HHS Public Access

Author manuscript

*Epilepsia*. Author manuscript; available in PMC 2026 January 30.

Published in final edited form as:

*Epilepsia*. 2026 April ; 67(4): 1587–1588. doi:10.1002/epi.70083.

# Deployable seizure forecasting requires clinically meaningful performance: Response to Stirling et al

**Chi-Yuan Chang, PhD**[1,2], **Robert Moss, BS**[3], **M. Brandon Westover, MD PhD**[1,2], **Daniel M. Goldenholz, MD, PhD**[1,2]

[1]Harvard Medical School, Boston MA

[2]Beth Israel Deaconess Medical Center, Boston, MA

[3]Seizure Tracker LLC, Springfield, VA

In contrast with our recent study[1], Stirling et al. state that Cycle forecasts outperform the Napkin method in EEG cohorts (n=24; hourly AUC ≈0.70 vs 0.47, daily AUC ≈0.58 vs 0.47) and in a diary cohort (n=808; hourly AUC ≈0.66 vs 0.50, daily AUC ≈0.58 vs 0.50).

We have several comments. First, we provide code from our analysis: https://github.com/GoldenholzLab/Deepman2.git. Second, the forecasting code we used in our study for computing Cycle was provided by the Stirling / Karoly labs – our lab generated these forecasts just as they do. Therefore, the main difference between their result and ours is likely the cohort of patients tested. Third, we recommend Stirling's group exclude the 101 diary patients (12.5%) and 8 EEG patients (33.3%) with seizure frequency >0.5/day as such patients are inappropriate for daily forecasts.

Our recent paper[1] focused on only diary-based seizure forecasting. Nevertheless, the recommendations in our rigorous benchmarks paper[2] are relevant to any seizure forecasting method and dataset (including EEG). Those are:

1. Report outcomes metrics as a function of SF.
2. Report both discrimination and calibration forecasting metrics.
3. A model must outperform the Napkin method across all of #1 and #2

**Corresponding author**: Daniel M. Goldenholz, 330 Brookline Ave, Baker 5, Boston MA 02215, daniel.goldenholz@bidmc.harvard.edu.

Conflicts of Interest Disclosure:

Dr. Chang has no conflict. Mr. Moss is the cofounder and owner of Seizure Tracker, LLC, and has received personal fees from Courtagen Life Sciences, Engage Therapeutics, Epitel, LivaNova, Marinus Pharmaceuticals, Neurelis, Neuropace, UCB, and grants from the Tuberous Sclerosis Complex Alliance. Seizure Tracker was paid for the effort to participate in this project via NIH funding. Dr. Goldenholz is an unpaid advisor for Epilepsy AI and Eysz. He has been provided speaker fees from AAN, AES, ACNS, and NNS. He also previously has been a paid consultant for Neuro Event Labs, IDR, LivaNova and Health Advances. Dr. Westover is a co-founder, scientific advisor, and consultant to Beacon Biosignals and has a personal equity interest in the company. He also receives royalties for authoring Pocket Neurology from Wolters Kluwer and Atlas of Intensive Care Quantitative EEG by Demos Medical.

Ethics Approval Statement

No ethical approval was needed for the present commentary.

Ethical Publication Statement

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.

4. If #3 fails, the model cannot be clinically useful.
5. If #3 succeeds, clinical utility is possible, but not assured.

The *magnitude of effect* on discrimination is only marginally better for Cycle than the Napkin and is nearly indistinguishable on calibration. Cycle is likely clinically indistinguishable from Napkin, despite statistical significance. Some might misread "statistically significant" to indicate "accurate." Indeed, studies from our lab[3] and theirs[4] suggest patients are eager to have forecasts, even inaccurate forecasts. Moreover, they report that patients desire forecasts for scheduling "travel activities"[4]. Perhaps the common misunderstanding of statistics[5], coupled with low accuracy forecasting could translate into increases in motor vehicle accidents and other injuries for people with epilepsy[6].

We view the findings Stirling et al. present as not a refutation, but a validation of our main conclusion – the Cycle method in both our analysis and theirs does not demonstrate a clinically significant superiority to the Napkin method. We continue to hope for better models.

## Funding:

DG and CC were supported by NINDS K23NS124656. MB was supported by grants from the NIH (RF1AG064312, RF1NS120947, R01AG073410, R01HL161253, R01NS126282, R01AG073598, R01NS131347, R01NS130119), and NSF (2014431).

## Data Availability

No new data is presented in this commentary.

## References

1. Chang C, 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 https://doi.org/10.1111/EPI.18677 (2025) doi:10.1111/EPI.18677.
2. Chang C-Y et al. Necessary for seizure forecasting outcome metrics: Seizure frequency and benchmark model. Epilepsy Res 208, 107474 (2024). [PubMed: 39522392]
3. Goldenholz DM, Eccleston C, Moss R & Westover MB Prospective validation of a seizure diary forecasting falls short. Epilepsia https://doi.org/10.1111/EPI.17984 (2024) doi:10.1111/EPI.17984.
4. Stirling RE et al. User experience of a seizure risk forecasting app: A mixed methods investigation. Epilepsy and Behavior 157, (2024).
5. Westover MB, Westover KD & Bianchi MT Significance testing as perverse probabilistic reasoning. BMC Med 9, (2011).
6. Joshi CN, Vossler DG, Spanaki M, Draszowki JF & Towne AR "Chance Takers Are Accident Makers": Are Patients With Epilepsy Really Taking a Chance When They Drive? Epilepsy Curr 19, 221 (2019). [PubMed: 31328536]