

RESEARCH ARTICLE

Prospective validation of a seizure diary forecasting falls short

Daniel M. Goldenholz^{1,2}  | Celena Eccleston^{1,2} | Robert Moss³ | M. Brandon Westover^{1,2,4,5}

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

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

³Seizure Tracker, Springfield, Virginia, USA

⁴Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA

⁵McCance Center for Brain Health, Boston, Massachusetts, USA

Correspondence

Daniel M. Goldenholz, Department of Neurology, Beth Israel Deaconess Medical Center, 330 Brookline Ave., Baker 5, Boston, MA 02215, USA.
Email: daniel.goldenholz@bidmc.harvard.edu

Funding information

National Science Foundation, Grant/Award Number: 2014431; National Institute of Neurological Disorders and Stroke, Grant/Award Number: K23NS124656, KL2TR002542, R01AG073410, R01NS102190, R01NS102574, R01NS107291, RF1AG064312 and RF1NS120947; American Academy of Sleep Medicine

Abstract

Objective: Recently, a deep learning artificial intelligence (AI) model forecasted seizure risk using retrospective seizure diaries with higher accuracy than random forecasts. The present study sought to prospectively evaluate the same algorithm.

Methods: We recruited a prospective cohort of 46 people with epilepsy; 25 completed sufficient data entry for analysis (median = 5 months). We used the same AI method as in our prior study. Group-level and individual-level Brier Skill Scores (BSSs) compared random forecasts and simple moving average forecasts to the AI.

Results: The AI had an area under the receiver operating characteristic curve of .82. At the group level, the AI outperformed random forecasting (BSS = .53). At the individual level, AI outperformed random in 28% of cases. At the group and individual level, the moving average outperformed the AI. If pre-enrollment (nonverified) diaries (with presumed underreporting) were included, the AI significantly outperformed both comparators. Surveys showed most did not mind poor-quality LOW-RISK or HIGH-RISK forecasts, yet 91% wanted access to these forecasts.

Significance: The previously developed AI forecasting tool did not outperform a very simple moving average forecasting in this prospective cohort, suggesting that the AI model should be replaced.

KEYWORDS

machine learning, prospective, seizure forecasting

1 | INTRODUCTION

Not knowing when the next seizure will happen reduces quality of life for people living with epilepsy. Roughly a decade ago, it was discovered that it is possible to provide seizure forecasts using invasive technology.¹ Since then, novel

approaches involving highly invasive²⁻⁵ and less invasive tools^{6,7} have been proposed. Using a retrospective study of 5419 unverified self-reported electronic diaries from Seizure Tracker, our group reported that 24-h forecasts from seizure diaries alone were possible using deep learning.⁸ The present study aimed to validate these findings prospectively.

2 | MATERIALS AND METHODS

2.1 | Patients

The protocol was deemed exempt by the Beth Israel Deaconess Medical Center Institutional Review Board. Participants were recruited by Seizure Tracker⁹ via email. Participants with (1) epilepsy, (2) age 18 years or older, (3) an active Seizure Tracker e-diary account, (4) at least three seizures recorded in their account, and (5) at least 3 months of previous e-diary data were eligible. Verified participants linked their e-diary and a RedCap^{10,11} survey account to the study. They completed an initial survey and then weekly surveys (verifying diary completion) for 5 months. They also maintained seizure e-diaries. For safety, only retrospective forecasts were provided monthly.

2.2 | AI forecaster

Using our pretrained deep learning algorithm⁸ (hereafter: AI), seizure forecasts were calculated for every day possible. The AI uses a recurrent neural network connected to a multilayer perceptron trained on 3806 users (Appendix A). All model parameters and hyperparameters remained unchanged from the original model.

The AI computes a probability of any seizures occurring within a 24-h period. The AI uses the 84-day trailing history of daily seizure counts leading up to that forecasted day as input. The tool was applied sequentially with a sliding window that moves forward 1 day at a time. Each patient could have up to 57 daily forecasts (8 weeks and 1 day), representing the prospective observation period. In some patients, this number was lower due to incomplete diary information (Appendix B). The 3-month pre-enrollment diaries were retained for additional analysis.

2.3 | Random forecaster

The daily AI forecast was compared with a permuted forecaster as a benchmark (hereafter “random”). The random forecaster is generated by permuting forecasts from the AI at the subject level. This can be thought of as shuffling a deck of cards, where each card is the AI forecast for a given day, and there is a different deck for each patient. A useful forecast should (at minimum) outperform a permuted forecaster.¹² Where appropriate, the average outcome metric from 1000 such permutations was used, such as for computing the Brier Score.

Key points

- A previously developed e-diary-based AI seizure forecasting tool was prospectively tested.
- Although by some metrics the tool was successful, the overall AI performance was unacceptably low.
- It was much easier to outperform a random forecast; it was much harder to outperform a moving average forecast.
- Using unverified diaries can skew forecasting metrics in favor of underperforming tools.

2.4 | Moving average forecaster

The daily AI forecast was also compared with a moving average forecaster, which accounted for the typical seizure rate from each patient. Moving average forecasts were computed by taking the total number of seizure days in each trailing 84-day history and dividing by 84 to obtain a simple estimate of daily risk of any seizures for the coming 24-h forecast (Appendix A). Of note, unlike a similar comparator used in our prior study (there called the “rate matched random” forecaster), this moving average forecaster uses total seizure days, not total seizure counts.⁸ This change was made to provide a more stringent comparator for the AI. Also of note, all summary results were computed using only the verified postenrollment period due to concerns about possible underreporting during the pre-enrollment period (see Discussion).

2.5 | Outcome metrics

Performance of each model was measured using area under the receiver operating characteristic curve (AUC), and the Brier Score. AUC values range between 0 and 1, with .5 representing a tool indistinguishable from coin flipping, and 1 representing a perfect discriminator. Brier Scores range between 0 and 1, with values closer to 0 representing higher accuracy. Our primary outcome (Appendix B) was comparing AI to the random forecasts using Brier Skill Scores (BSSs). BSS of 1 indicates the AI algorithm is perfect, 0 indicates the AI is not better than the reference forecast, and -1 indicates the reference forecast is perfect.

BSS was computed both at the group level and at the individual participant level. When using as reference test the random forecaster to calculate BSS, “group

level” means that random forecasts were generated by randomly shuffling the AI predictions across all patients and randomly reassigning them. Note that this means that forecasts from one patient may be randomly reassigned to other patients. By contrast, calculating BSS at the “individual level” relative to random forecasting means that random forecasts are all from the same patients, albeit in a randomly shuffled order. This means that the group- and individual-level BSS scores are not directly comparable, and the median of the individual-level BSS scores need not match the group-level BSS score. Additional BSS values were computed using the moving average as an alternative reference.

Calibration curves were generated for the AI, random, and moving average forecasters using equally spaced bins. Confidence intervals (CIs) for AUC and BSS values were obtained by 1000 bootstrapped samples, selecting patients with replacement.

Code is available here: <https://github.com/GoldenholzLab/deepManCode>.

3 | RESULTS

Of 46 recruited participants, one was ineligible, three were seizure-free, and 11 provided insufficient diary data. Within the remaining 31, there were three drop-outs and eight who missed some of the weekly diary completeness responses. Only 25 patients had sufficient contiguous data to perform forecasts based on 3 months of prospectively collected history. Forecastable diary days (Appendix C) ranged 15–57 (median = 57) days. Total seizures per patient ranged from 1 to 56 (median = 13). Participant characteristics are summarized in Table 1.

3.1 | Group-level results

The following represent group-level metrics (Figure 1). CIs were obtained via 1000 bootstrapped samples with replacement at the patient level. The AUC for AI was .82 (95% CI = .72–.90), and for the permuted AI (i.e., random forecast) it was .50 (95% CI = .46–.54). The Brier Score for AI was .14. The AI performed significantly better than the random forecaster at the group level, with a BSS (AI vs. random) of .53 (95% CI = .27–.70). However, the AUC of the moving average forecaster was also .82 (95% CI = .72–.89), which was not significantly different from the AI (Mann–Whitney U , $p = .13$), and the BSS of the AI relative to the moving average forecaster was $-.01$ (95% CI = $-.04$ to $.02$), suggesting minimal difference in performance.

TABLE 1 Baseline characteristics of participants in the prospective study.

Characteristic	n	%
Patients	31	
Females	14	45
Physician-confirmed epilepsy	31	100
EEG-confirmed epilepsy		
Yes	27	87
Unsure	4	13
Handedness		
Right	23	74
Left	6	19
Mixed	2	6
Epilepsy type		
Generalized	8	26
Focal	11	35
Focal + generalized	8	26
Unknown	4	13
Epilepsy location		
Frontal	1	3
Temporal	6	19
Parietal	0	0
Occipital	1	3
Multifocal	2	6
Unknown	21	68
Epilepsy cause		
Structural	9	29
Genetic	6	19
Infectious	1	3
Metabolic	0	0
Immune	0	0
Unknown	15	48
Prior epilepsy surgery	16	52

Note: Thirty-one patients had sufficient information to proceed to analysis; however, six did not have sufficient data for analysis involving forecasts made only from 3 months of prospectively collected history.

Abbreviation: EEG, electroencephalography.

3.2 | Individual-level results

In seven patients (28%) the AI was superior (i.e., individual BSS > 0) to the random forecaster, whereas for nine patients (36%) the AI was superior to the moving average. The individual BSSs (mean permuted AI forecasts¹² as comparator) were median .00 (95% CI = $-.03$ to $.20$). These values were notably lower than the group-level BSS values (see Appendix I). Individual BSSs with moving average as comparator were median $-.01$ with 95% confidence range $-.08$ to $.17$. Individual-level AI

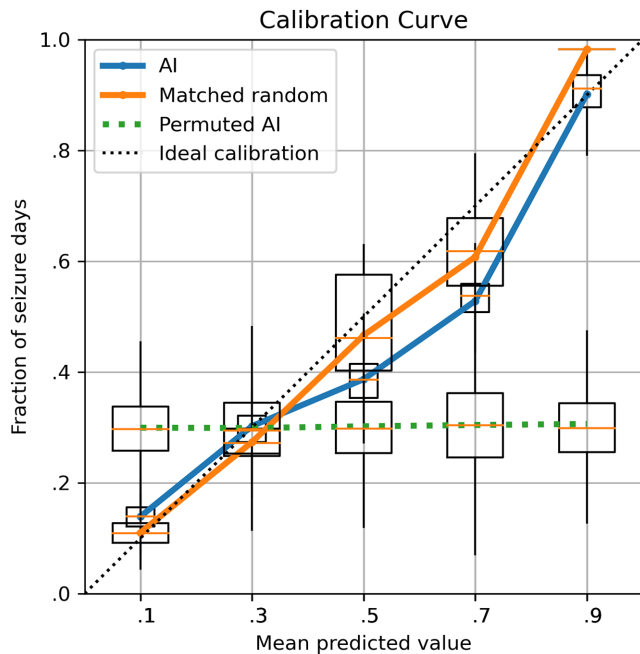


FIGURE 1 Calibration curves. The prospective seizure forecasts (pooled across all patients) are compared to the actual observed seizures for (1) the artificial intelligence (AI), (2) the rate-matched random forecast, and (3) random permutations of the AI. Confidence intervals are shown by bootstrapping 1000 times (choosing patients with replacement). A perfectly calibrated (dashed line) forecast would always forecast the correct percentage of observed seizures. In this figure, the AI and random forecast deviate from the ideal somewhat, whereas the permuted reference is very poorly calibrated (as expected).

AUC values were very poor quality ($AUC = .43 \pm .21$), as were individual-level moving average values ($AUC = .43 \pm .13$).

Complete diaries with AI and moving average forecasts were plotted (Appendices D and E). There were 25 patients reporting fewer than three seizures in the pre-enrollment period (see Appendix D). Time-in-warning analysis was conducted (Appendix G).

The above analyses were also recomputed using the full set of 31 patients using the 3-month pre-enrollment diaries (Appendix F). This showed the AI was superior to random and moving average at the group level, and superior to the moving average at the individual level in 14 patients (45%). However, pre-enrollment data seizure rate was dramatically lower than the enrollment seizure rates, suggesting severe underreporting.

The initial surveys ($n = 46$), filled out prior to any forecasting, included questions related to seizure forecasting (Appendix G). Many (52%) patients stated they would not mind poor-quality HIGH-RISK forecasts, and many (52%) did not mind poor-quality LOW-RISK forecasts, yet almost all (91%) wanted access to forecasts. In the setting of LOW-RISK forecasts, 80% said they would not change their

behavior, yet in HIGH RISK only 28% would not change; many stated that they would avoid risk-taking behavior (54%).

4 | DISCUSSION

Our results prospectively attempted validation of a deep learning seizure forecasting system⁸ that is based entirely on seizure diaries. At the group level (considering all forecasts from all patients equally), one may mistakenly believe that the AI has strong potential (our study found a group-level AI AUC of .82, similar to .86 in our prior study). Using a random permutation surrogate as our comparator, the AI performed statistically better than chance. However, a simple moving average forecaster turns out to perform just as well as the AI. Moreover, at the individual level (summarizing each patient separately first, then aggregating results), the AI outperforms the random permutation and the moving average in a small minority of cases, showing very poor overall individual-level performance in AUC and Brier Scores. The present work mirrors the previous retrospective study⁸; however, it focuses on the individual patient level with physician-curated, verified complete diaries. Here, we used moving average number of seizure days as a comparator, rather than a moving average number of total seizures as used previously.⁸ This resulted in a comparator that was more robust to brief seizure clusters and was harder to beat. By reporting multiple metrics in different ways using a more robust comparator, this study highlights deficiencies of the present AI algorithm and of certain outcome metrics. Clearly, the AI is not better than moving average forecasts; however, when missing data are present, the AI outperforms the moving average.

Qualitatively, the data (Appendices D–F) suggest that at least one driver of periods of better forecasts relates to the AI being better able to forecast multiday clusters of seizures compared with the random permutation or the moving average. These clusters may reflect multiday seizure susceptibility periods, although they do not appear to be periodic,^{3,13} and they do not fit the classical definition of seizure clusters.^{14,15}

Unlike our retrospective study⁸ that did not have verified complete diaries, the prospective study utilized weekly verified diaries from patients with clinical data confirming their epilepsy diagnosis. The misalignment of results between the former study and the present one may reflect the difference between the self-report and closely monitored self-report. In the case of the former, some events may be missed (underreporting¹⁶), but in the case of the latter, some dubious events may be included

(overreporting¹). There are no rigorous studies of overreporting, which is challenging to accurately quantify. Here, the verified diaries have dramatically higher rates during the prospective phase compared to the pre-enrollment 3-month periods (see Appendix D), strongly suggesting underreporting.

The apparent underreporting from the pre-enrollment period appears to reflect that without supervision, diaries might be incomplete. Our study required for enrollment the existence of a Seizure Tracker account with at least 3 months of data prior to enrollment; however, we did not verify or demand that such diaries were complete. This oversight is significant, because during the observed portion of the study we asked the participants weekly whether their diaries were complete, and the seizure rates were consistently much higher (see Appendix D). Importantly, multiple lines of evidence^{13,17–21} show that, contrary to what we observed in our cohort, unverified seizure diaries often do reproduce patterns confirmed in verified systems; thus, unsupervised seizure diaries may not always suffer from underreporting bias. Nevertheless, future studies will need to either confirm with participants that pre-enrollment diaries are complete or obtain longer duration observation periods and use only data obtained during confirmed time frames.

One might suspect that patients with very high seizure rates would be unlikely to benefit from seizure forecasts at all. On the other hand, our cohort included only patients who wanted to be involved in a forecasting study (there was no compensation for this study), and 39% of them had very high seizure rates. Patient preferences (Appendix G) may even support inaccurate forecasts rather than no forecasts. It is worthwhile to note that the preferences reported were obtained prior to obtaining any forecasts from our team; therefore, these can be viewed as the opinion of optimistic patients who had just enrolled in a study. Nevertheless, forecasts are likely the most important in patients with less frequent seizure days (based on the need to make temporary changes in behavior), and the present algorithm did not excel in this area. More study is needed to better understand what the characteristics are of patients who would be most interested in seizure forecasts, and who would benefit most. It should be emphasized that in the absence of a nearly perfect forecast system, patients should never be encouraged to engage in risky behavior during periods of forecasted low risk.

The present study has several limitations. First, some people with epilepsy have very low (e.g., 1–2 seizures per year) or very high (i.e., \geq daily) seizure rates.²² Such patients would not be likely to benefit from the current generation of daily forecasting tools. Second, it can be challenging for patients to maintain a seizure diary,²³ thus

limiting tools of this nature to patients and caregivers willing to maintain a diary. Third, our prior⁸ and present study did not have available electroencephalographic (EEG) data to augment forecasts. Although speculative, including EEG data may enhance the performance of these models. Fourth, the 5-month prospective duration of the present study may be too short to make definitive conclusions about the utility of the AI algorithm. To address this deficiency, our group will be conducting a larger study soon with a longer observation period to allow for sufficiently large windows of investigator-verified seizure diaries. Fifth, there was a presumed dramatic underreporting in the pre-enrollment period. In our future study, we will not include a pre-enrollment period due to the challenges in verifying that they are complete. Finally, the choice of reference standard comes at a cost. Our average permutation (a.k.a. random) forecaster standard could not be realistically provided to patients in real time. Conversely, our second reference standard was the moving average forecaster. This can be implemented in a real-time system, making it a realistic comparator. A comparison of the calibration curve (Figure 1) shows very poor calibration of the permuted AI, but decent calibration of moving average and AI. In using both, we highlight the advantages and disadvantages of each.

Based on some of the lessons learned here, we suggest future prospective studies should include several things. First, the typical seizure rate of each participant should be reported. Second, candidate forecasting tools should have high individual-level performance metrics in isolation and compared to a robust comparator. Third, we recommend the moving average number of seizure days as the preferred comparator for calculating BSS. Finally, we discourage the use of pooled group-level metrics (i.e., considering all forecasts across patients equally), due to the possibility of obscuring underperforming models. Rather, patient-level metrics should be computed, and population statistics should be reported based on those. Moreover, the reporting of individual Brier Score and BSS values are best reported in combination with individual seizure frequency, given the correlation between the two.

We hope that future advances in wearables⁶ and minimally invasive tools^{7,24} can synergistically be applied to diary-based forecasting tools to achieve higher accuracy and wider patient appeal.

AUTHOR CONTRIBUTIONS

Daniel M. Goldenholz, M. Brandon Westover, and Robert Moss contributed to conception and design of the study; Celena Eccleston, Daniel M. Goldenholz, and Robert Moss contributed to acquisition and analysis of data. Daniel M. Goldenholz drafted a significant portion of the manuscript and figures.

ACKNOWLEDGMENTS

M.B.W. is a cofounder of Beacon Biosignals, and Director for Data Science for the McCance Center for Brain Health.

FUNDING INFORMATION

D.M.G. was supported by NINDS KL2TR002542 and K23NS124656. M.B.W. received funding support from the American Academy of Sleep Medicine through an AASM Foundation Strategic Research Award, the NIH (R01NS102190, R01NS102574, R01NS107291, RF1AG064312, RF1NS120947, R01AG073410), and the NSF (2014431).

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest for any of the authors. 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

Data are available on reasonable request.

ORCID

Daniel M. Goldenholz  <https://orcid.org/0000-0002-8370-2758>

REFERENCES

- 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.
- Leguia MG, Andrzejak RG, Rummel C, Fan JM, Mirro EA, Tchong TK, et al. Seizure cycles in focal epilepsy. *JAMA Neurol* [Internet]. 2021;78(4):454–63. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33555292>
- 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>
- Proix T, Truccolo W, Leguia MG, Tchong 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. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/33341149>
- Nasseri M, Pal Attia T, Joseph B, Gregg NM, Nurse ES, Viana PF, et al. Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning. *Sci Rep* [Internet]. 2021;11(1):21935. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/34754043>
- Karoly PJ, Stirling RE, Freestone DR, Nurse ES, Maturana MI, Halliday AJ, et al. Multiday cycles of heart rate are associated with seizure likelihood: an observational cohort study. *EBioMedicine* [Internet]. 2021 [cited 2021];72:103619. Available from: <https://pubmed.ncbi.nlm.nih.gov/34649079/>
- 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 inpatient models. *Epilepsia*. 2023;64(Suppl 4):S124–S133. doi: [10.1111/epi.17252](https://doi.org/10.1111/epi.17252).
- 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. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32567720>
- Casassa C, Rathbun Levit E, Goldenholz DM. Opinion and special articles: self-management in epilepsy: web-based seizure tracking applications. *Neurology*. 2018;91(21):e2027–e2030.
- Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81.
- 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.
- 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. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30219655>
- Haut SR. Seizure clusters: characteristics and treatment. *Curr Opin Neurol* [Internet]. 2015;28(2):143–50. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25695133>
- 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.
- Elger CE, Hoppe C. Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. *Lancet Neurol* [Internet]. 2018;17(3):279–88. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29452687>
- Goldenholz DM, Westover MB. Flexible realistic simulation of seizure occurrence recapitulating statistical properties of seizure diaries. *Epilepsia* [Internet]. 2023;64(2):396–405. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/36401798>
- 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.
- 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(June):145–51. <https://doi.org/10.1016/j.eplepsyres.2017.07.013>
- LaGrant B, Goldenholz DM, Braun M, Moss RE, Grinspan ZM. Patterns of recording epileptic spasms in an electronic seizure diary compared with video-EEG and historical cohorts. *Pediatr Neurol* [Internet]. 2021;122:27–34. <https://doi.org/10.1016/j.pediatrneurol.2021.04.008>

21. 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.
22. 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. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/30187007>
23. Fisher RS, Blum DE, DiVentura B, Vannest J, Hixson JD, Moss R, et al. Seizure diaries for clinical research and practice: limitations and future prospects. *Epilepsy Behav*. 2012;24(3):304–10.
24. Stirling RE, Grayden DB, D'Souza W, Cook MJ, Nurse E, Freestone DR, et al. Forecasting seizure likelihood with wearable technology. *Front Neurol*. 2021;12(July):1–12.

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, Eccleston C, Moss R, Westover MB. Prospective validation of a seizure diary forecasting falls short. *Epilepsia*. 2024;65:1730–1736. <https://doi.org/10.1111/epi.17984>