Received: 11 July 2022 | Revised: 12 November 2022 | Accepted: 17 November 2022

DOI: 10.1111/epi.17471

Epilepsia®

RESEARCH ARTICLE

Flexible realistic simulation of seizure occurrence recapitulating statistical properties of seizure diaries

Daniel M. Goldenholz[1,2] | **M. Brandon Westover**[1,2,3,4]

[1]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

[2]Department of Neurology, Harvard Medical School, Boston, Massachusetts, USA

[3]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA

[4]McCance Center for Brain Health, Boston, Massachusetts, USA

Correspondence
Daniel M. Goldenholz, BIDMC Epilepsy Division, Baker 5, Boston, MA 02215, USA.
Email: daniel.goldenholz@bidmc.harvard.edu

Funding information
American Academy of Sleep Medicine Foundation, Grant/Award Number: AASM Foundation Strategic Research Award; American Federation for Aging Research, Grant/Award Number: Breakthroughs in Gerontology Grant; Glenn Foundation for Medical Research; National Institute of Neurological Disorders and Stroke, Grant/Award Number: KL2TR002542, R01AG073410, R01NS102190, R01NS102574, R01NS107291, RF1AG064312 and RF1NS120947; National Science Foundation, Grant/Award Number: 2014431

Abstract

Objective: A realistic seizure diary simulator is currently unavailable for many research needs, including clinical trial analysis and evaluation of seizure detection and seizure-forecasting tools. In recent years, important statistical features of seizure diaries have been characterized. These include (1) heterogeneity of individual seizure frequencies, (2) the relation between average seizure rate and standard deviation, (3) multiple risk cycles, (4) seizure clusters, and (5) limitations on inter-seizure intervals. The present study unifies these features into a single model.

Methods: Our approach, Cyclic Heterogeneous Overdispersed Clustered Open-source L-relationship Adjustable Temporally limited E-diary Simulator (CHOCOLATES) is based on a hierarchical model centered on a gamma Poisson generator with several modifiers. This model accounts for the aforementioned statistical properties. The model was validated by simulating 10 000 randomized clinical trials (RCTs) of medication to compare with 23 historical RCTs. Metrics of 50% responder rate (RR50) and median percent change (MPC) were evaluated. We also used CHOCOLATES as input to a seizure-forecasting tool to test the flexibility of the model. We examined the area under the receiver-operating characteristic (ROC) curve (AUC) for test data with and without cycles and clusters.

Results: The model recapitulated typical findings in 23 historical RCTs without the necessity of introducing an additional "placebo effect." The model produced the following RR50 values: placebo: 17 ± 4%; drug 38 ± 5%; and the following MPC values: placebo: 13 ± 6%; drug 40 ± 4%. These values are similar to historical data: for RR50: placebo, 21 ± 10%, drug: 43 ± 13%; and for MPC: placebo: 17 ± 10%, drug: 41 ± 11%. The seizure forecasts achieved an AUC of 0.68 with cycles and clusters, whereas without them the AUC was 0.51.

Significance: CHOCOLATES represents the most realistic seizure occurrence simulator to date, based on observations from thousands of patients in different contexts. This tool is open source and flexible, and can be used for many applications, including clinical trial simulation and testing of seizure-forecasting tools.

KEYWORDS
epilepsy, seizure diary, simulation, statistics

1 | INTRODUCTION

Seizure diaries represent a record of the most salient symptom of epilepsy and are to date the most practical proxy for degree of illness.[1] Having a realistic simulation of seizure diaries would be useful in many areas of epilepsy research. In the domain of randomized controlled trials (RCTs), simulations could be used for statistical power calculations,[2–5] assessment of different trial parameters,[3] outcome metrics,[4,6,7] and the impact of enrolling different types of patients on trial success.[8] For seizure detection, simulation could help assess risk levels for different devices[9,10] and algorithms.[11] For neuromodulation devices,[12] a realistic simulator could help facilitate optimization of stimulation parameters. For seizure forecasting,[13–15] simulation is critical to develop a "null hypothesis" upon which new forecasting tools can be tested. For prevention of sudden unexpected death in epilepsy (SUDEP),[16] a realistic simulator would provide developers of potential alerting devices and treatments with a tool to plan studies and improve safety.

Therefore, there have been several attempts to model seizure diaries over the years.[17,18] Most models have been based on a single small data set, or even no data at all. Often a Poisson model was assumed, which although easy to compute, is known to poorly represent seizure diaries. Several other studies use variations of Poisson models or negative binomial models, although typically these models encompass only one or two statistical features of diaries (for more details about prior models, see this study[18]). In the past few years, larger data sets have become available,[19] and several additional statistical features across different seizure diaries have been well described. These features are (1) statistical characteristics of the large-scale heterogeneity in average seizure rates across subjects,[19] (2) a strong relationship between average seizure rate and standard deviation across patients and data sets (the "L-relationship"),[20] (3) multiple coexisting risk cycles,[21,22] (4) seizure clusters,[23,24] and (5) limitations on inter-seizure intervals.[19,25,26] Thus far no simulator has incorporated all these features into a unified model.

Our approach, Cyclic Heterogeneous Overdispersed Clustered Open-source L-relationship Adjustable Temporally-limited E-diary Simulator, or CHOCOLATES, accounts for each of the above-mentioned properties. It produces synthetic diaries that mimic features of a population of people with epilepsy at the individual and population levels.

Key points

- We introduce a new open-source simulation tool called Cyclic Heterogeneous Overdispersed Clustered Open-source L-relationship Adjustable Temporally limited E-diary Simulator (CHOCOLATES).
- CHOCOLATES accounts for five, well-documented statistical features of seizure diaries.
- CHOCOLATES represents the most realistic seizure diary simulator to date.
- CHOCOLATES can recapitulate historical clinical trial results without a "placebo effect."
- CHOCOLATES can be used to test ideas about seizure forecasting.

2 | METHODS

2.1 | Model architecture

The simulator design is summarized in Figure 1, with example data in Figure 2. A complete description of the mathematics of the model can be found in Appendix A. Briefly, a seizure frequency is chosen randomly for the individual, which is then used to generate parameters for the base risk rate generator. This base risk is modulated by randomly selected risk cycles for that individual. The modulated risk rate is used to produce *diary A*. The timing of actual seizures is used to modify the modulated risk to allow for clusters. This new modulated cluster risk is used to produce *diary B*. If there are samples of *diary B* that have too many seizures based on the maximum limitations, those are truncated to form the *final diary*.

The monthly seizure frequency is chosen randomly with a log-normal distribution, tuned to produce a distribution mimicking a population of 1186 subjects.[19] The main seizure diaries are produced via a modified version of a gamma-Poisson distribution (equivalent to a negative binomial[27]). The gamma-Poisson parameters were optimized using a modified stochastic gradient descent approach to account for an appropriate seizure-rate distribution (feature #1) and the L-relationship[20] (feature #2). A modulation signal is composed by randomly selecting a set of sine waves from a distribution optimized to reproduce cycle results reported previously[21,22,28] (feature

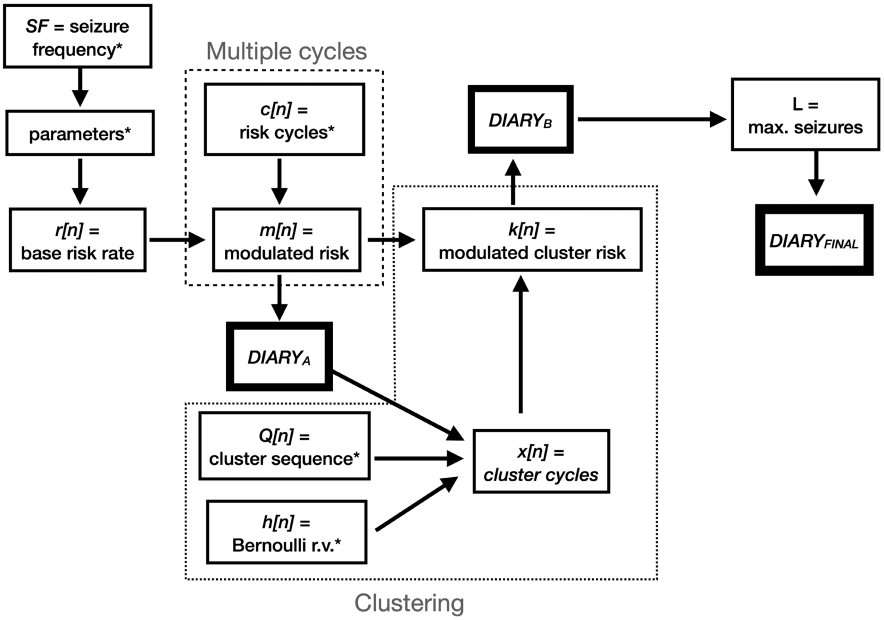

FIGURE 1 Simulation architecture. The conceptual workflow for the generation of a synthetic diary with Cyclic Heterogeneous Overdispersed Clustered Open-source L-relationship Adjustable Temporally limited E-diary Simulator (CHOCOLATE). First, a seizure frequency is generated. This number influences the parameters of the base risk rate. That rate is modulated by the multiple cycles. The modulated risk is used to produce $DIARY_A$, which in turn is used for the clustered diary, $DIARY_B$. Maximum seizure counts are limited before the $DIARY_{FINAL}$ is produced. Note, some of the stages have an asterisk (*) next to them. These indicate stages that include user modifiable parameters

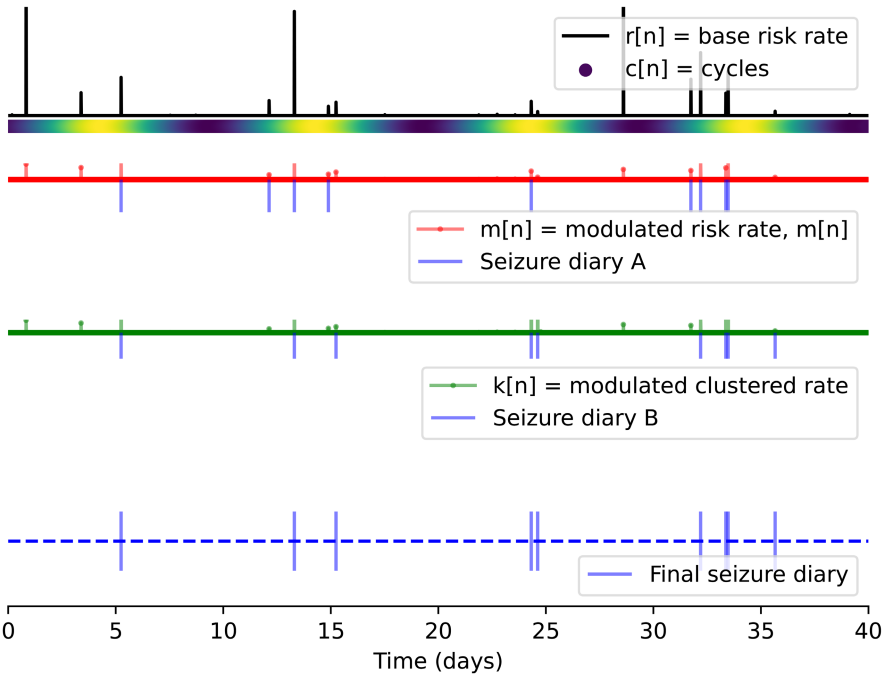

FIGURE 2 An example of a synthetic diary being generated. Top row: The base risk rate, *r[n]*. Under *r[n]*, a bar of cycling colors indicates a sinusoidal risk cycle signal, *c[n]*. Warmer (brighter) colors of *c[n]* represent higher risk susceptibility and cooler (darker) colors represent lower. To appreciate the modulating effect of *c[n]* when multiplied by *r[n]*, look at the second row, *m[n]*. Second row: The risk cycle signal modulates the base risk rate to produce the modulated risk rate *m[n]*. This signal is used to produce *Seizure Diary A* (vertical lines). Third row: *Diary A* and *m[n]* are used to produce the modulated cluster risk rate *k[n]*, which in turn is used to produce *Diary B* (vertical lines). Fourth row: The *Final Diary* (vertical lines) is produced by requiring that *Diary B* have no more than a maximum number per unit time. Note: Scales of each plot differ for clarity; therefore relative heights should be considered rather than absolute heights. In all the plots of diaries (Diary A, Diary B, and Final Diary), each vertical line represents one seizure

#3). The modulation parameters were optimized using a modified stochastic gradient descent algorithm to match the results from a multiple cycle analysis published from two data sets: 256 responsive neurostimulation patients[28] and from 1118 self-reported Seizure Tracker diaries.[22] A clustering algorithm that accounts for known characteristics[23,24,29–31] (feature #4) is then applied (see Appendix A). A maximum seizure limitation is imposed on diaries as well (feature #5). This limitation arises from the observation that 95% of generalized convulsions last less than 2 min,[19,25,26] >90% of 1 037 909 self-reported seizures last less than 5 min,[19] and the International League Against Epilepsy (ILAE) recognizes seizures longer than 5 min (or multiple seizures over 5 min without return to clinical baseline) to be status epilepticus.[25] It is challenging to quantify a typical duration for return to baseline.[32] Anecdotally, seizures can include a very short inter-seizure interval with apparent return to baseline while intracranial monitoring reveals ongoing electrographic status epilepticus. In the absence of additional studies, we propose a simple rule to avoid confusion with status epilepticus: we stipulate a maximum of one seizure in a 10-min interval. This simple rule ensures that the inter-seizure interval be at minimum twice the minimum status epilepticus duration. The code for this software is freely available at https://github.com/GoldenholzLab/CHOCOLATE.

2.2 | Simulating randomized controlled trials

To simulate RCTs, a standard parallel placebo-controlled design was used, modeled after typical RCTs used for medication and devices in epilepsy.[3,6,33,34] The RCT simulation used modeling assumptions mirroring prior studies.[2,3,6,8] Each of 10 000 trials comprised 200 synthetic CHOCOLATES patients, treated with "placebo" and "drug" in a 1:1 ratio. The "drug" was assumed to have a 30% efficacy[35] in all patients (a number typically seen across many clinical drug trials). The "placebo" was assumed to have an intrinsic 0% efficacy (i.e., no psychological "placebo effect" expected[36,37]). Effect of treatment was modeled by probabilistically removing seizures during the treatment period based on drug efficacy. In the absence of any seizure-rate variability, placebo-treated patients would see no change in seizure rate, and drug-treated patients would see a 30% reduction in seizure rate.[38] However, CHOCOLATES includes several features that produce variability in seizure rates; thus the outcomes are not easily predicted without simulation. Moreover, even in the absence of a modeled "placebo effect," synthetic patients in the placebo arm could appear to improve due to variability alone. There were 8 weeks of baseline (i.e., no treatment) and 12 weeks of treatment. Typical outcome metrics used by regulatory agencies were reported, including 50%-responder rate (RR50) and median percentage change (MPC).[39] Both metrics involve the percentage change from baseline seizure rate of each trial participant to be calculated. The results were compared to the RR50 and MPC from 23 historical anti-seizure medication trials.[40–62] To illustrate the flexibility of CHOCOLATES, additional trial simulations using different numbers of patients, drug strengths, and placebo strengths were evaluated as well (Appendix B).

2.3 | Evaluating a seizure-forecasting tool

To explore one potential use for CHOCOLATES, it was used to train a machine-learning algorithm. A training set of 10 000 seizure diaries ranging from 6 to 24 months in duration was produced. One diary sample was produced per day. A deep learning algorithm (fully connected eight-layer multilayer perceptron with dropout) was given the task of predicting risk of seizure in the coming 24-h given 3 months of history from that patient. Once trained, the algorithm was tested on two different datasets. First a set of 1000 test patients were generated in the same manner as the training set ("features on"). Second, a set of 1000 test patients were generated similarly but with the clustering and cycle features disabled ("features off"). A receiver-operating characteristic (ROC) curve and area under the curve (AUC) was produced based on the algorithm's ability to correctly predict the next day's seizure risk in each of the two test sets (this AUC represents all forecasts from all patients in one calculation).

3 | RESULTS

A demonstration of the stages of CHOCOLATES is shown with example outputs (Figure 2).

A plot of the distribution of seizure rates (Figure 3A) demonstrates the heterogeneity of seizure frequencies across a population of 50 000 daily sampled synthetic patients. Highlighted are the median seizure frequency of the population (2.9 seizures/month) and the median seizure frequency of patients who have >4 seizures/month (8.6 seizures/month). These two values are comparable to the 2.7 seizures/month in adults and 3.5 seizures/month in children reported in a study of 10 186 subjects,[19] and the average baseline seizure rate of study populations in the 23 RCTs, which was 8.4, whereas

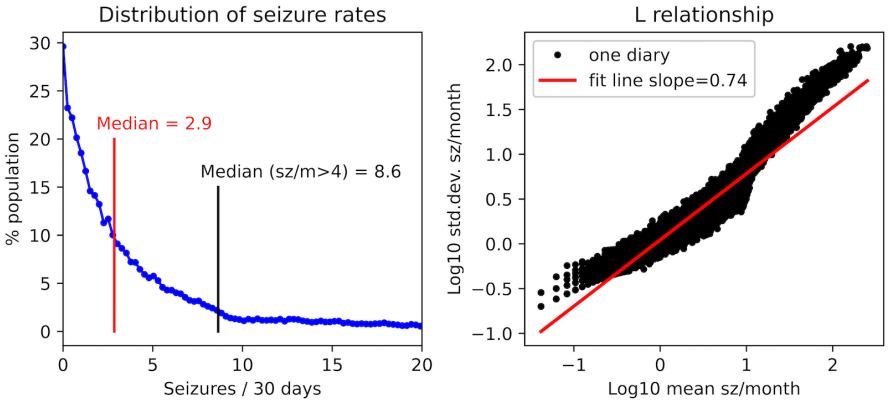

FIGURE 3 Seizure rate properties. (A) A population histogram of average monthly seizure rate from 50 000 synthetic patients. The median seizure rate from the population is highlighted. Also highlighted is the median seizure rate from a subpopulation if the seizure rate must be >4 seizures/month. A prior study[19] found the population median to be 2.7 for adults and 3.5 for children. (B) The L-relationship is shown. Each dot represents one of the 50 000 simulated patients. The position of the dot represents the logarithm of the mean and of the standard deviation of the monthly seizure rate from that patient. A fit line is plotted through the entire population of points, with a y intercept = 0 and the slope, which is shown. A prior study[20] found this slope to be 7.3–8.3

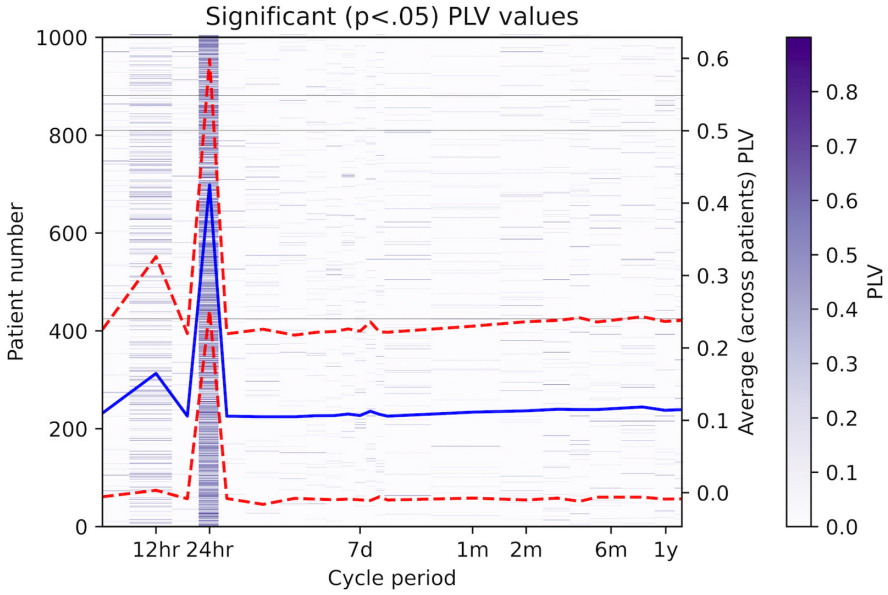

FIGURE 4 Multiple cycle properties. The result of simulating 1000 synthetic patients for 5 years is shown. Each patient is represented along the y axis. On the x axis, various period lengths are shown. The phase locking value (PLV) is calculated for each period length for each patient. Using a permutation method, statistically significant ($p < .05$) PLV values are shown as colored lines (if not statistically significant, only white is shown). Superimposed is an average PLV across all patients (blue curve) with 1 standard deviation above and below highlighted with dashed red curves. Prominently, significant 12- and 24-h cycles are noted in most patients, with longer duration, coexisting cycles seen in many patients at different period lengths

the average minimum eligibility seizure rate was 3.7 in those RCTs.[40–62]

The L-relationship represents a strong linear correlation between the logarithm of the mean seizure rate and the logarithm of the standard deviation of the seizure rate, measured as a slope of the fit line with a y intercept of zero. The slope was found to be 0.71–0.83 across multiple data sets,[20] which included intracranial data from the NeuroVista project,[63] physician-curated diaries from the Human Epilepsy Project, and user-curated self-reported diaries from Seizure Tracker.[19] Figure 3B shows the slope of the L-relationship to be 0.77 from CHOCOLATES.

A plot of the multiple co-existing cycles seen in the CHOCOLATES population is shown in Figure 4. This figure plots 1000 synthetic patients sampled hourly for

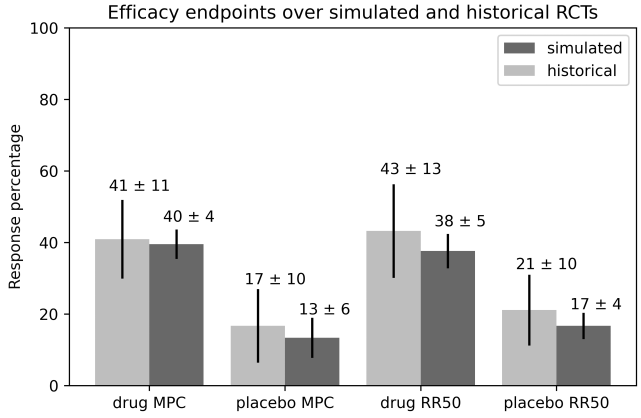

FIGURE 5 Randomized controlled trial simulations. After simulating 10 000 trials comprising 200 patients each (1:1 for 0% efficacy placebo to 30% efficacy drug), presented here are the simulated 50% responder rates (RR50) and the median percentage change (MPC) values. These are similar to values from 23 historical trials, also summarized here

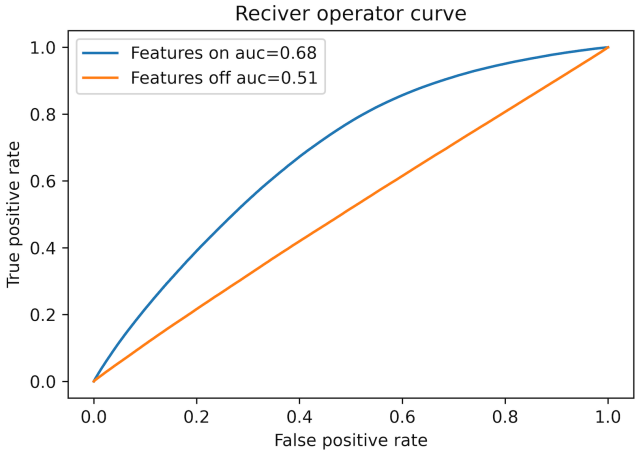

FIGURE 6 Testing a seizure forecasting tool. The receiver operating characteristic curves are plotted for two test sets: 1000 patients with clusters and cycles enabled ("with features") and 1000 patients clusters and cycles disabled ("without features"). These test sets were fed into a seizure-forecasting deep-learning algorithm trained from an independent 10 000 patients with clusters and cycles enabled. The plots demonstrate that when the features were enabled, the deep-learning algorithm was able to forecast the risk of seizure with higher accuracy, measured by area under the curve (AUC), representing all forecasts across patients

4 years. The fixed-cycle phase locking value (PLV) was computed for seizure cycle durations between 12 h and 1 year.[22,64] A conservative statistical permutation test was used to identify outlier ($p < .05$) PLVs with 1000 permutations.[64] All PLVs that were significant are drawn on the plot. Superimposed is the average ± standard deviation of PLVs across patients. The figure shows strong peaks at 12 h and 24 h for most patients, whereas significant peaks can be seen occasionally in a seemingly random fashion in some individual patients ranging up to 1-year periods.

The clustering capabilities of CHOCOLATES were tested by simulating 10 000 patients with and without the cluster tool enabled, sampled hourly for 20 years. For testing, clusters were defined using the traditional two or more seizures in 8 h definition.[24] When the CHOCOLATES clustering feature was turned on, 47% of patients had >20% of their seizures in the form of clusters, whereas when it was off, this was only 13%. This aligns well with a study that found 52% of non–seizure-free patients to exhibit clusters ($n = 137$).[29] When clusters were enabled, a median of 80% of seizures were isolated, and 20% were clustered (across patients). When clusters were disabled, 83% of seizures were isolated and 17% were clustered (across patients). These results align well with the observation that some clusters are the results of spurious timing of random events.[23,65]

When simulating RCTs (Figure 5), the model produced RR50: 17 ± 4% placebo, 38 ± 5% drug; and MPC: 13 ± 6% placebo, 40 ± 4 drug. This compared favorably to historical data: RR50 21 ± 10% placebo, 43 ± 13% drug; and MPC: 17 ± 10% placebo, and 41 ± 11% drug.[65] Additional simulations are presented in Appendix B, with alternative parameters to demonstrate the flexibility of CHOCOLATES to simulate many possible outcomes.

The seizure-forecasting tool (Figure 6) achieved an AUC = 0.68 with features on, and an AUC = 0.51 with features off, indicating that the presence of clusters and cycles facilitates seizure forecasting, whereas their absence makes forecasting very challenging. This forecasting example shows that CHOCOLATES can be used to interrogate potential mechanisms for forecasting by modifying features of simulated diaries.

4 | DISCUSSION

This study investigated a new realistic seizure diary simulator, CHOCOLATES, based on data from multiple studies of people with epilepsy. It accounts for many statistical features found in seizure diaries and was able to recapitulate results from typical anti-seizure medication RCTs. Nothing in the design or parameter specifications of CHOCOLATES mathematically required that it replicate RCT results; therefore this result may represent a form of model validation. In addition, it was able to shed light on the cause for the seemingly high seizure rates (roughly eight seizures/month) of patients in historical RCTs, showing that the heavy-tailed distribution (seen in Figure 3A) of seizure rates results in higher typical values when setting inclusion criteria at baseline

rates of ≥4 seizures/month. CHOCOLATES also demonstrated how seizure-forecasting tools (such as deep learning) can be explored more systematically with an unlimited supply of virtual patients that have easily adjustable statistical features. As shown here, investigators can simulate patients with or without cycles, with or without clusters, and control the details of each. A set of patients without some feature (i.e., a null distribution) can be produced when testing forecasting tools. The simulator accounts for (1) the heterogeneity of seizure rates,[19] (2) the L-relationship between rate and variability in rate,[20] (3) multiple co-existing cycles,[21,22] (4) seizure clusters,[23] and (5) maximum seizure counts.[19,25,26] Unlike many prior simulator tools (see tab. 1 in Tharayil et al.[18]), CHOCOLATES accounts for multiple statistical results reported from many different data sets derived from many different data collection modalities. Thus CHOCOLATES may be more generalizable and more realistic than previous models.

Because the tool is open-source and flexible, it can be used in a variety of situations. For RCTs, CHOCOLATES can be used for computing statistical power,[2–5] designing trial parameters,[3] comparing trial outcome metrics,[4,6,7] and early prediction of trial success.[8] For seizure-detection devices[9,10] and algorithms,[11] CHOCOLATES could help assess risk by comparing known "true" seizures to simulated detections accounting for sensitivity and false alarm rates. For neuromodulation devices,[12] CHOCOLATES could help optimize parameters. For seizure forecasting,[13–15] CHOCOLATES can produce unlimited supplies of realistic "random" data sets upon which new forecasting tools can be tested and compared. Data sets can be controlled for desired properties (such as types and relative quantities of cycles, duration of clusters, and so on). As shown in our example, the impact of manipulating specific features on forecasting accuracy can be tested. For preventing SUDEP,[16] CHOCOLATES could simulate a targeted epilepsy population for proposed devices, and the number needed to treat could be estimated. Clearly, many uses are possible.

The simulator comes with certain drawbacks. First and foremost, it is no better than the data that was used to design it. Therefore, the previously reported studies that included statistical results about seizure diaries may contain biases, calculation errors, or omissions. Those studies were taken at face value for the construction of CHOCOLATES. As examples of unknown quantities, skew and kurtosis of inter-seizure intervals have not been reported previously. Future studies will be needed to determine expected values. Second, any model system must make simplifying assumptions. Epilepsy is a collection of related conditions with more than100 underlying causes.[66] We believe that the key findings of most epilepsy conditions can be simulated with CHOCOLATES, but this remains to be further validated over time as data sets become larger and more comprehensive. Third, an important drawback of CHOCOLATES is that it comes with several parameters that have been tuned to obtain previously published results related to the five key statistical features. The parameters for the cluster module was not optimized precisely due to the wide range of expected values in the literature,[23,24,29–31] with prevalence of clustering between 13% and 76%. Our current default parameters matched the number of clustering patients with a study that found 52% of patients showing clustering over 1 year.[29] With newer studies of seizure clustering, it may be necessary to further tune the model, although the default cluster parameters are easily adjustable. Previously, it has been noted that self-reported clinical seizure diaries suffer from inaccuracies due to under-reporting[67] and over-reporting.[63] These inaccuracies are effectively built into CHOCOLATES, given that the model was built from data derived from these imperfect diaries. It is noted that data from the Human Epilepsy Project (physician verified self-reported data),[20] responsive neurostimulation,[21,28] and prolonged intracranial recordings[63] were also accounted for in CHOCOLATES, partially mitigating this effect.

Could a simpler model suffice? The answer will depend on the application. Sometimes, it may be adequate to ignore known statistical features of seizure diaries. In the case of a clinical trial simulation, sometimes a very simple model may suffice.[1] If the simulation accounts for the intra-individual variability seen in seizure rates, a more realistic model is preferable.[8] Similarly, when considering the impact of clustering or multiple cycles on RCT outcomes, CHOCOLATES would be very helpful. Conversely, none of this complexity is required if one is only interested in, for example, average seizure frequency. Because CHOCOLATES is an open-source model it can be reduced and simplified to meet the demands of individual studies. The full model was designed to have characteristics in common with measured human data.

CHOCOLATES represents the state-of-the-art in realistic seizure diary simulation, accounting for empirically derived features of seizure diaries across patient-reported, intracranial, and physician-curated data sets. We hope that this contribution will further the study of epilepsy and accelerate research toward more effective treatments.

AUTHOR CONTRIBUTIONS

DG designed and conceived the study, conducted the analysis, and drafted the manuscript. MBW designed and conceived the study, helped with interpretation of analysis, and edited the manuscript.

ACKNOWLEDGEMENTS

DMG was supported by National Institute of Neurological Disorders and Stroke (NINDS) KL2TR002542 and K23NS124656. MBW received funding support from the Glenn Foundation for Medical Research and the American Federation for Aging Research through a Breakthroughs in Gerontology Grant; the American Academy of Sleep Medicine through an AASM Foundation Strategic Research Award; and grants from the National Institutes of Health (NIH; R01NS102190, R01NS102574, R01NS107291, RF1AG064312, RF1NS120947, and R01AG073410) and National Science Foundation (NSF; 2014431). MBW is a co-founder of Beacon Biosignals and Director of Data Science for the McCance Center for Brain Health.

CONFLICT OF INTEREST

The authors report no competing interests.

ORCID

Daniel M. Goldenholz https://orcid.org/0000-0002-8370-2758
M. Brandon Westover https://orcid.org/0000-0003-4803-312X

REFERENCES

1. Karoly P, Goldenholz DM, Cook M. Are the days of counting seizures numbered? Curr Opin Neurol. 2018;31(2):162–8.
2. Goldenholz DM, Tharayil JJ, Kuzniecky R, Karoly P, Theodore WH, Cook MJ. Simulating clinical trials with and without intracranial EEG data. Epilepsia Open. 2017;2(2):156–61.
3. Goldenholz DM, Tharayil J, Moss R, Myers E, Theodore WH. Monte Carlo simulations of randomized clinical trials in epilepsy. Ann Clin Transl Neurol. 2017;4(8):544–52.
4. Goldenholz DM, Goldenholz SR, Moss R, French J, Lowenstein D, Kuzniecky R, et al. Does accounting for seizure frequency variability increase clinical trial power? Epilepsy Res [Internet]. 2017;137:145–51. https://doi.org/10.1016/j.eplepsyres.2017.07.013
5. Goldenholz DM, Sun H, Ganglberger W, Westover MB. Sample size analysis for machine learning clinical validation studies. medRxiv [Internet]. 2021;2021.10.26.21265541. http://medrxiv.org/content/early/2021/10/27/2021.10.26.21265541.abstract
6. Oliveira A, Romero JM, Goldenholz DM. Comparing the efficacy, exposure, and cost of clinical trial analysis methods. Epilepsia [Internet]. 2019;60(12):e128–32. https://doi.org/10.1111/epi.16384
7. Romero J, Chiang S, Goldenholz DM. Can machine learning improve randomized clinical trial analysis? Seizure [Internet]. 2021;91:499–502.
8. Romero J, Goldenholz DM. Statistical efficiency of patient data in randomized clinical trials of epilepsy treatments. Epilepsia [Internet]. 2020;61(8):1659–67.
9. Shum J, Friedman D. Commercially available seizure detection devices: a systematic review. J Neurol Sci [Internet]. 2021;428:117611.
10. Vander T, Stroganova T, Doufish D, Eliashiv D, Gilboa T, Medvedovsky M, et al. What is the optimal duration of home-video-EEG monitoring for patients with <1 seizure per day? A simulation study Front Neurol. 2022;13:938294.
11. Ahmad I, Wang X, Zhu M, Wang C, Pi Y, Khan JA, et al. EEG-based epileptic seizure detection via machine/deep learning approaches: a systematic review. Comput Intell Neurosci [Internet]. 2022;2022:6486570.
12. Ryvlin P, Rheims S, Hirsch LJ, Sokolov A, Jehi L. Neuromodulation in epilepsy: state-of-the-art approved therapies. Lancet Neurol. 2021;20(12):1038–47.
13. Goldenholz DM, Goldenholz SR, Romero J, Moss R, Sun H, Westover B. Development and validation of forecasting next reported seizure using e-diaries. Ann Neurol [Internet]. 2020;88(3):588–95.
14. Proix T, Truccolo W, Leguia MG, Tcheng TK, King-Stephens D, Rao VR, et al. Forecasting seizure risk in adults with focal epilepsy: a development and validation study. Lancet Neurol [Internet]. 2021;20(2):127–35.
15. Stirling RE, Cook MJ, Grayden DB, Karoly PJ. Seizure forecasting and cyclic control of seizures. Epilepsia [Internet]. 2021;62 Suppl 1(S1):S2–14.
16. Harden C, Tomson T, Gloss D, Buchhalter J, Cross JH, Donner E, et al. Practice guideline summary: sudden unexpected death in epilepsy incidence rates and risk factors: report of the guideline development, dissemination, and implementation Subcommittee of the American Academy of neurology and the American Epilepsy Society. Neurology. 2017;88(17):1674–80.
17. Balish M, Albert PS, Theodore WH. Seizure frequency in intractable partial epilepsy: a statistical analysis. Epilepsia. 1991;32(5):642–9.
18. Tharayil JJ, Chiang S, Moss R, Stern JM, Theodore WH, Goldenholz DM. A big data approach to the development of mixed-effects models for seizure count data. Epilepsia. 2017;58(5):835–44.
19. Ferastraoaru V, Goldenholz DM, Chiang S, Moss R, Theodore WH, Haut SR. Characteristics of large patient-reported outcomes: where can one million seizures get us? Epilepsia Open [Internet]. 2018;3(3):364–73.
20. Goldenholz DM, Goldenholz SR, Moss R, French J, Lowenstein D, Kuzniecky R, et al. Is seizure frequency variance a predictable quantity? Ann Clin Transl Neurol. 2018;5(2):201–7.
21. 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 [Internet]. 2018;9(1):1–10. https://doi.org/10.1038/s41467-017-02577-y
22. 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 [Internet]. 2018;17(11):977–85.
23. Chiang S, Haut SR, Ferastraoaru V, Rao VR, Baud MO, Theodore WH, et al. Individualizing the definition of seizure clusters based on temporal clustering analysis. Epilepsy Res. 2020;163:106330.
24. Haut SR. Seizure clusters: characteristics and treatment. Curr Opin Neurol [Internet]. 2015;28(2):143–50.
25. Trinka E, Cock H, Hesdorffer D, Rossetti AO, Scheffer IE, Shinnar S, et al. A definition and classification of status

epilepticus - report of the ILAE task force on classification of status epilepticus. Epilepsia. 2015;56(10):1515–23.
26. Theodore WH, Porter RJ, Albert P, Kelley K, Bromfield E, Devinsky O, et al. The secondarily generalized tonic-clonic seizure: a videotape analysis. Neurology. 2012;44(8):1403–3.
27. Klugman SA. Loss models: from data to decisions: from data to decisions, fifth edition book + solutions manual set (Wiley series in probability and statistics). 5th ed. Hoboken: Wiley; 2019.
28. Leguia MG, Andrzejak RG, Rummel C, Fan JM, Mirro EA, Tcheng TK, et al. Seizure cycles in focal epilepsy. JAMA Neurol [Internet]. 2021;78(4):454–63.
29. Detyniecki K, O'Bryan J, Choezom T, Rak G, Ma C, Zhang S, et al. Prevalence and predictors of seizure clusters: a prospective observational study of adult patients with epilepsy. Epilepsy Behav [Internet]. 2018;88:349–56.
30. Bauman K, Devinsky O. Seizure clusters: morbidity and mortality. Front Neurol. 2021;12:159.
31. Jafarpour S, Hirsch LJ, Gaínza-lein M, Kellinghaus C, Detyniecki K. Seizure cluster: definition, prevalence, consequences, and management. Seizure. 2019;68:9–15.
32. Fisher RS, Engel JJ. Definition of the postictal state: when does it start and end? Epilepsy Behav [Internet]. 2010;19(2):100–4.
33. Perucca E. Antiepileptic drugs: evolution of our knowledge and changes in drug trials. Epileptic Disord [Internet]. 2019;21(4):319–29.
34. Friedman D, French JA. Clinical trials for therapeutic assessment of antiepileptic drugs in the 21st century: obstacles and solutions. Lancet Neurol [Internet]. 2012;11(9):827–34. https://doi.org/10.1016/S1474-4422(12)70177-1
35. Rheims S, Perucca E, Cucherat M, Ryvlin P. Factors determining response to antiepileptic drugs in randomized controlled trials. A systematic review and meta-analysis. Epilepsia. 2011;52(2):219–33.
36. Goldenholz DM, Moss R, Scott J, Auh S, Theodore WH. Confusing placebo effect with natural history in epilepsy: a big data approach. Ann Neurol. 2015;78(3):329–36.
37. Goldenholz DM, Strashny A, Cook M, Moss R, Theodore WH. A multi-dataset time-reversal approach to clinical trial placebo response and the relationship to natural variability in epilepsy. Seizure. 2017;53:31–6.
38. Rheims S, Cucherat M, Arzimanoglou A, Ryvlin P. Greater response to placebo in children than in adults: a systematic review and meta-analysis in drug-resistant partial epilepsy. PLoS Med. 2008;5(8):1223–37.
39. Siddiqui O, Hershkowitz N. Primary efficacy endpoint in clinical trials of antiepileptic drugs: change or percentage change. Drug Inf J [Internet]. 2010;44(3):343–50. https://doi.org/10.1177/009286151004400316
40. Naritoku DK, Warnock CR, Messenheimer JA, Borgohain R, Evers S, Guekht AB, et al. Lamotrigine extended-release as adjunctive therapy for partial seizures. Neurology. 2007;69(16):1610–8.
41. Lim KS, Lotay N, White R, Kwan P. Efficacy and safety of retigabine/ezogabine as adjunctive therapy in adult Asian patients with drug-resistant partial-onset seizures: a randomized, placebo-controlled phase III study. Epilepsy Behav. 2016;61:224–30.
42. Arroyo S, Anhut H, Kugler AR, Lee CM, Knapp LE, Garofalo EA, et al. [3]Pregabalin add-on treatment: a randomized, double-blind, placebo-controlled, dose-response study in adults with partial seizures. Epilepsia. 2004;45(1):20–7.
43. Biton V, Di Memmo J, Shukla R, Lee YY, Poverennova I, Demchenko V, et al. Adjunctive lamotrigine XR for primary generalized tonic-clonic seizures in a randomized, placebo-controlled study. Epilepsy Behav. 2010;19(3):352–8.
44. Brodie MJ, Lerche H, Gil-Nagel A, Elger C, Hall S, Shin P, et al. Efficacy and safety of adjunctive zonisamide therapy for refractory partial seizures. Epilepsy Res. 2010;75(2–3):75–83.
45. Porter RJ, Partiot A, Sachdeo R, Nohria V, Alves WM. Randomized, multicenter, dose-ranging trial of retigabine for partial-onset seizures. Neurology. 2007;68(15):1197–204.
46. Chung S, Sperling MR, Biton V, Krauss G, Hebert D, Rudd GD, et al. Lacosamide as adjunctive therapy for partial-onset seizures: a randomized controlled trial. Epilepsia. 2010;51(6):958–67.
47. Ryvlin P, Werhahn KJ, Blaszczyk B, Johnson ME, Lu S. Adjunctive brivaracetam in adults with uncontrolled focal epilepsy: results from a double-blind, randomized, placebo-controlled trial. Epilepsia. 2014;55(1):47–56.
48. Halász P, Kälviäinen R, Mazurkiewicz-Beldzińska M, Rosenow F, Doty P, Hebert D, et al. Adjunctive lacosamide for partial-onset seizures: efficacy and safety results from a randomized controlled trial. Epilepsia. 2009;50(3):443–53.
49. Krauss GL, Serratosa JM, Villanueva V, Endziniene M, Hong Z, French J, et al. Randomized phase III study 306: adjunctive perampanel for refractory partial-onset seizures. Neurology. 2012;78(18):1408–15.
50. French JA, Krauss GL, Biton V, Squillacote D, Yang H, Laurenza A, et al. Adjunctive perampanel for refractory partial-onset seizures: randomized phase III study 304. Neurology. 2012;79(6):589–96.
51. French JA, Kugler AR, Robbins JL, Knapp LE, Garofalo EA. Dose-response trial of pregabalin adjunctive therapy in patients with partial seizures. Neurology. 2003;60(10):1631–7.
52. Lee BI, Yi S, Hong SB, Kim M-K, Lee SA, Lee SK, et al. Pregabalin add-on therapy using a flexible, optimized dose schedule in refractory partial epilepsies: a double-blind, randomized, placebo-controlled, multicenter trial. Epilepsia. 2009;50(3):464–74.
53. Ben-Menachem E, Biton V, Jatuzis D, Abou-Khalil B, Doty P, Rudd GD. Efficacy and safety of oral lacosamide as adjunctive therapy in adults with partial-onset seizures. Epilepsia. 2007;48(7):1308–17.
54. Berkovic SF, Knowlton RC, Leroy RF, Schiemann J, Falter U, Levetiracetam N01057 Study Group. Placebo-controlled study of levetiracetam in idiopathic generalized epilepsy. Neurology. 2007;69(18):1751–60.
55. Biton V, Berkovic SF, Abou-Khalil B, Sperling MR, Johnson ME, Lu S. Brivaracetam as adjunctive treatment for uncontrolled partial epilepsy in adults: a phase III randomized, double-blind, placebo-controlled trial. Epilepsia. 2014;55(1):57–66.
56. Klein P, Schiemann J, Sperling MR, Whitesides J, Liang W, Stalvey T, et al. A randomized, double-blind, placebo-controlled, multicenter, parallel-group study to evaluate the efficacy and safety of adjunctive brivaracetam in adult patients with uncontrolled partial-onset seizures. Epilepsia. 2015;56(12):1890–8.
57. French JA, Krauss GL, Steinhoff BJ, Squillacote D, Yang H, Kumar D, et al. Evaluation of adjunctive perampanel in patients with refractory partial-onset seizures: results of randomized global phase III study 305. Epilepsia. 2013;54(1):117–25.

58. Baulac M, Leon T, O'Brien TJ, Whalen E, Barrett J. A comparison of pregabalin, lamotrigine, and placebo as adjunctive therapy in patients with refractory partial-onset seizures. Epilepsy Res. 2010;91(1):10–9.
59. Beydoun A, Uthman BM, Kugler AR, Greiner MJ, Knapp LE, Garofalo EA. Safety and efficacy of two pregabalin regimens for add-on treatment of partial epilepsy. Neurology. 2005;64(3):475–80.
60. Elger CE, Brodie MJ, Anhut H, Lee CM, Barrett JA. Pregabalin add-on treatment in patients with partial seizures: a novel evaluation of flexible-dose and fixed-dose treatment in a double-blind, placebo-controlled study. Epilepsia. 2005;46(12):1926–36.
61. Biton V, Sackellares JC, Vuong A, Hammer AE, Barrett PS, Messenheimer JA. Double-blind, placebo-controlled study of lamotrigine in primary generalized tonic-clonic seizures. Neurology. 2005;65(11):1737–43.
62. French JA, Abou-Khalil BW, Leroy RF, Yacubian EM, Shin P, Hall S, et al. Randomized, double-blind, placebo-controlled trial of ezogabine (retigabine) in partial epilepsy. Neurology. 2011;76(18):1555–63.
63. 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.
64. Leguia MG, Rao VR, Kleen JK, Baud MO. Measuring synchrony in bio-medical timeseries. Chaos [Internet]. 2021;31(1):13138.
65. Romero J, Larimer P, Chang B, Goldenholz SR, Goldenholz DM. Natural variability in seizure frequency: implications for trials and placebo. Epilepsy Res [Internet]. 2020;162:106306.
66. Wyllie E, Gidal BE, Goodkin HP, Jehi L, Loddenkemper T. Wyllie's treatment of epilepsy: principles and practice.
67. Elger CE, Hoppe C. Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. Lancet Neurol [Internet]. 2018;17(3):279–88.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Goldenholz DM, Westover MB. Flexible realistic simulation of seizure occurrence recapitulating statistical properties of seizure diaries. Epilepsia. 2023;64:396–405. https://doi.org/10.1111/epi.17471