Received: 3 December 2025 | Revised: 4 February 2026 | Accepted: 5 February 2026

DOI: 10.1002/epi.70158

Epilepsia®

RESEARCH ARTICLE

Are seizure forecasts and cycles better than chance? What chance?

Ralph G. Andrzejak[1] | **Martin Brešar**[2] | **Mark P. Richardson**[3,4] | **Pedro F. Viana**[3,4,5] | **Richard Rosch**[3,6]

[1]Department of Engineering, Universitat Pompeu Fabra, Barcelona, Spain

[2]Department of Systems and Control, Jožef Stefan Institute, Ljubljana, Slovenia

[3]Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, UK

[4]Epilepsy Centre, King's College Hospital NHS Foundation Trust, London, UK

[5]Centro de Estudos Egas Moniz, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

[6]Wellcome Centre for Imaging Neuroscience, University College London, London, UK

Correspondence
Ralph G. Andrzejak, Department of Engineering, Universitat Pompeu Fabra, Carrer Roc Boronat 138, 08018 Barcelona, Catalonia, Spain.
Email: ralph.andrzejak@upf.edu

Funding information
NIHR, Grant/Award Number: CL-2023-17-004; NIHR Development and Skills Enhancement Award, Grant/Award Number: NIHR305718; Spanish Ministry of Science and Innovation and the State Research Agency, Grant/Award Number: PID2020-118196GBI00/MICIU/AEI/10.13039/501100011033; Slovenian Research and Innovation Agency, Grant/Award Number: P2-0001

Abstract

Objective: There is a growing synergy between the lines of research on cycles in epilepsy and seizure forecasting. It has been conjectured, for instance, that incorporating information about significant seizure cycles into forecasting algorithms can lead to a better-than-chance forecasting performance. However, significance and better-than-chance are each typically evaluated against only a single null hypothesis, for example, that forecasts are generated by a Poisson process. We here argue that this should be considered only a first step. Our objective is to demonstrate the importance of testing complementary null hypotheses that represent alternative chance models.

Methods: To ensure controlled conditions, we use synthetic data generated from simple mathematical models. Samples drawn from gamma distributions are used to generate sequences of random seizure times and random forecasts. We then determine the strength of cycles as a function of the cycle duration and calculate the sensitivity and fraction of time under alarm obtained for the random forecasting algorithm. In both analyses, we apply numerical, surrogate-based null-hypothesis testing methods. In the latter case, this includes a straightforward approach to correcting for multiple testing on nonindependent data.

Results: Counterintuitively, the random seizure-time sequences contain multiple prominent cycles, which are judged highly significant by the Rayleigh test. Moreover, randomly forecasting random seizure times results in a sensitivity of 79% at a fraction of time under alarm of only 42%, clearly outperforming a Poisson-like predictor. In both cases, however, the flexibility and versatility of surrogate-based null-hypothesis tests allow us to successfully reveal that all results can be explained by chance models.

Significance: Before reaching conclusions on real cycles in epilepsy, the forecastability of seizures, and genuine capacity of forecasting algorithms, it is essential to test and reject several complementary null hypotheses. Many

Social media summary: When "better than chance" misleads you: a cautionary tale on seizure cycles, forecasting, and why multiple null hypotheses must be tested.

conclusions might not withstand such rigorous tests, allowing the community to focus on those that do.

KEYWORDS
null-hypothesis testing, seizure cycles, seizure forecasting, surrogates

1 | INTRODUCTION

The pursuit of reliable epileptic seizure forecasting methods has often been compared to a journey on a long and winding road.[1–3] Over the past decade and a half, this endeavor has gained new momentum.[3–7] Progress has been fueled by advances in technology enabling electroencephalographic (EEG) recordings in patients with epilepsy on ultralong time scales,[8–24] increasing accessibility of electronic seizure diaries,[13–16,25–27] as well as by general developments in wearable devices measuring a broad spectrum of biosignals.[22,24,27–30] In parallel, key advances in computing hardware have substantially enhanced the capacity of signal analysis[8–14,18,22,23,27,31–35] and machine learning algorithms.[8,11,13,14,19–21,24,26,28–31,34,36–38] Further impetus was generated by approaches based on complex networks[33,39] and dynamical systems theory.[18,35] Importantly, these developments also revealed further evidence for cyclical variation in the occurrence rates of seizures and interictal epileptiform activity across daily, multidien, and yearly rhythms.[3,5,9–19,22,23,25,27,32] Rather than being a byproduct, these cycles are of central importance: their phases, and relationships between them, can help to forecast seizure risk.[3,5,9–11,13,14,18,19,23,25,27,32]

Studies on seizure cycles and seizure forecasting routinely use tailored statistical tests to compare their results against what can be expected under certain null hypotheses.[8–45] In particular for forecasting, rather than aiming at a binary decision on whether to reject the null hypothesis, these statistical tests are often used to estimate a performance baseline. If this baseline is surpassed, it is commonly concluded that the forecast is better-than-chance. Here, we challenge this conclusion by showing that random forecasters can perform better-than-chance, which may seem contradictory or paradoxical, but is not. The point is that better-than-chance is underspecified. It is commonly acknowledged that rejecting a null hypothesis does not prove the alternative hypothesis, let alone any specific hypothesis included in that alternative. However, it is equally important to recognize that rejecting a specific null hypothesis provides evidence against just one particular chance model. In no way does this imply that other null hypotheses, representing alternative chance models, would be rejected as well. It is therefore important—after rejecting some null hypothesis—to define and test further null hypotheses.

Key points

- Forecasting random seizure-time sequences by means of random predictions can yield areas under the receiver operating curve of up to .76.
- Seizure cycles extracted from random seizure-time sequences can appear highly significant when evaluated with the Rayleigh test.
- It is essential to test several complementary null hypotheses about cycle strength and seizure forecasting performance.
- Numerical, surrogate-based null-hypothesis tests offer the necessary flexibility and versatility for this purpose.
- Rigorous null-hypothesis testing can improve the characterization of cycles in epilepsy and support advances in seizure forecasting.

The aim of this paper is therefore to advance research on cycles in epilepsy and seizure forecasting by further developing rigorous statistical tests for both of these research areas. We first illustrate that rejecting single null hypotheses is insufficient to establish the existence of real cycles, forecastability of seizures, or genuine forecasting capacity. We argue that it is thus essential to always test several complementary null hypotheses about different sources of cycle strength and forecasting performance. We show that this aim can readily be achieved using numerical, surrogate-based null-hypothesis tests that offer the necessary flexibility and versatility for such rigorous statistical testing.

2 | MATERIALS AND METHODS

2.1 | Study design

Our study has two parts, one on seizure forecasting and one on seizure cycles. In both parts, we use synthetic random data generated by gamma distributions with different parameters. Under such controlled conditions the ground

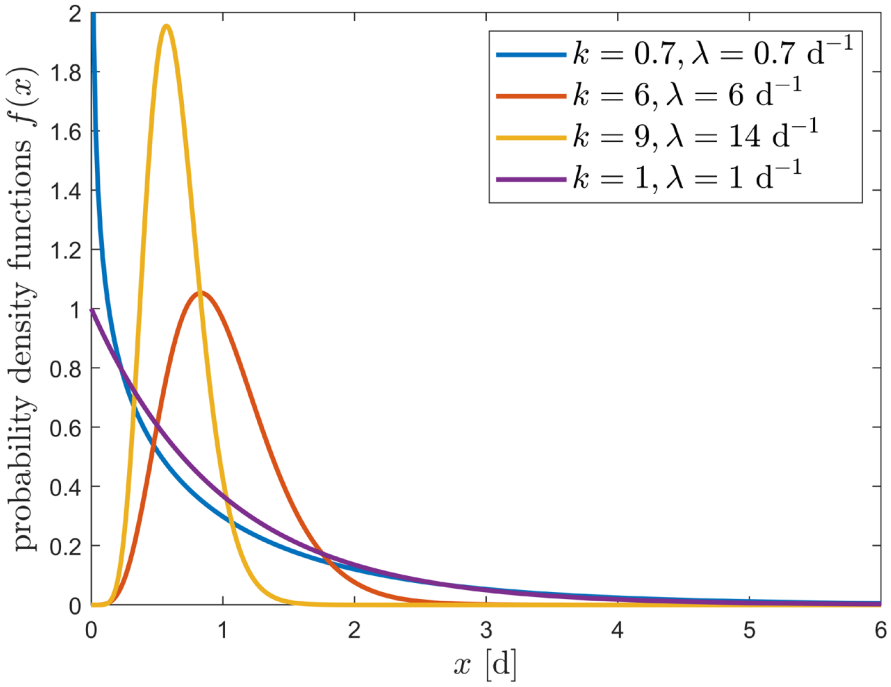

FIGURE 1 Gamma distributions: simple and flexible. Probability density functions of gamma distributions for different values of the shape parameter k and rate parameter λ are shown. The particular combinations of k and λ shown here are those used in our simulations. This includes an exponential distribution obtained for $k = 1$.

truth is known, that is, we can be certain whether a given null hypothesis is true or false. In the first part, we generate sequences of random seizure times and apply a simple random forecasting procedure to them. The forecasting procedure is defined ad hoc to be applicable prospectively in real time and to yield the standard performance metrics: sensitivity and fraction of time under alarm. In the second part, we quantify the cycle strength in random seizure times as a function of the cycle duration. For this purpose, we use the mean resultant length, also known as phase-locking value or synchronization index, and assess its significance using the Rayleigh test. At the core of our methodology in both parts are numerical, surrogate-based null hypothesis tests. These allow us to estimate the values of the sensitivity, fraction of time under alarm, and cycle strength expected under the chance model of the respective null hypothesis. These surrogate-based tests can also account for multiple testing on nonindependent data, which arises when assessing cycle strength across different durations. To contextualize and motivate this numerical approach, we briefly discuss analytical approaches based on the chance models of sensorless forecasting procedures and the Poisson process.

2.2 | Gamma distributions

Our synthetic data consist of random seizure times and random forecasts. We generate both using random samples from gamma distributions,[46] a two-parameter family of continuous probability distributions commonly used to model waiting times until an event occurs. By adjusting the shape parameter k and rate parameter λ of a gamma distribution within this family, one can generate data with different characteristics (Figure 1). Fixing $k = 1$ leads to the one-parameter subfamily of exponential distributions underlying the Poisson process with rate λ. Thus, the choice of gamma distributions combines parsimony and flexibility. The shape parameter k is dimensionless, and because we use gamma distributions to generate random time intervals, the rate parameter λ here has units of inverse time. For a given distribution, the mean and SD are $\mu = \frac{k}{\lambda}$ and $\sigma = \frac{\sqrt{k}}{\lambda}$, respectively, and have units of time. Time t is continuous and measured in days, denoted by d. For conciseness, "N random values" refers to an independent and identically distributed sample of size N.

2.3 | Generating random seizure times

To generate one realization of a random seizure-time sequence, we draw random intervals I_j, for $j = 1, \ldots N$ from a gamma distribution with parameters k_{sz} and λ_{sz}. Here "sz" stands for "seizures." We define the time of the zeroth seizure as $C_0 = 0$ d and determine the remaining times by $C_n = C_0 + \sum_{j=1}^{n} I_j$ for $n = 1, \ldots, N$. We use two parameter settings. ***A***: $k_{\mathrm{sz}} = 6, \lambda_{\mathrm{sz}} = 6\ \mathrm{d}^{-1}, N = 365$ and ***B***: $k_{\mathrm{sz}} = \frac{7}{10}, \lambda_{\mathrm{sz}} = \frac{7}{10}\ \mathrm{d}^{-1}, N = 365$ (Figure 1). Thereby, in both settings the mean of the gamma distribution is $\mu_{sz} = \frac{k_{\mathrm{sz}}}{\lambda_{\mathrm{sz}}} = 1$ d, and the expected time span between the zeroth and N-th seizure is $N \cdot \mu_{\mathrm{sz}} = 365$ d; in other words, the sequences have an expected duration of 1 year and

a mean rate of one seizure per day. However, the SD of the intervals between seizures is $\sigma_{sz} = \frac{\sqrt{k_{sz}}}{\lambda_{sz}} \approx .41$ d in setting ***A***, versus $\sigma_{sz} \approx 1.20$ d in setting ***B***. Moreover, the shape of the gamma distribution obtained for the parameters in ***B*** leads to a mixture of frequent short intervals and occasional long intervals. In contrast, for setting ***A*** typically neither very short nor very long intervals occur (Figures 1 and 2). These different degrees of dispersion in the random seizure-time sequences allow us to illustrate different effects with regard to seizure forecasting (setting ***A*** in Section 3.1) and seizure cycles (Setting ***B*** in Section 3.2).

2.4 | Random seizure forecasting procedure

In this section, we introduce a simple model of a seizure forecasting procedure that, due to its random nature, cannot be expected to yield any trustworthy forecast. First of all, the procedure has no access to any signals from the patient, such as EEG recordings or biosignals captured by wearables. The only input it receives is the seizure times $C_0, \ldots, C_N$ at the moments of their occurrences. Thereby, the procedure is prospective and could be used in real-time applications. The observation interval is bounded

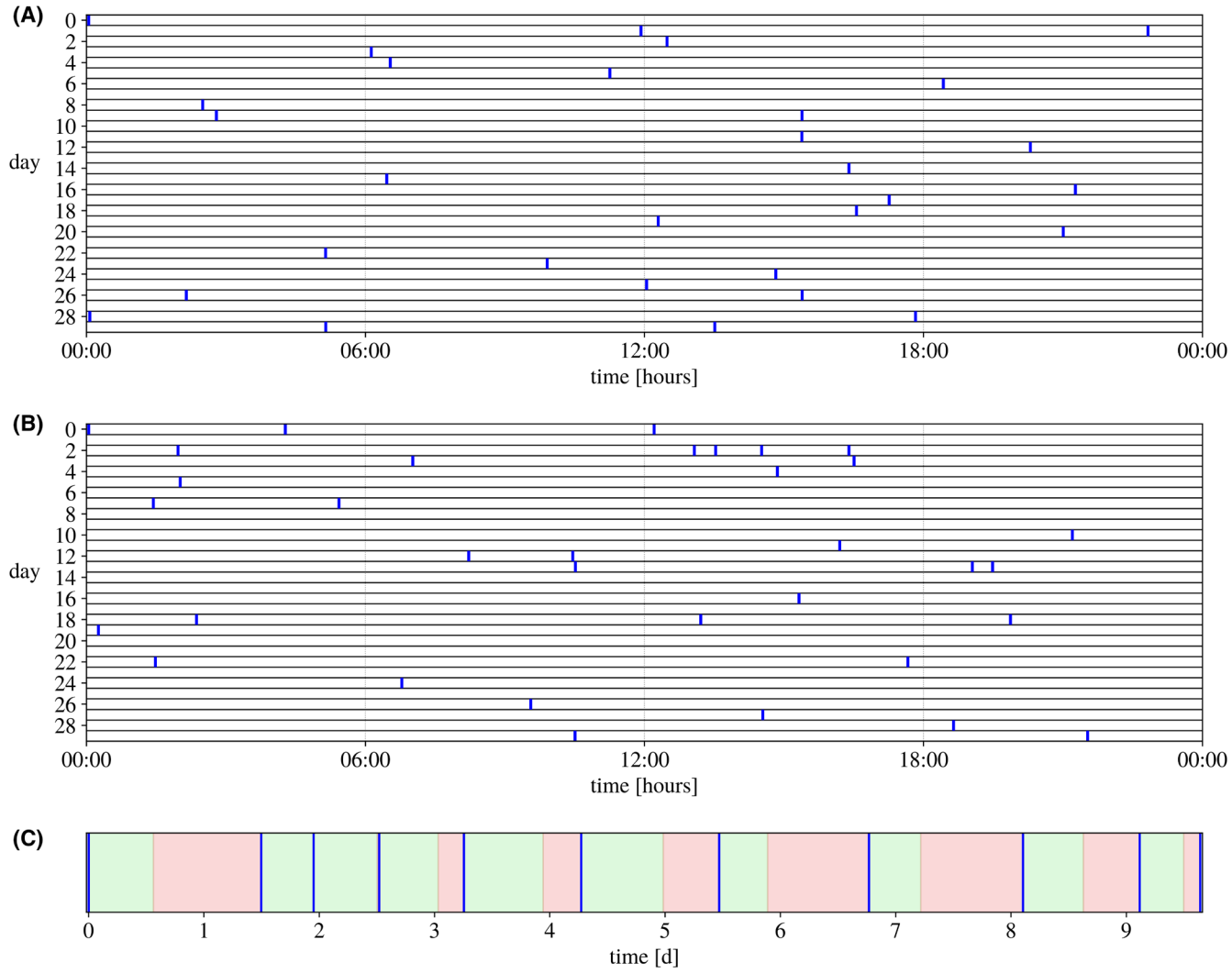

FIGURE 2 Example seizure-time sequences and illustration of random seizure forecasting procedure. (A) Seizure times generated with parameters of setting ***A***. Times of individual seizures are indicated by blue tick marks. Subsequent intervals of 24 h are wrapped into different rows for a compact representation. For an intuitive reading of this plot, we here assume that the time of the zeroth seizure C_0 is at midnight. In consequence, each row corresponds to a calender day, and the abscissa shows time of day. In real life, C_0 could be at any time of day. Note that this figure contains only those of the 365 seizures that occur during the interval of 30 days displayed here. (B) Same as panel A, but for seizure times generated with parameters of setting ***B***. (C) Illustration of the random seizure forecasting procedure using the zeroth to 10th seizure of panel A as an example. Blue tick marks indicate seizure times $C_0, C_1, \ldots, C_{10}$. Green and red periods indicate safe and alarm periods, respectively. In this example, the second and third seizures fall into the safe period (false negative forecasts). All other seizures fall into alarm periods (true positive forecasts).

by the times of the zeroth and last seizure, thus having a duration of $T = C_N - C_0$. Upon the occurrence of a seizure, the duration of the subsequent seizure-free interval is forecasted by drawing a random value from a gamma distribution with shape parameter k_{fc} and rate parameter λ_{fc} (Figure 2C). Here "fc" stands for "forecast." Recall that gamma distributions form a family of distributions, which vary substantially with their parameters. Therefore, as long as $k_{\mathrm{fc}} \neq k_{\mathrm{sz}}$ and/or $\lambda_{\mathrm{fc}} \neq \lambda_{\mathrm{sz}}$, different distributions are used to generate seizure times and forecasts.

To emphasize the prospective nature of the forecasting procedure, we describe the timing of each step and indicate which are triggered by seizure events in real time. Because the observation starts at time C_0, the zeroth seizure cannot be forecasted. Instead, at C_0 the count of true positive predictions s and the total time under alarm a are initialized to zero. The duration of the first seizure-free interval, denoted by Q_1, is then forecasted. This duration is used to define the first *safe period*, covering the interval $(C_0, Q_1]$. At this point, the forecasting procedure is in a state where two outcomes are possible, and it waits to observe which occurs. The first outcome is a false negative prediction, occurring if the first seizure falls within the first safe period, that is, $C_1 \leq Q_1$. In this case, the safe period is aborted at $t = C_1$. The second outcome is that the safe period is not interrupted by a seizure. In this case, the first *alarm period* is initiated immediately after the first safe period at time $t = Q_1$. The alarm is kept active until the first seizure occurs at $C_1 > Q_1$, no matter how long that takes. Therefore, false positive predictions are not possible, and the second outcome always results in a true positive prediction, incrementing the corresponding count s. Now that a seizure has occurred at time C_1, the first alarm period is terminated and the total time under alarm a is incremented by the first alarm's duration $a = a + (C_1 - Q_1)$. Thus, the seizure either aborts the safe period or terminates the alarm period, whichever is active. After resetting the procedure at the time of the first seizure, the duration of the second seizure-free interval is forecasted. This is done by again drawing a random value Q_2 from the gamma distribution with fixed k_{fc} and λ_{fc}. Accordingly, the interval $(C_1, C_1 + Q_2]$ defines the second safe period, evaluated as before. The procedure is iterated until the last seizure at time C_N. We should note that, for simplicity, we neglect the duration of the seizures. It could be incorporated by starting the next safe period not at C_1 but only after a short time span modeling the seizure's duration. Similarly, unlike in previous studies,[8,25,31,37,38,40–42] we do not require a minimal time between the onset of the alarm period and the seizure occurrence time for true positive predictions. This could be done by extending the end of each safe period by that minimal time.

Performance metrics are defined as follows. The count of true positive predictions is divided by the total number of seizures to quantify *sensitivity*: $S = \frac{s}{N}$. One could define specificity via true and false negative predictions given by the number of safe periods that are not aborted and that are aborted by a seizure, respectively. This is problematic, however, because the safe periods vary in length. We therefore assess *specificity* by the fraction of time under alarm,[8,10,14,19–25,27–30,42] defined by the total time under alarm divided by the duration of the observation interval: $A = \frac{a}{T}$. Both S and A are dimensionless and bounded within $[0, 1]$. A single measure of performance is defined by $F = S - A$ and is bounded between $(-1, 1]$. $F = -1$ is impossible, because $A = 1$ implies $S = 1$.

2.5 | Sensorless seizure forecasting procedures

Naïve predictors, such as moving average-based schemes,[26] provide an important performance benchmark. Sensorless forecasters[42] take this idea further; they have no access to any signals from the patient and are, moreover, uninformed about seizure times in the past or present. These forecasters might use some assumption about the distribution of interseizure intervals to constrain the distribution of intervals between alarms. However, they cannot link the timing of alarms to the most recent seizure, nor can they reset their state when a seizure occurs, as that information is unavailable. Therefore, even if the forecaster precisely knew the true distribution of interseizure intervals, it could only draw random values from this distribution and raise alarms at random times. Each alarm activates an alarm period that is later terminated according to predefined rules. Suppose that, in consequence, a fraction of $0 \leq A \leq 1$ of the time is covered by alarm periods. Averaged across independent runs of the procedure applied to a fixed sequence of seizure times, the fraction of seizures expected to occur during an active alarm is given by A. Therefore, the sensitivity has the same expected value as the fraction of time under alarm: $\langle S \rangle = \langle A \rangle$.[42]

2.6 | The null hypothesis of a Poisson process

A prominent null hypothesis about seizure forecasting procedures is that alarms are raised by a Poisson process with a constant rate λ.[10,21,24,31,36,38,40–42] This is a plausible null hypothesis and should be considered first. Loosely speaking, under this null hypothesis, denoted by $\mathcal{H}^{\mathrm{fc}}_{\mathrm{Pois}(\lambda)}$, the momentary probability of an alarm is

always the same. In more precise terms, for a Poisson process with rate λ, the number of alarms in any time interval of length y follows a Poisson distribution with the mean given by the product λy.[46] Furthermore, the Poisson process has no memory. Having observed whatever number of alarms in the past, one always expects a mean number of λy alarms in any future interval of length y. Finally, the two following statements are equivalent. Alarms are raised by a Poisson process with rate λ. Intervals between alarms are random values drawn from an exponential distribution with its rate parameter λ equal to the rate of the Poisson process.

Accordingly, under $\mathcal{H}^{\text{fc}}_{\text{Pois}(\lambda)}$ any input—such as signals from the patient—is irrelevant, as it would be ignored by the Poisson process. If, furthermore, seizures are not used to deactivate alarm periods, this results in a specific case of a sensorless algorithm. In this case, analytical approaches allow deriving equations for the expected values of the performance metrics.[40–42] For example, the expected sensitivity can be expressed as a function of the alarm rate λ and fixed durations of alarm periods, prediction horizons,[40,41] or detection intervals.[42] However, these analytical approaches cannot be applied to evaluate our random forecasting procedure. First, it is not sensorless. It disrupts the safe or alarm periods upon the occurrence of a seizure. Second, it has memory. After a transition from a safe period to an alarm period, it must wait for a seizure to return to the safe period. Only then does the next transition to an alarm period become possible. Thus, by construction of our forecasting procedure, the null hypothesis $\mathcal{H}^{\text{fc}}_{\text{Pois}(\lambda)}$ is false (see also Ozcan and Erturk,[31] Wang et al.,[36] Gao et al.,[38] Winterhalder et al.[41]). We therefore need other null hypotheses and ways to test them.

2.7 | Surrogate-based null hypotheses tests

The framework of surrogates offers a flexible way to test null hypotheses numerically[47–49] and can readily be applied to evaluate seizure forecasting procedures or cycles in epilepsy.[12,43–45] In general, a surrogate is constructed by randomizing some original data, either by shuffling the original values or by replacing them with random ones. The randomization is constrained so that the surrogate preserves those, and only those, properties of the original data that are consistent with their null hypothesis.

Consider a forecasting procedure for which you analyze a continuous data stream consisting of EEG recordings and biosignals measured from the patient. You use a method you believe is effective, although you cannot be certain. The seizure times are known so that the methods of Section 2.4 can be applied to determine safe and alarm periods. As results, you obtain the sensitivity S_{ori}, the fraction of time under alarm A_{ori}, and the corresponding performance value $F_{\text{ori}} = S_{\text{ori}} - A_{\text{ori}}$, where the subscripts stand for "original." As stated above, the null hypothesis $\mathcal{H}^{\text{fc}}_{\text{Pois}(\lambda)}$ is, by construction, false for our random forecasting procedure. However, we can test the closely related null hypothesis $\mathcal{H}^{\text{fc}}_{\text{expo}(\lambda)}$ that the forecasted durations of the safe periods are random values from an exponential distribution with rate parameter λ. To run this test, we replace the safe periods forecasted by the original procedure with surrogate safe periods whose durations are determined by random values from an exponential distribution with rate parameter λ_{sur}. Here "sur" stands for "surrogate." The resulting performance metrics are denoted by $S_{\text{sur},w}$, $A_{\text{sur},w}$, and $F_{\text{sur},w}$. The index w indicates that the process is repeated $w = 1, \ldots, W$ times. For every w, we use new random safe periods, that is, an independent surrogate realization. If $\mathcal{H}^{\text{fc}}_{\text{expo}(\lambda)}$ is false, F_{ori} is expected to be significantly higher than the $F_{\text{sur},w}$ values. For a significance level of $\alpha^{\text{fc}}_{\text{expo}(\lambda)}$, one has to use at least $W = 1/\alpha^{\text{fc}}_{\text{expo}(\lambda)} - 1$ surrogates. Using more surrogates increases the resolution of the $p^{\text{fc}}_{\text{expo}(\lambda)}$-value estimate. If for V of the W surrogates one finds $F_{\text{ori}} > F_{\text{sur},w}$, then $p^{\text{fc}}_{\text{expo}(\lambda)} = \frac{W-V+1}{W+1}$.

2.8 | Quantifying the strength of seizure cycles

We follow a common approach[11–19,22,25,27,32] and quantify the strength of cycles in seizure-time sequences by the mean resultant length, a measure known under alternative names, such as phase-locking value and synchronization index, among others (see Mardia and Jupp,[50] Andrzejak et al.,[51] and references therein). Given one realization of a sequence and cycle duration m to test for, we take the seizure times $C_{n=1,\ldots,N}$ modulo m and transfer them to circular data in the range $[0, 2\pi)$ using $\phi_n(m) = \left(C_n \bmod m\right) \cdot \frac{2\pi}{m}$. We exclude C_0, because the zeroth seizure is also excluded from the forecast. We then quantify the cycle strength by the mean resultant length $\overline{R}_N(m) = \frac{1}{N}\left|\sum_{n=1}^{n=N} e^{i\phi_n(m)}\right|$. Here $|\cdot|$ denotes the absolute value, and i is the imaginary unit. The upper limit $\overline{R}_N(m) = 1$ is reached if all $\phi_n(m)$ are identical. We assess the significance of $\overline{R}_N(m)$ by the Rayleigh test,[12,15–17,22,25] whose null hypothesis $\mathcal{H}_{\text{unif}(N)}$ is that the data underlying the mean resultant length are N random values from a uniform distribution on the circle.[50] Under this $\mathcal{H}_{\text{unif}(N)}$, the expected value of the mean resultant length approaches its lower bound of zero as N tends to infinity (see also Andrzejak et al.[51]).

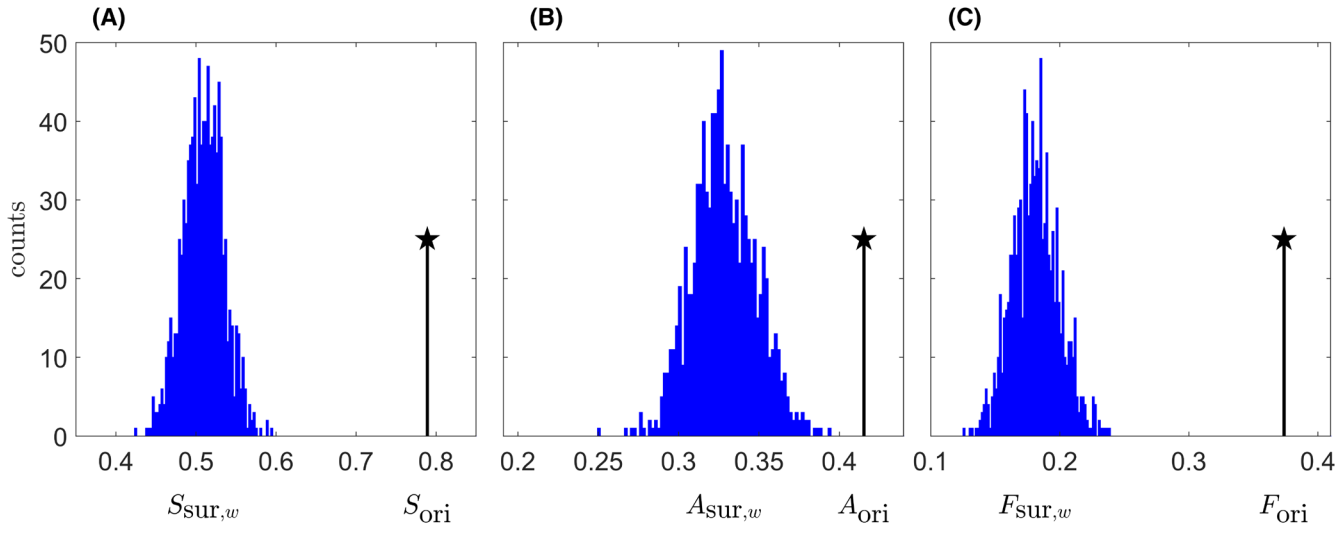

FIGURE 3 One random predictor outperforms another one. Results are shown for seizure times generated with the parameters of setting ***A***. (A) Black vertical line with star marks the value obtained for the original seizure forecasting procedure S_{ori} using $k_{\mathrm{fc}} = 9$ and $\lambda_{\mathrm{fc}} = 14\ \mathrm{d}^{-1}$. Blue: Histogram of $S_{\mathrm{sur},w=1,\ldots,W}$ values obtained for $W = 999$ surrogates testing the null hypothesis $\mathcal{H}^{\mathrm{fc}}_{\mathrm{expo}(\lambda)}$ at $\lambda = \lambda_{\mathrm{sur}} = .75$. We use a bin size of $\frac{1}{N}$ to account for the discrete nature of the sensitivity values. (B) Same as panel A, but here for $A_{\mathrm{sur},w}$, and A_{ori}. The bin size is adapted such that highest counts are comparable to those in panel A. (C) Same as panel B, but here for $F_{\mathrm{sur},w}$ and F_{ori}.

3 | RESULTS

3.1 | Application and evaluation of random seizure forecasting procedure

In this section, we apply the random seizure forecasting procedure to seizure-time sequences generated using parameter setting ***A***. The parameters of the forecasting procedure are set to $k_{\mathrm{fc}} = 9$ and $\lambda_{\mathrm{fc}} = 14\ \mathrm{d}^{-1}$, resulting in safe periods with mean duration $\mu_{\mathrm{fc}} \approx .64$ d and SD $\sigma_{\mathrm{fc}} \approx .21$ d. Accordingly, $k_{\mathrm{fc}} \neq k_{\mathrm{sz}}$; $\lambda_{\mathrm{fc}} \neq \lambda_{\mathrm{sz}}$; $\mu_{\mathrm{fc}} \neq \mu_{\mathrm{sz}}$; $\sigma_{\mathrm{fc}} \neq \sigma_{\mathrm{sz}}$, that is, we use different gamma distributions to generate seizure times and to forecast them (Figure 1). The values of k_{fc} and λ_{fc} are chosen ad hoc and not fine-tuned to produce the following results. To mimic real-world conditions, we show results for one fixed realization of the random seizure-time sequence and one fixed realization of the random seizure forecasting procedure. We find a sensitivity of $S_{\mathrm{ori}} = .79$ and a fraction of time under alarm of $A_{\mathrm{ori}} = .42$, and thereby $F_{\mathrm{ori}} = .37$ (Figure 3). Recall that for a sensorless forecasting procedure, S_{ori} and A_{ori} have the same expected value (Section 2.5). However, our random forecasting procedure is not sensorless, because we use the seizure times to reset it. Our example shows that in this case, $S_{\mathrm{ori}} > A_{\mathrm{ori}}$ can be obtained.

To test whether the results for $S_{\mathrm{ori}}, A_{\mathrm{ori}}$, and F_{ori} are consistent with the null hypothesis $\mathcal{H}^{\mathrm{fc}}_{\mathrm{expo}(\lambda)}$, that is, the durations of the safe periods are random values from an exponential distribution, we use the surrogate-based method of Section 2.7. To this end, we first have to set the rate parameter of the exponential distribution. One could argue for using $\lambda_{\mathrm{sur}} = \frac{14}{9}\ \mathrm{d}^{-1}$ so that the mean of the distribution $\mu_{\mathrm{sur}} = \frac{1}{\lambda_{\mathrm{sur}}}$ coincides with that of the original forecasting procedure $\mu_{\mathrm{fc}} = \frac{k_{\mathrm{fc}}}{\lambda_{\mathrm{fc}}} = \frac{9}{14}$ d. However, in real-world applications, the latter is not known. Estimating it by observing the durations of the original safe periods is not straightforward, because part of them is interrupted by seizures. Simulating real-world conditions, we therefore start with a first ad hoc guess of $\lambda_{\mathrm{sur}} = \frac{3}{4}\ \mathrm{d}^{-1}$. We run $W = 999$ realizations of the surrogate forecasting procedure, which is more than sufficient for a significance level of $\alpha^{\mathrm{fc}}_{\mathrm{expo}(\lambda)} = .05$. We find that F_{ori} is higher than all $F_{\mathrm{sur},w=1,\ldots,W}$ (Figure 3C). In consequence, for $\lambda = \lambda_{\mathrm{sur}} = \frac{3}{4}\ \mathrm{d}^{-1}$, we reject $\mathcal{H}^{\mathrm{fc}}_{\mathrm{expo}(\lambda)}$ with $p^{\mathrm{fc}}_{\mathrm{expo}(\lambda)} = .001$. However, the surrogates have not only lower sensitivity ($S_{\mathrm{sur},w} < S_{\mathrm{ori}}$; Figure 3A), but also a lower fraction of time under alarm ($A_{\mathrm{sur},w} < A_{\mathrm{ori}}$; Figure 3B). Therefore, the significant difference between the surrogate and original performance may be caused by using too low a value of λ_{sur}. To address this concern, we run the analysis for a range of λ_{sur} values. As the test is performed for a single surrogate at each λ_{sur} rather than for W realizations, we suppress the index w. Comparing the resulting point set $S_{\mathrm{sur}}(\lambda_{\mathrm{sur}})$ versus $A_{\mathrm{sur}}(\lambda_{\mathrm{sur}})$ with the point of S_{ori} versus A_{ori} in Figure 4A provides final compelling evidence against $\mathcal{H}^{\mathrm{fc}}_{\mathrm{expo}(\lambda)}$.

Rather than immediately concluding that the forecast procedure performs better-than-chance, one should take advantage of the flexibility of the surrogate-based null-hypothesis test. We therefore repeat the analysis underlying Figure 4A using gamma distributions with a variety of shape parameters k_{sur}, resulting in point

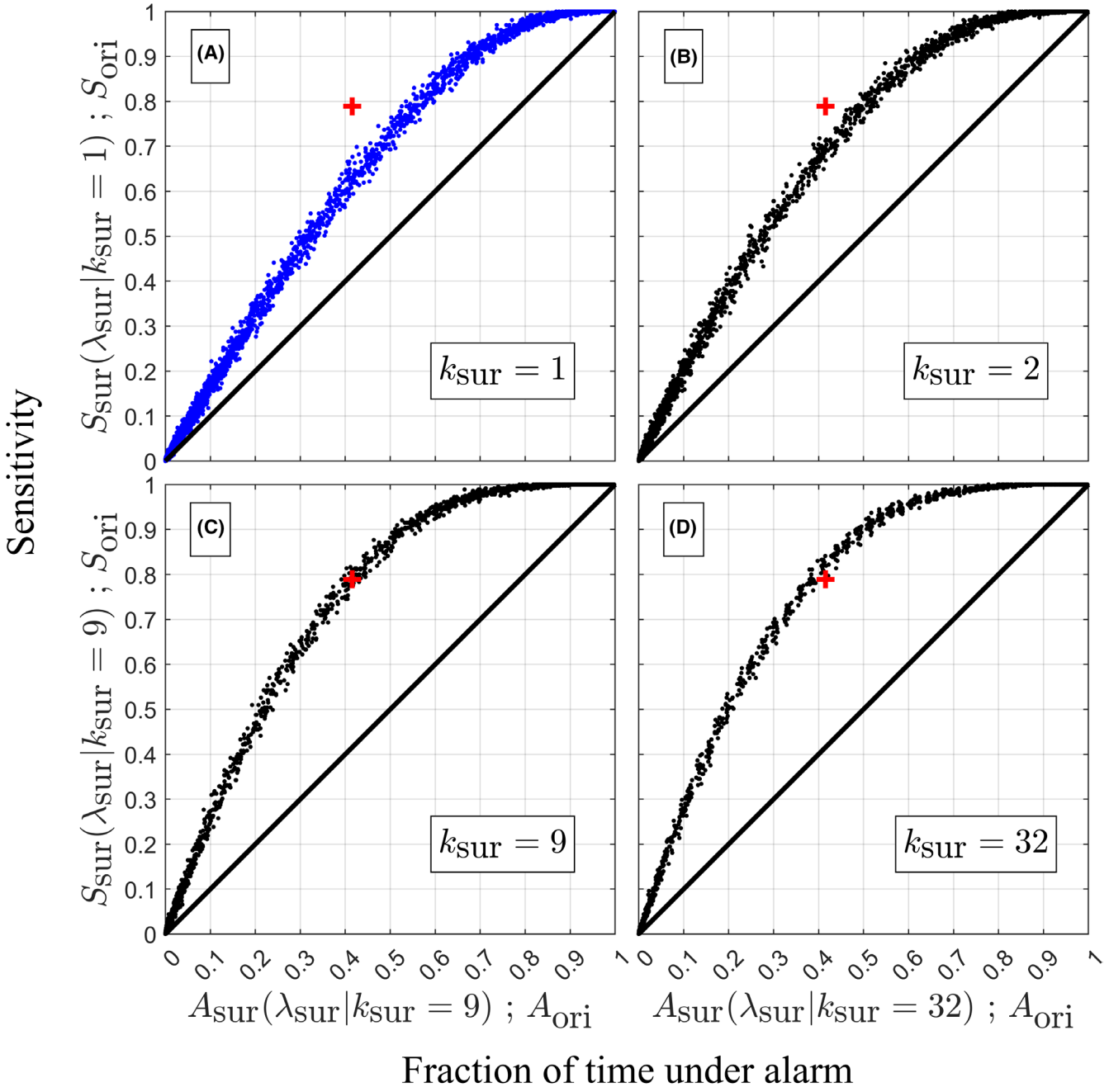

FIGURE 4 Our random predictor outperforms some—but not all—other random predictors. Results are shown for seizure times generated with the parameters of setting $\boldsymbol{A}$. (A) Red cross: Sensitivity S_{ori} versus fraction of time under alarm A_{ori} obtained for the original seizure forecasting procedure using $k_{fc} = 9$ and $\lambda_{fc} = 14\ d^{-1}$. Blue dots: $S_{sur}(\lambda_{sur})$ versus $A_{sur}(\lambda_{sur})$, testing the null hypothesis $\mathcal{H}^{fc}_{expo(\lambda)}$. The rate parameter λ_{sur} is varied in small steps so that both the sensitivity and fraction of time under alarm span their full range of [0, 1]. Recall that the gamma distribution with $k = 1$ is the exponential distribution. In this case, $\mathcal{H}^{fc}_{gam(k,\lambda)}$ and $\mathcal{H}^{fc}_{expo(\lambda)}$ coincide, and the blue dots in panel A corresponds to $S_{sur}(\lambda_{sur}|\, k_{sur} = 1)$ versus $A_{sur}(\lambda_{sur}|\, k_{sur} = 1)$. (B) Same as panel A, but black dots show $S_{sur}(\lambda_{sur}|\, k_{sur} = 2)$ versus $A_{sur}(\lambda_{sur}|\, k_{sur} = 2)$, testing the null hypothesis $\mathcal{H}^{fc}_{gam(k,\lambda)}$ using shape parameter $k_{sur} = 2$. (C, D) Same as panel B but for $k_{sur} = 9$ and $k_{sur} = 32$, respectively. In panels A and B, the original result (red cross) is outside of the point set of the surrogates. In particular, it is closer to the upper left corner of perfect performance ($S = 1$ and $A = 0$). Taken together, this provides compelling evidence against the respective null hypothesis. In panel C, the original result is inside the point set of the surrogates. In panel D, it is again outside, but is it now closer to the lower right corner of worst performance ($S = 0$ and $A = 1$). In both cases, the corresponding null hypothesis is thus not rejected. The areas under the curves fitted to the sets $S_{sur}(\lambda_{sur}|\, k_{sur})$ versus $A_{sur}(\lambda_{sur}|\, k_{sur})$ are .65, .69, .74, .76 for panels A, B, C, and D, respectively.

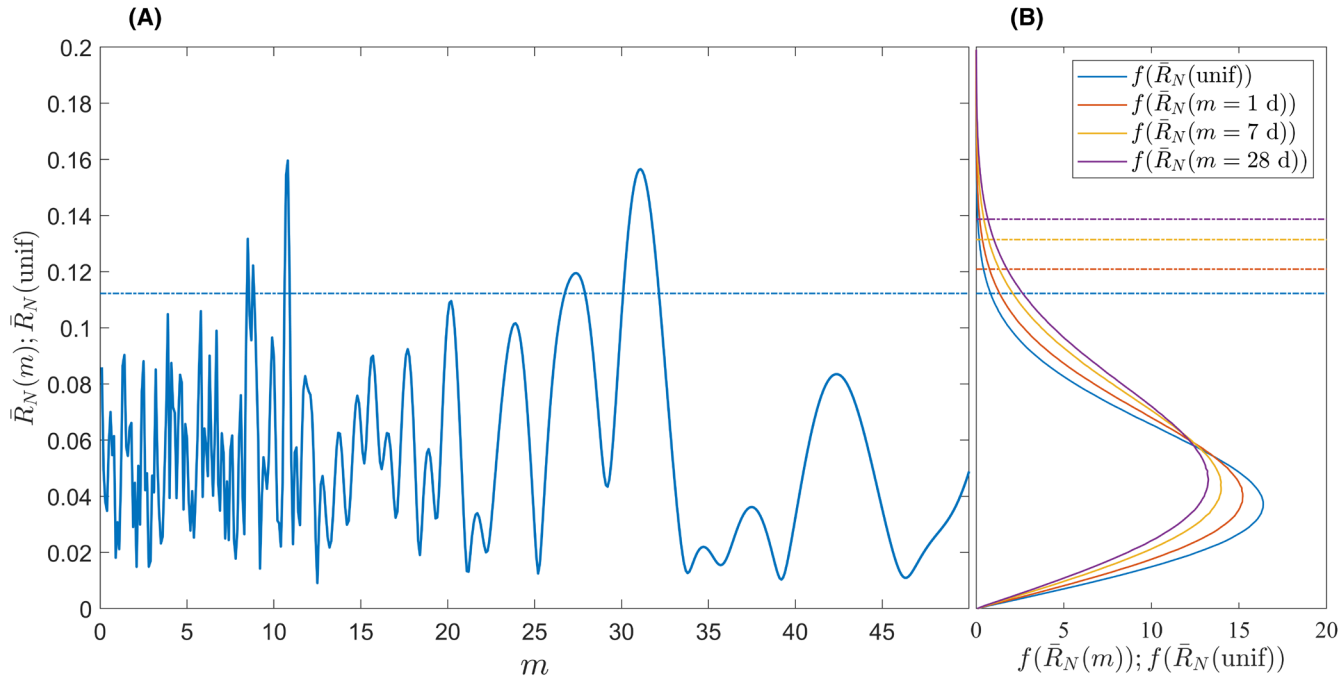

FIGURE 5 Are prominent seizure cycles in random seizure-time sequences significant? (A) Mean resultant length $\overline{R}_N(m)$ in dependence on the cycle duration m for an example realization of random seizure times generated using the parameters of setting ***B***. The horizontal line shows the critical value of the Rayleigh test at a significance level of $\alpha_{\text{unif}(N)} = .01$. (B) $f\left(\overline{R}_N(m)\right)$ for the m values indicated in the legend, along with $f\left(\overline{R}_N(\text{unif})\right)$. Recall that the latter represents the null distribution of the Rayleigh test. Note that we plot the probability density functions on the abscissa and the variables on the ordinate. To numerically estimate $f\left(\overline{R}_N(m = 1\ \text{d})\right)$, we generate 10^8 realizations of random seizure-time sequences. For each realization, the $\phi_n(m = 1\ \text{d})$ are extracted for $n = 1, \ldots N$, and the resulting $\overline{R}_N(m = 1\ \text{d})$ value is determined. The probability density function curve is a fine frequency polygon obtained from the 10^8 independent values of $\overline{R}_N(m = 1\ \text{d})$. The same procedure is used for $f\left(\overline{R}_N(m = 7\ \text{d})\right)$ and $f\left(\overline{R}_N(m = 28\ \text{d})\right)$, using the respective values of m. The same procedure is also used for $f\left(\overline{R}_N(\text{unif})\right)$. However, here the ϕ_n are not extracted from seizure times, but instead we use 10^8 realizations of $N = 365$ random values from the uniform distribution on the circle. Accordingly, in this case, there is no cycle duration variable m. The horizontal lines show the 99th percentile for each distribution. In each case, this corresponds to the critical value at a significance level of $\alpha^{\text{sz}}_{\text{gam}(k,\lambda),N,m} = .01$. By construction, the blue lines in panels A and B are at the same height.

sets $S_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ versus $A_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ (Figure 4B–D). In this way, we test the family of null hypotheses $\mathcal{H}^{\text{fc}}_{\text{gam}(k,\lambda)}$ that the durations of the safe periods are random values from a gamma distribution with parameters k and λ. For $k_{\text{sur}} = 2$, the set $S_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ versus $A_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ comes closer to the point S_{ori} versus A_{ori}, without reaching it yet (Figure 4B). For $k_{\text{sur}} = k_{\text{fc}} = 9$, the point S_{ori} versus A_{ori} is inside the set $S_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ versus $A_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ (Figure 4C). At $k_{\text{sur}} = 32$, the set $S_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ versus $A_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ has surpassed the point S_{ori} versus A_{ori} (Figure 4D). Thus, for $k_{\text{sur}} \geq k_{\text{ori}} = 9$, the null hypothesis $\mathcal{H}^{\text{fc}}_{\text{gam}(k,\lambda)}$ that the duration of the safe periods are random values from a gamma distribution with parameters k and λ is not rejected, correctly revealing that a chance model can reach—or even surpass—the performance of the original random seizure forecasting procedure. Of note, the areas under the curves fitted to the $S_{\text{sur}}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ versus $A_{\text{sur},w}\left(\lambda_{\text{sur}} | k_{\text{sur}}\right)$ sets are always higher than .5, the value commonly regarded as the chance level (Figure 4). Thus, although all curves reflect the performance metrics of random predictors, they surpass this threshold.

3.2 | Prominent cycles in random seizure-time sequences

In this section, we use setting ***B*** to generate seizure-time sequences and quantify the strength of cycles in dependence on the cycle duration m.[9,12,13,15,17,25] For the example sequence in Figure 2B, the seizures do not seem to be concentrated at any time of day. Across the entire sequence and using a cycle duration of 1 day, we get $\overline{R}_N(m = 1\ \text{d}) = .031$, resulting in a nonsignificant $p_{\text{unif}(N)} = .71$ for the Rayleigh test. Although the mean rate is one seizure per day, this is to be expected due to the high variance of intervals between seizures.

Calculating $\overline{R}_N(m)$ for a range of m, however, we do find prominent peaks, many of which are significant when assessed by the Rayleigh test (Figure 5A). Can we

conclude that these peaks reflect true cycles in the seizure times? No. First, calculating $\overline{R}_N(m)$ across a range of m is multiple testing. It is well known that if a null-hypothesis test is conducted multiple times using samples for which the null hypothesis is true and that are independent across tests, one expects to reject a proportion of tests given by the significance level α. One might consider adjusting the significance level using standard multiple-testing corrections. However, this is complicated in that the $\overline{R}_N(m)$ are not independent. They are all based on the same seizure-time sequence, and only the cycle duration m is varied. More importantly, as the following analysis shows, the null hypothesis of the Rayleigh test $\mathcal{H}_{\mathrm{unif}(N)}$ is false.

To show this, we have to first determine the likelihood of different $\overline{R}_N(m)$ values when calculated for $\phi_n(m)$ obtained from random seizure times C_n. For this purpose, we numerically estimate the probability density function $f\left(\overline{R}_N(m)\right)$ (Figure 5B). Here, not only m, but also $k_{\mathrm{sz}}, \lambda_{\mathrm{sz}}, N$ play the role of parameters that jointly determine the shape of $f\left(\overline{R}_N(m)\right)$. We furthermore determine the distribution $f\left(\overline{R}_N(\mathrm{unif})\right)$ expected under the null hypothesis $\mathcal{H}_{\mathrm{unif}(N)}$ of the Rayleigh test. Figure 5B shows marked differences between $f\left(\overline{R}_N(m)\right)$ and $f\left(\overline{R}_N(\mathrm{unif})\right)$ for various example cycle durations m. Hence, for our data $\mathcal{H}_{\mathrm{unif}(N)}$ is false. In particular, with increasing cycle duration m, the right tails of the $f\left(\overline{R}_N(m)\right)$ get heavier. Accordingly, for the parameters of setting $\boldsymbol{B}$ and using the Rayleigh test, the likelihood to detect significant cycles in fully random seizure-time sequences increases with cycle duration.

Consider the null hypothesis $\mathcal{H}^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$ that the $\overline{R}_N(m)$ values are determined from N seizure times C_n where the intervals between them are random values from a gamma distribution with parameters k and λ, and the times are taken modulo m. In this case, the null distribution is given by the previously introduced $f\left(\overline{R}_N(m)\right)$. One can thus use $f\left(\overline{R}_N(m)\right)$ to estimate the probability that a random sample from it has a value equal to or greater than a certain $\overline{R}_N(m)$.[12,16] This probability is the $p^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$ value of the result $\overline{R}_N(m)$ under the null hypothesis $\mathcal{H}^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$. The highest peak of the $\overline{R}_N(m)$ curve shown in Figure 5A is $\overline{R}_N(m = 10.8\ \mathrm{d}) = .160$, resulting in $p^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m} = .0013$. As a side note, because the $f\left(\overline{R}_N(m)\right)$ depend on m, the highest peak in the $\overline{R}_N(m)$ curve is not necessarily the one leading to the lowest $p^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$ value. For the example shown in Figure 5A, that happens to be the case. However, in general, the $p^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$ should be evaluated across all m to find the minimum. Evidently, this minimum arises from dependent tests across multiple values of m, and the following question arises. Under the null hypothesis $\mathcal{H}^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$, what is the probability that the minimal $p^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$ value across the $\overline{R}_N(m)$ curve is as low as or lower than the one observed? This probability directly represents the $p^{\mathrm{sz,corrected}}_{\mathrm{gam}(k,\lambda),N,m}$ value corrected for multiple testing. It can be determined numerically by generating multiple realizations of seizure-time sequences, computing the resulting $\overline{R}_N(m)$ curves, and extracting the minimum $p^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$ values. For the example in Figure 5A, we use 10 000 realizations to estimate $p^{\mathrm{sz,corrected}}_{\mathrm{gam}(k,\lambda),N,m} = .217$, correctly revealing that the $\overline{R}_N(m)$ curve and its most prominent peak at 10.8 d are consistent with the null hypothesis $\mathcal{H}^{\mathrm{sz}}_{\mathrm{gam}(k,\lambda),N,m}$. We note that changing k_{sz}, λ_{sz}, or N can readily produce higher overall $\overline{R}_N(m)$ values than those in Figure 5A. However, $p^{\mathrm{sz,corrected}}_{\mathrm{gam}(k,\lambda),N,m}$ still reveals these values to be insignificant.

4 | DISCUSSION

In Section 3.1, we use a random procedure to forecast sequences of random seizure times. Nonetheless, the resulting area under the curve spanned by the sensitivity versus the fraction of time under alarm is clearly above .5, the value generally considered as the chance level. Under our controlled conditions, the cause of this deviation is straightforward to identify. The forecasting procedure is not sensorless; each seizure either aborts the safe period or terminates the alarm period, whichever is active. Consequently, by construction of our forecasting procedure, the sensitivity and the fraction of time under alarm have different expected values, such that the area under their curve exceeds .5. In real-world conditions, other properties of the forecasting procedure may lead to analogous deviations from the chance level, the causes of which, however, can be difficult to identify. This problem is not specific to the sensitivity, the fraction of time under alarm, and the area under their curve. Any discriminative performance metrics should detect these deviations. A good and common practice to address this problem is to use surrogate-based approaches to estimate baselines for seizure forecasting performance metrics.[8,10,13,14,18,20–25,27,28,30,33,35,37] Nevertheless, the full potential of these approaches, which derives from their flexibility and versatility, is often not fully exploited. In our example of Section 3.1, flexibility arises because the two-parameter family of gamma distributions allows us to define and test a family of null hypotheses. Using this flexibility by testing hypotheses for different combinations of the two parameters was key to revealing the random nature of our forecasting procedure. If gamma distributions had not been suitable, one could have relied on the surrogate-based approach's versatility; any non-negative distribution—each representing a different null

hypothesis—could have been tested instead of gamma distributions.

As an alternative to using parametric distributions, surrogates can be generated by resampling the original data. In either case, it is essential to be aware of all implicit assumptions of the null hypothesis. For example, randomly shuffling the intervals between seizures without any constraints assumes stationarity and statistical independence between successive interval lengths. Consider for instance a year-long sequence of interseizure intervals. Suppose that short and long intervals are statistically more likely followed by short and long intervals, respectively. Further suppose that during the first 6 months the overall mean of the intervals is shorter than during the second 6 months. Finally, suppose that, except for this temporal correlation and nonstationarity, the sequence of intervals is random. Because unconstrained random shuffling destroys such temporal correlation and nonstationarity, the surrogate null hypothesis is false. Of note, one of these properties would be sufficient to render it false. Consequently, the surrogate performance metrics cannot be expected to match the original ones. Again, the versatility of surrogates comes into play. For instance, by constraining the shuffling to preserve the temporal correlations in the interval sequence, the assumption of statistical independence between successive intervals can be dropped.

In Section 3.2, we use the mean resultant length to estimate the strength of cycles for random seizure times taken modulo the cycle duration. We demonstrate that the Rayleigh test null hypothesis is generally invalid in this case, highlighting the need for caution when taking plausible null hypotheses for granted. Taking random seizure times modulo the cycle duration certainly leads to random circular data. However, random circular data of sample size N can be inconsistent with a random sample of size N from a uniform distribution on the circle, that is, the null hypothesis of the Rayleigh test. One might argue that the null distributions in Figure 5B are overall not very different from each other. However, their tails are markedly different. This is critical in the context of hypothesis testing. In simplified terms, a p -value tells you how far you reach into the tail of your null distribution. Given a certain result, it can be highly significant with regard to one null distribution and nonsignificant with regard to another. In other words, the same result can lead to a true positive rejection of one null hypothesis and a true negative rejection of another null hypothesis. For our data, the null hypothesis of the Rayleigh test is false and should be rejected. The same holds true for the omnibus Hodges–Ajne test or other tests that have the null hypothesis in common with the Rayleigh test but use other statistics to test it. The flexibility of surrogate-based approaches is again key not only for testing other null hypotheses but also for correcting for multiple nonindependent tests across different cycle lengths. Given the growing evidence for coexisting cycles at various time scales,[3,5,9–19,22,23,25,27,32] surrogate-based approaches can thus be crucial for discriminating real cycles from spurious ones. A complementary approach to ours is taken by Leguia et al.[12] Whereas our study uses controlled conditions to generate random seizure-time sequences that are stationary and lack true cycles, Leguia et al. consider synthetic random data constructed to be nonstationary and to contain multidien cycles. It is shown that surrogates are also essential for detecting both the presence and the correct duration of cycles.

An intriguing direction for future work is to develop surrogates to represent not only clinical but also subclinical seizures, with regard to both forecasting and cycle strength. For our synthetic data, all seizures are treated as equivalent events. However, it is straightforward to generalize our approach to marked point processes. Each seizure event would have not only a time stamp but also a mark. The mark could be categorical, encoding whether a seizure is clinical or subclinical, or it could be continuous-valued, encoding the magnitude of the seizure. Different stochastic processes could be used to generate the time stamps and marks of the seizure surrogates.

5 | CONCLUSIONS

When assessing the results of studies on cycles in epilepsy and seizure forecasting, it is essential to test several complementary null hypotheses about these results. Numerical, surrogate-based null-hypothesis tests offer the necessary flexibility and versatility for this purpose. A potential concern is that it may always be possible to find some null hypothesis that can explain a given observed result, potentially making this approach overly conservative. However, if real cycles in epilepsy exist, seizures are forecastable, and algorithms possess genuine forecasting capacity, any reasonable combination of null hypotheses can be expected to be rejected. If, because of limitations in data quantity or quality, one fails to reject null hypotheses that are actually false, these tests can nonetheless provide useful information. In addition to benchmarks derived from naïve predictors,[26] null-hypothesis tests can yield performance baselines, allowing discrimination between promising and less promising approaches.

AUTHOR CONTRIBUTIONS

Ralph G. Andrzejak designed the study. Ralph G. Andrzejak and Martin Brešar performed the analysis. All authors evaluated and interpreted the results. All authors drafted the manuscript and edited the final version.

ACKNOWLEDGMENTS

R.G.A. acknowledges funding from the Spanish Ministry of Science and Innovation and the State Research Agency (grant No. PID2020-118196GBI00/MICIU/AEI/10.13039/501100011033). M.B. acknowledges funding from the Slovenian Research and Innovation Agency (research core funding No. P2-0001). R.R. is supported by the UK National Institute of Health and Care Research (NIHR; CL-2023-17-004). P.F.V. is supported by the NIHR Development and Skills Enhancement Award (NIHR305718). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

CONFLICT OF INTEREST STATEMENT

None of the authors has any conflict of interest to disclose. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

DATA AVAILABILITY STATEMENT

The source code needed to run the analysis underlying our results is available at https://doi.org/10.34810/data2899.

ORCID

Ralph G. Andrzejak https://orcid.org/0000-0002-7400-2098
Martin Brešar https://orcid.org/0000-0002-1674-3314
Mark P. Richardson https://orcid.org/0000-0001-8925-3140
Pedro F. Viana https://orcid.org/0000-0003-0861-8705
Richard Rosch https://orcid.org/0000-0002-0316-5818

REFERENCES

1. Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. Brain. 2007;130(2):314–33.
2. Mormann F, Andrzejak RG. Seizure prediction: making mileage on the long and winding road. Brain. 2016;139(6):1625–7.
3. Baud MO, Proix T, Gregg NM, Brinkmann BH, Nurse ES, Cook MJ, et al. Seizure forecasting: bifurcations in the long and winding road. Epilepsia. 2023;64:S78–S98.
4. Kuhlmann L, Lehnertz K, Richardson MP, Schelter B, Zaveri HP. Seizure prediction—ready for a new era. Nat Rev Neurol. 2018;14(10):618–30.
5. Karoly PJ, Rao VR, Gregg NM, Worrell GA, Bernard C, Cook MJ, et al. Cycles in epilepsy. Nat Rev Neurol. 2021;17(5):267–84.
6. Andrzejak RG, Zaveri HP, Schulze-Bonhage A, Leguia MG, Stacey WC, Richardson MP, et al. Seizure forecasting: where do we stand? Epilepsia. 2023;64:S62–S71.
7. Lehnertz K, Broehl T, von Wrede R. Epileptic-network-based prediction and control of seizures in humans. Neurobiol Dis. 2023;181:106098.
8. 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(6):563–71.
9. 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(4):1066–78.
10. Karoly PJ, Ung H, Grayden DB, Kuhlmann L, Leyde K, Cook MJ, et al. The circadian profile of epilepsy improves seizure forecasting. Brain. 2017;140(8):2169–82.
11. Baud MO, Kleen JK, Mirro EA, Andrechak JC, King-Stephens D, Chang EF, et al. Multi-day rhythms modulate seizure risk in epilepsy. Nat Commun. 2018;9(1):88.
12. Leguia MG, Rao VR, Kleen JK, Baud MO. Measuring synchrony in bio-medical timeseries. Chaos. 2021;31:013138.
13. Leguia MG, Andrzejak RG, Rummel C, Fan JM, Mirro EA, Tcheng TK, et al. Seizure cycles in focal epilepsy. JAMA Neurol. 2021;78(4):454–63.
14. Leguia MG, Rao VR, Tcheng TK, Duun-Henriksen J, Kjær TW, Proix T, et al. Learning to generalize seizure forecasts. Epilepsia. 2023;64:S99–S113.
15. Viana PF, Duun-Henriksen J, Glasstëter M, Dümpelmann M, Nurse ES, Martins IP, et al. 230 days of ultra long-term subcutaneous EEG: seizure cycle analysis and comparison to patient diary. Ann Clin Transl Neurol. 2021;8(1):288–93.
16. Karoly PJ, Goldenholz DM, Freestone DR, Moss RE, Grayden DB, Theodore WH, et al. Circadian and circaseptan rhythms in human epilepsy: a retrospective cohort study. Lancet Neurol. 2018;17(11):977–85.
17. Gregg NM, Nasseri M, Kremen V, Patterson EE, Sturges BK, Denison TJ, et al. Circadian and multiday seizure periodicities, and seizure clusters in canine epilepsy. Brain Commun. 2020;2(1):fcaa008.
18. Maturana MI, Meisel C, Dell K, Karoly PJ, D'Souza W, Grayden DB, et al. Critical slowing down as a biomarker for seizure susceptibility. Nat Commun. 2020;11(1):2172.
19. Stirling RE, Maturana MI, Karoly PJ, Nurse ES, McCutcheon K, Grayden DB, et al. Seizure forecasting using a novel subscalp ultra-long term EEG monitoring system. Front Neurol. 2021;12:713794.
20. Viana PF, Pal Attia T, Nasseri M, Duun-Henriksen J, Biondi A, Winston JS, et al. Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: individualized intrapatient models. Epilepsia. 2023;64:S124–S133.
21. Pal Attia T, Viana PF, Nasseri M, Duun-Henriksen J, Biondi A, Winston JS, et al. Seizure forecasting using minimally invasive, ultra-long-term subcutaneous EEG: generalizable cross-patient models. Epilepsia. 2023;64:S114–S123.
22. Gregg NM, Pal Attia T, Nasseri M, Joseph B, Karoly P, Cui J, et al. Seizure occurrence is linked to multiday cycles in diverse physiological signals. Epilepsia. 2023;64(6):1627–39.

23. Khambhati AN, Chang EF, Baud MO, Rao VR. Hippocampal network activity forecasts epileptic seizures. Nat Med. 2024;30(10):2787–90.
24. Nasseri M, Stirling RE, Viana PF, Cui J, Nurse E, Karoly PJ, et al. Forecasting epileptic seizures with wearable devices: a hybrid short-and long-horizon pseudo-prospective approach. Epilepsia. 2025;66(9):3293–308. https://doi.org/10.1111/epi.18466
25. Karoly PJ, Cook MJ, Maturana M, Nurse ES, Payne D, Brinkmann BH, et al. Forecasting cycles of seizure likelihood. Epilepsia. 2020;61(4):776–86.
26. Chang CY, Moss R, Westover MB, Goldenholz DM. Rigorous evaluation of five models for e-diary-only seizure forecasting: retrospective and prospective datasets do not outperform the napkin method. Epilepsia. 2025. https://doi.org/10.1111/epi.18677
27. Xiong W, Stirling RE, Payne DE, Nurse ES, Kameneva T, Cook MJ, et al. Forecasting seizure likelihood from cycles of self-reported events and heart rate: a prospective pilot study. EBioMedicine. 2023;93:104656. https://doi.org/10.1016/j.ebiom.2023.10465
28. Meisel C, El Atrache R, Jackson M, Schubach S, Ufongene C, Loddenkemper T. Machine learning from wristband sensor data for wearable, noninvasive seizure forecasting. Epilepsia. 2020;61(12):2653–66.
29. Ding TY, Gagliano L, Jahani A, Toffa DH, Nguyen DK, Bou Assi E. Epileptic seizure forecasting with wearable-based nocturnal sleep features. Epilepsia Open. 2024;9(5):1793–805.
30. Halimeh M, Jackson M, Loddenkemper T, Meisel C. Training size predictably improves machine learning-based epileptic seizure forecasting from wearables. Neuroscience Informatics. 2025;5(1):100184.
31. Ozcan AR, Erturk S. Seizure prediction in scalp EEG using 3D convolutional neural networks with an image-based approach. IEEE Trans Neural Syst Rehabil Eng. 2019;27(11):2284–93.
32. Kashyap A, Müller P, Miron G, Meisel C. Critical dynamics and interictal epileptiform discharge: a comparative analysis with respect to tracking seizure risk cycles. Front Netw Physiol. 2024;4:1420217.
33. Rings T, von Wrede R, Lehnertz K. Precursors of seizures due to specific spatial-temporal modifications of evolving large-scale epileptic brain networks. Sci Rep. 2019;9(1):10623.
34. Kalousios S, Müller J, Yang H, Eberlein M, Uckermann O, Schackert G, et al. ECG-based epileptic seizure prediction: challenges of current data-driven models. Epilepsia Open. 2025;10(1):143–54.
35. Wilkat T, Rings T, Lehnertz K. No evidence for critical slowing down prior to human epileptic seizures. Chaos. 2019;29(9):091104.
36. Wang Y, Cui W, Yu T, Li X, Liao X, Li Y. Dynamic multi-graph convolution-based channel-weighted transformer feature fusion network for epileptic seizure prediction. IEEE Trans Neural Syst Rehabil Eng. 2023;31:4266–77.
37. Lopes F, Pinto MF, Dourado A, Schulze-Bonhage A, Dümpelmann M, Teixeira C. Addressing data limitations in seizure prediction through transfer learning. Sci Rep. 2024;14(1):14169.
38. Gao Y, Liu A, Wang L, Qian R, Chen X. A self-interpretable deep learning model for seizure prediction using a multi-scale prototypical part network. IEEE Trans Neural Syst Rehabil Eng. 2023;31:1847–56.
39. Fruengel R, Bröhl T, Rings T, Lehnertz K. Reconfiguration of human evolving large-scale epileptic brain networks prior to seizures: an evaluation with node centralities. Sci Rep. 2020;10(1):21921.
40. Maiwald T, Winterhalder M, Aschenbrenner-Scheibe R, Voss HU, Schulze-Bonhage A, Timmer J. Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. Physica D: Nonlinear Phenomena. 2004;194(3–4):357–68.
41. Winterhalder M, Maiwald T, Voss H, 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(3):318–25.
42. Snyder DE, Echauz J, Grimes DB, Litt B. The statistics of a practical seizure warning system. J Neural Eng. 2008;5(4):392.
43. 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(1):010901.
44. 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(6):061915.
45. Andrzejak RG, Chicharro D, Elger CE, Mormann F. Seizure prediction: any better than chance? Clin Neurophysiol. 2009;120(8):1465–78.
46. DeGroot MH, Schervish MJ. Probability and statistics. Fourth ed. Boston: Pearson; 2012.
47. Theiler J, Eubank S, Longtin A, Galdrikian B, Farmer JD. Testing for nonlinearity in time series: the method of surrogate data. Physica D Nonlinear Phenom. 1992;58(1–4):77–94.
48. Andrzejak RG, Kraskov A, Stögbauer H, Mormann F, Kreuz T. Bivariate surrogate techniques: necessity, strengths, and caveats. Phys Rev E. 2003;68(6):066202.
49. Lancaster G, Iatsenko D, Pidde A, Ticcinelli V, Stefanovska A. Surrogate data for hypothesis testing of physical systems. Phys Rep. 2018;748:1–60.
50. Mardia KV, Jupp PE. Directional statistics. New Jersey: Wiley; 2000.
51. Andrzejak RG, Espinoso A, García-Portugués E, Pewsey A, Epifanio J, Leguia MG, et al. High expectations on phase locking: better quantifying the concentration of circular data. Chaos Interdiscip J Nonlinear Sci. 2023;33:091106.

How to cite this article: Andrzejak RG, Brešar M, Richardson MP, Viana PF, Rosch R. Are seizure forecasts and cycles better than chance? What chance? Epilepsia. 2026;00:1–13. https://doi.org/10.1002/epi.70158