Epilepsy & Behavior 99 (2019) 106408

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Letter to the Editor

Commentary on "Predicting seizure freedom after epilepsy surgery, a challenge in clinical practice"

To the Editor,

The recent article entitled "Predicting seizure freedom after epilepsy surgery, a challenge in clinical practice" [1] has raised some important questions, but we have some concerns.

The study aimed to assess how accurately clinicians predict surgical outcomes and to compare their performance to two surgical outcome prediction algorithms called modified seizure freedom score (m-SFS) and epilepsy surgery nomogram (ESN) [2,3]. For the former aim, they recapitulated findings seen in many other studies; even within the presence of rich presurgical clinical data, clinicians are poor at forecasting surgical success. For the latter aim, performance between the three techniques was compared. There was no statistically significant difference between ESN and clinicians, nor between m-SFS and ESN. Nevertheless, the authors conclude that "there is a moderate statistical difference favoring the m-SFS to the clinicians (p 0.0960 and 0.0514, for 2 and 5 years)". The statement appears in the abstract, results, and figure legend of the article. That statement raises two problems.

The first problem is one of logic. If A is no different from B, and B is no different from C, common sense leads us to expect that A is no different than C. That logical syllogism fails in the above statement by the authors. This logic can fail when using statistical significance testing to make inferences [4], e.g., when there are insufficient data to accurately estimate an effect and/or when the effect size is very small. The authors discuss neither of these possibilities.

The second problem is one of statistical convention. By convention, an effect is considered statistically significant when $p < 0.05$. By this convention, m-SFS cannot be said to differ statistically from clinicians. Even if the authors' data had shown a $p < 0.05$, reliance on p-values alone in drawing scientific conclusions is itself problematic. Proposed solutions range from making the conventional threshold stricture (e.g., $p < 0.005$) [5], to discarding p-values entirely [6]. In any case, concluding that a simple algorithm outperforms clinical judgment in surgical decision-making requires considerably stronger evidence than the authors provide.

In addition to the above, the receiver operating characteristic (ROC) curve in Fig. 6b is almost entirely below the $y = x$ diagonal line labeled "Doctors" in orange. This line is labeled in the legend with an area under the curve (AUC) = 0.539. We wonder if the authors mislabeled this number because in order for it to be correct, the orange curve would need to be mostly above the $y = x$ line. Otherwise, the AUC is representing a "paradoxical" situation, in which case doctor's predictions are the opposite of the truth. Another possibility is that the "Doctors" curve may have been plotted inaccurately, and that the AUC value is correct. We checked the plot using WebPlotDigitizer [7] and found that perhaps the authors were reporting 1-AUC, because the AUC represented by the orange curve is closer to 0.462. In any case, all three prediction methods perform very close to chance level (AUC 0.5). Whether they differ at all from chance level is not clear, as confidence bands (or other appropriate measures of uncertainty) are not provided.

Overall, the authors' conclusions in the article were appropriately muted, not suggesting that m-SFS nor ESN be clinically employed. We agree here, because the data presented in the article supports the concept that clinicians, ESN, and m-SFS are all unreliable oracles of seizure-freedom among patients who undergo epilepsy surgery.

The data available in the article (and elsewhere) do not support the suggestion by the authors that the best strategy is to "use the available technology, the clinician's knowledge, the presentation of cases in comprehensive multidisciplinary group conferences along with the available statistical analyses to open a discussion with the patient as a possible candidate of epilepsy surgery." Indeed, that suggestion represents expert opinion only, and has little support in the literature, as alluded to by the authors, by reference, and by the data presented. There currently is no reliable, validated method to produce predictions of long-term seizure-free outcomes, and the best that one can do is to present the outcome statistics that 40–50% of patients who undergo epilepsy surgery attain long-term seizure remission [8]. Interestingly in 1947, the long-term surgical success rate was reported by Penfield as 53% [9]. It is sobering to admit that all the available data, tools, and expert opinions accrued since then have not increased this number.

Funding

DG and HS have no funding to disclose.

BW has support from NIH K23NS090900, R01NS102190 and R01NS107291.

References

[1] Gracia CG, Chagin K, Kattan MW, Ji X, Kattan MG, Crotty L, et al. Predicting seizure freedom after epilepsy surgery, a challenge in clinical practice. Epilepsy Behav 2019;95:124–30. https://doi.org/10.1016/j.yebeh.2019.03.047.

[2] Jehi L, Yardi R, Chagin K, Tassi L, Lo Russo G, Worrell G, et al. Development and validation of nomograms to provide individualised predictions of seizure outcomes after epilepsy surgery: a retrospective analysis. Lancet Neurol 2015;14:283–90. https://doi.org/10.1016/S1474-4422(14)70325-4.

[3] Garcia Gracia C, Yardi R, Kattan MW, Nair D, Gupta A, Najm I, et al. Seizure freedom score: a new simple method to predict success of epilepsy surgery. Epilepsia 2015; 56:359–65. https://doi.org/10.1111/epi.12892.

[4] Westover MB, Westover KD, Bianchi MT. Significance testing as perverse probabilistic reasoning. BMC Med 2011;9:20. https://doi.org/10.1186/1741-7015-9-20.

[5] Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, et al. Redefine statistical significance. Nat Hum Behav 2018;2:6–10. https://doi.org/10.1038/s41562-017-0189-z.

[6] Ziliak ST, McCloskey DN. The cult of statistical significance: how the standard error costs us jobs, justice, and lives (economics, cognition, and society). 1rst ed. University of Michigan Press; 2008.

[7] Rohatgi A. WebPlotDigitizer. https://apps.automeris.io/wpd/, Accessed date: 5 July 2019.

https://doi.org/10.1016/j.yebeh.2019.07.009

[8] Malmgren K, Edelvik A. Long-term outcomes of surgical treatment for epilepsy in adults with regard to seizures, antiepileptic drug treatment and employment. Seizure 2017;44:217–24. https://doi.org/10.1016/j.seizure.2016.10.015.
[9] Penfield W, Steelman H. The treatment of focal epilepsy by cortical excision. Ann Surg 1947;126:740–62.

Daniel Goldenholz
Harvard Beth Israel Deaconess Medical Center, Department of Neurology, United States of America
Corresponding author at: 330 Brookline Ave, Baker 5, Boston, MA 02215, United States of America.
E-mail address: daniel.goldenholz@bidmc.harvard.edu.

Haoqi Sun
Brandon Westover
Harvard Massachusetts General Hospital, Department of Neurology, United States of America
E-mail address: hsun8@mgh.harvard.edu.
E-mail address: mwestover@mgh.harvard.edu.

25 June 2019