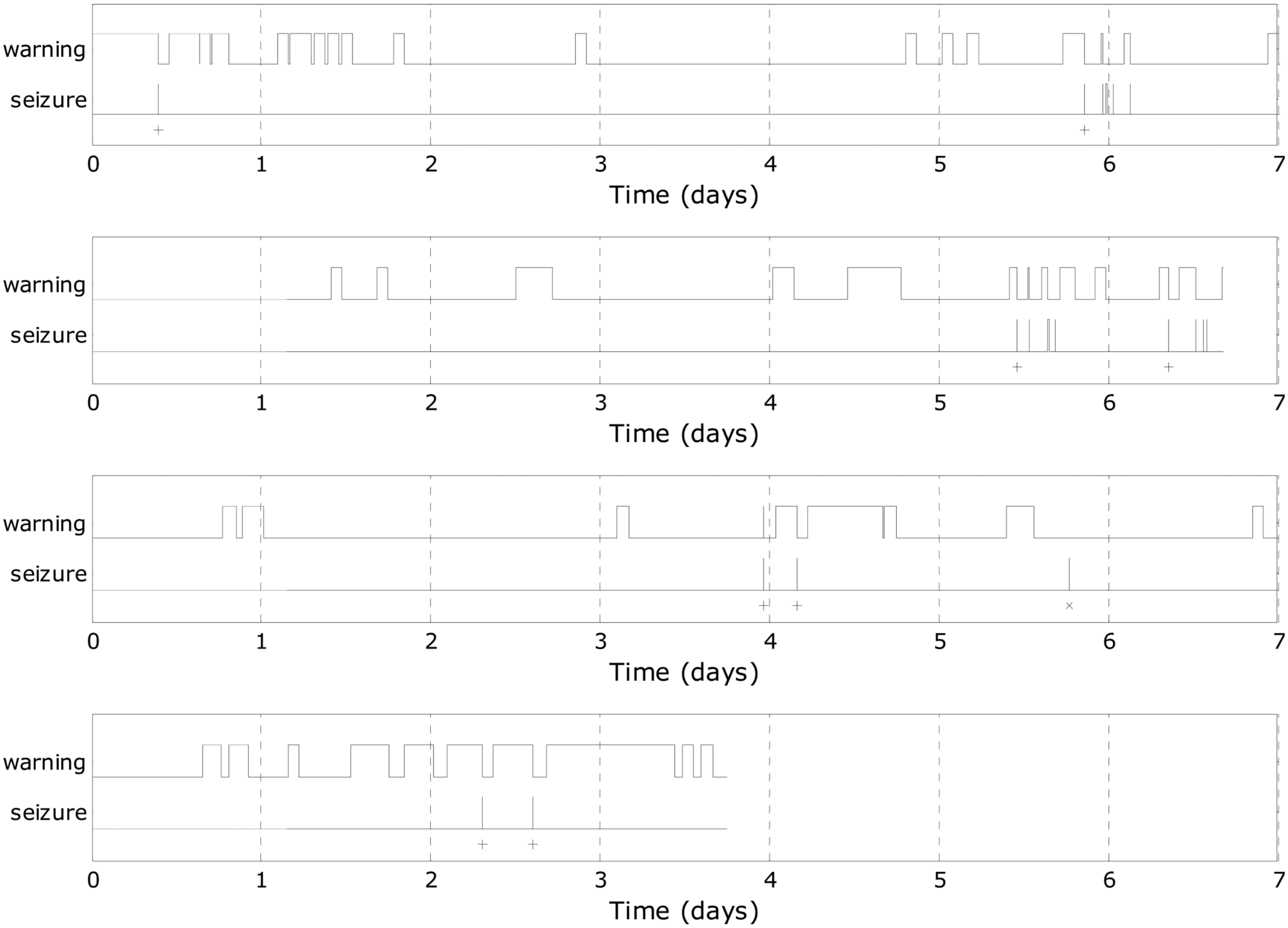

Figure 2.

Actual candidate predictor outputs after persistence over a week-long recording of four patients in relation to their leading seizures ('x' indicates false negative, and '+' indicates true positive).

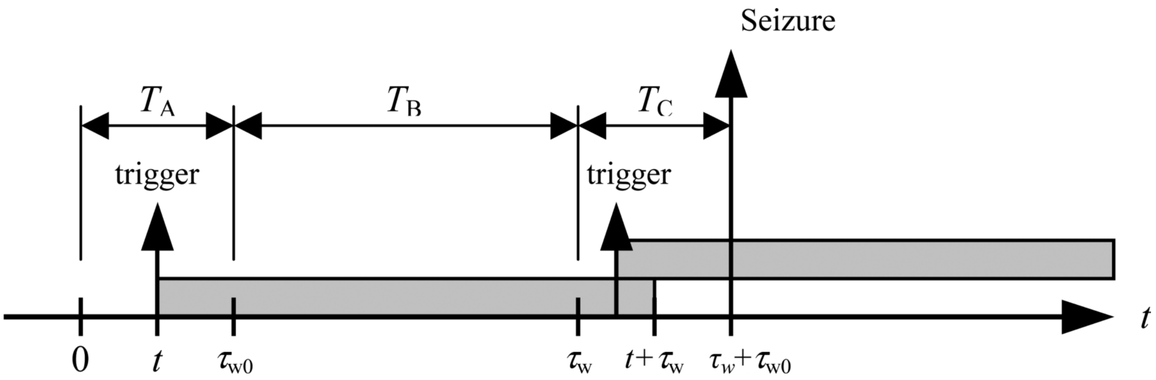

Figure A.1.
Successfully retriggered warning light with time origin shifted.

Table 1

Patient and data characteristics. CP=complex partial; GTC=generalized tonic-clonic.

Patient	Gender	Seizure Type(s)	Ictal Onset Zone(s)	Side of Resection	Recording Duration (hours)
A	F	CP, GTC	Inferior temporal gyrus of the left temporal lobe	Left	422.7
B	F	CP	Temporal, parietal and frontal lobes	Left	131.1
C	M	CP	Multifocal, more frequently left mesial temporal	None (indeterminate)	187.4
D	M	CP, GTC	Right frontal	Right	90.0

Table 2

Observed prediction metrics and significance levels for the four exemplary patients.

Patient	Total # of Seizures	Sensitivity S_n	Warning Proportion	Warning Rate (per hour)	P-Value
A	8	87.5%	29.0%	0.099	0.001
B	2	100.0%	19.1%	0.076	0.036
C	5	60.0%	26.5%	0.091	0.118
D	2	100.0%	53.3%	0.111	0.502

Table D.1

Monte Carlo estimates of chance prediction metrics. P-value is the probability of chance explaining the S_n=80% of a matched reference algorithm. The expected values according to the Poisson formulas are $E[S_{nc}] = 0.27241$, $E[\rho_w] = 0.275$, $E[r_w] = 0.155$/hr, and $p = 0.02153$.

Seizures vs. Chance Predictor Distributions	Quantity of Interest	Monte Carlo Simulation
Uniform, Poisson	$E[S_{nc}]$	0.27053
	$E[\rho_w]$	0.275
	$E[r_w]$	0.156/hr
	p-value	0.02008
Gamma, Poisson	$E[S_{nc}]$	0.27390
	$E[\rho_w]$	0.275
	$E[r_w]$	0.155/hr
	p-value	0.02480
Gamma, Gamma	$E[S_{nc}]$	0.27558
	$E[\rho_w]$	0.275
	$E[r_w]$	0.131/hr
	p-value	0.03100