

Journal Club: Criteria for Defining Interictal Epileptiform Discharges in EEG

John R. McLaren, MD, Jin Jing, PhD, M. Brandon Westover, MD, PhD, and Fábio A. Nascimento, MD

Neurology® 2022;99:430-432. doi:10.1212/WNL.0000000000200991

Correspondence

Dr. McLaren
jmclaren2@
mgh.harvard.edu

EEG misinterpretation is a major contributor to epilepsy misdiagnosis.^{1,2} To help improve interictal epileptiform discharge (IED) identification, the International Federation of Clinical Neurophysiology (IFCN) proposed a set of 6 operational criteria³ (Table 1) that was subsequently validated with Class III evidence demonstrating IFCN criteria in sensor space (visual analysis) and analysis in source space (spatial filtering method applied to transform scalp signals into regions of interest) have high specificity (>95%) and sensitivity (81%–85%) for IED identification, similar to expert scorings.⁴

Hypothesis and Design

The authors investigated the diagnostic accuracy and interrater agreement (IRA) for IED identification using the IFCN criteria in sensor space and source space against a benchmark specificity of 95% for clinical significance. The hypothesis was that these methods would provide high diagnostic accuracy and reduce EEG overinterpretation compared with nonstandardized visual analysis. The study was a large, blinded, retrospective review of EEGs from 100 patients with and without epilepsy.

Methods

The authors selected EEGs of consecutive patients undergoing video-EEG monitoring at 2 Danish centers between 2012 and 2017. Patients were included if they were aged 1 year or older and had habitual paroxysmal events (epileptic or nonepileptic) with interictal sharp transients on EEG. For those with epileptic seizures, the sharp transients had to be concordant with the recorded ictal event. Patients were excluded if their recordings were diagnostically inconclusive or if they had both epileptic and nonepileptic events in the same recording. Interictal sharp transients, previously marked on initial diagnostic interpretation, were reviewed in 10- to 20-second epochs and selected for further evaluation by 2 authors who did not participate in the further in-study EEG analysis. These epochs were included only if both experts confirmed the sharp transients met selection criteria (transient, pointed peak, phase reversal in bipolar montage, and concordant with the recorded seizure, if epileptic). Seven separate experts then evaluated the samples in 3 different rounds more than 1 month apart, blinded to all other patient data.

In the *first round*, raters scored each marked transient as having the presence or absence of each IFCN criterion in sensor space. In the *second round*, raters analyzed EEG samples in source space in a different randomized order. IEDs were identified if the sharp transients met all 3 of the following: changes in time, distribution in space, and exclusion of artifacts and normal variants. In the *third round*, raters labeled the sharp transients as either epileptiform or nonepileptiform without any specific criteria, as if in their clinical practice. This was considered expert scoring.

Diagnostic accuracy of each extracted feature was achieved using the consensus majority scoring. For the IFCN sensor space criteria, a cutoff value of 2–6 fulfilled criteria for qualifying as an IED. Analysis in source space and expert scoring resulted in dichotomous scores, labeled

From the Department of Neurology (J.R.M., J.J., M.B.W., F.A.N.), Massachusetts General Hospital, Harvard Medical School, Boston, and Department of Neurology (F.A.B.), Washington University School of Medicine, St. Louis, MO.

Go to [Neurology.org/N](https://www.neurology.org/N) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Table 1 IFCN Proposed Operational Definition of IEDs to Help Standardize EEG Visual Analysis

1. Diphasic or triphasic waves with sharp or spiky morphology (i.e., pointed peak)
2. Different wave duration than the ongoing background activity either shorter or longer
3. Asymmetry of the waveform: a sharply rising ascending phase and a more slowly decaying descending phase, or vice versa
4. The transient is followed by an associated slow afterwave
5. The background activity surrounding IEDs is disrupted by the presence of the IEDs
6. The distribution of the negative and positive potentials on the scalp suggests a source of the signal in the brain, corresponding to a radial, oblique, or tangential orientation of the source.

Abbreviations: IED = interictal epileptiform discharge; IFCN = International Federation of Clinical Neurophysiology.

as either IEDs or nonepileptiform sharp transients. The gold standard used to determine whether patients experience epileptic or nonepileptic events originated from video-EEG data recording these patients' habitual spells.

Statistical Analysis

The receiver operating characteristic (ROC) curve⁵ was calculated based on cutoff values of 2–6 IFCN criteria in sensor space. ROC curves evaluate the performance of diagnostic tests and, more generally, the accuracy of a statistical model that classifies outcomes in a dichotomous fashion.⁶ The Wilson⁷ method was used to calculate 95% confidence intervals for sensitivity, specificity, and accuracy. These values were then compared between the different IED identification methods using the McNemar test.⁸ The IRA was calculated using the Gwet⁹ agreement coefficient (AC1) and interpreted as poor (<0.02), fair (0.2–0.4), moderate (0.4–0.6), substantial (0.6–0.8), or almost perfect (>0.8). The IRAs of the IED identification methods were then compared using a bootstrap analysis.

Results

The AUC for IFCN criteria in sensor space was 0.977 for the consensus majority scoring and median 0.941 for individual raters. The ROC curve highlights that IFCN criteria in sensor space with cutoff values of 4 and 5 and analysis in source space performed similar to expert scoring with no statistically significant difference in accuracy ($p > 0.157$). IFCN criteria with a cutoff of 5 resulted in higher specificity ($p = 0.025$) but lower sensitivity ($p = 0.005$) compared with a cutoff of 4. The 2 methods that achieved specificity higher than expected 95% threshold were IFCN sensor space with a cutoff of 5 and analysis in source space. Similarly, using the expert consensus as gold standard rather than patients' video-EEG data, these 2 methods achieved the highest accuracies (93%). The 3 methods with the best diagnostic performance (IFCN sensor with cutoffs of 4 and 5 and source space) had an IRA that was moderate to substantial (AC1: 0.490–0.608) without a significant difference in IRA between them ($p > 0.900$). IRA for individual criterion was found to be substantial for 1, 2, and 4, moderate for 6, and fair for 3 and 5 (Table 1 (links.lww.com/WNL/C164)).

Interpretation

The design where EEGs were preselected before they were serially rated by experts in an independent fashion provides a solid foundation to address the authors' research question. The use of ROC graphs is appropriate because they recognize the inherent tradeoff between sensitivity and false positive rate when making binary judgments in the face of uncertainty (when compared with studies that artificially impose a particular decision threshold as being optimal).⁵ For IRA calculations, the authors used the Gwet agreement coefficient (AC) to measure the amount of beyond-chance agreement between experts. The Gwet AC has advantages over the more commonly used Cohen kappa IRA statistic, which suffers from counterintuitive behavior (Cohen kappa paradox) in the presence of high levels of agreement.¹⁰ They then used a bootstrap method to compute p values, testing whether differences in IRAs achieved using the various IED identification methods were statistically significant.

This study has additional strengths. It was well powered (>0.8) with a total of 100 patients, achieving specificity of 95% with a significance level of 5%. EEGs included in the study originated from consecutive patients, minimizing selection bias. Achieving a pretest probability of approximately 50%, similar numbers of patients with epileptic and nonepileptic events were included. Regarding expert consensus, there were 7 raters with substantial EEG experience (median 14 years). All experts were blinded to clinical data, reducing history bias. The authors who initially selected the EEG samples for in-study review did not participate in the rating process, removing potential confirmation bias. Last, EEG analytic methods were compared against classification of the patients' habitual paroxysmal events based on video-EEG recordings. This gold standard has significant advantages of being objective and externally validated. However, it also has disadvantages—namely concerning patient selection bias. This method is unable to account for EEG data from a significant number of patients evaluated for seizures—those who never require video-EEG monitoring. In addition, patients without habitual episodes captured during video-EEG monitoring and those with

inconclusive long-term monitoring were not included in this study. We believe that using a combined gold standard including both patients' video-EEG data and expert consensus¹¹ would be advantageous.¹² The latter has the advantage of being immune to patient selection; however, it carries a disadvantage of being grounded in expert experience rather than an externally validated and objective source. Ultimately, both methods complement each other and should be considered as a highly reliable gold standard for future studies.¹²

Other limitations of the study include the lack of assessment of the pattern of repetition throughout the recordings because the IFCN criteria do not include such a criterion. Moreover, it was shown that the IRA associated with some IFCN criteria (namely 3, 5, and 6) was fair to moderate, potentially compromising the utility of these criteria in identifying IEDs and their global implementation. Another limitation was the lack of randomization between review rounds because all raters underwent the same sequential order of review (sensor space, source space, and expert review), exposing raters to potential confirmation bias. This effect was minimized by presenting the samples in randomized order, spacing each round of analysis by 1 month, and blinding the raters to the gold standard until all rounds were completed.

Several studies have since expanded our understanding of the use of the IFCN criteria. For example, the combination of IFCN criteria 1, 4, and 6 generates high accuracy and IRA on identifying IEDs,¹³ and it is required that discharges meeting fewer IFCN criteria repeat in the EEG study to maintain optimal diagnostic accuracy.¹⁴ Last, teaching the IFCN criteria to trainees improves their diagnostic accuracy and IRA.¹⁵

In summary, the authors validated the use of the IFCN criteria (4 or 5/6) and source space analysis to identify IEDs with an accuracy similar to expert scoring. The implementation of the IFCN criteria and source analysis in clinical practice should contribute to decreased error and misdiagnosis of epilepsy, especially in challenging cases and/or in scenarios where IRA is suboptimal. Given known educational gaps in neurology residency EEG training,¹⁶ we believe residents should also receive dedicated training on how to use the IFCN criteria and source analysis. This training would be based on objective criteria, thus potentially bypassing the occasional issue faced by some institutions where there is insufficient faculty time to teach. Potential challenges to an implementation of these methods include the need to thoroughly train readers on their use and the subjective judgment required to score each IFCN criterion.

Study Funding

No targeted funding reported.

Disclosure

F.A. Nascimento is a former member of the Neurology Resident and Fellow Section Editorial Team. The other authors report no relevant disclosures. Go to [Neurology.org/N](https://www.neurology.org/N) for full disclosures.

Publication History

Received by *Neurology* January 27, 2022. Accepted in final form June 3, 2022. Submitted and externally peer reviewed. The handling editor was Whitley Aamodt, MD, MPH.

Appendix Authors

Name	Location	Contribution
John R. McLaren, MD	Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston	Drafting/revision of the article for content, including medical writing for content; study concept or design; and analysis or interpretation of data
Jin Jing, PhD	Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston	Drafting/revision of the article for content, including medical writing for content
M. Brandon Westover, MD, PhD	Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston	Drafting/revision of the article for content, including medical writing for content
Fábio A. Nascimento, MD	Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston; Department of Neurology, Washington University School of Medicine, St. Louis, MO	Drafting/revision of the article for content, including medical writing for content; study concept or design; and analysis or interpretation of data

References

- Benbadis SR, Tatum WO. Overinterpretation of EEGs and misdiagnosis of epilepsy. *J Clin Neurophysiol*. 2003;20(1):42-44.
- Benbadis SR. Errors in EEGs and the misdiagnosis of epilepsy: importance, causes, consequences, and proposed remedies. *Epilepsy Behav*. 2007;11(3):257-262.
- Kane N, Acharya J, Beniczky S, et al. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. *Clin Neurophysiol Pract*. 2017;2:170-185.
- Kural MA, Duez L, Sejer Hansen V, et al. Criteria for defining interictal epileptiform discharges in EEG: a clinical validation study. *Neurology*. 2020;94(20):e2139-e2147.
- Fawcett T. An introduction to ROC analysis. *Pattern Recognition Lett*. 2006;27(8):861-874.
- Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol*. 2018;63(7):07TR01.
- Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc*. 1927;22(158):209-212.
- McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 1947;12:153-157.
- Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008;61(pt 1):29-48.
- Wongpakaran N, Wongpakaran T, Wedding D, et al. A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol*. 2013;13:61.
- Jing J, Herlopian A, Karakis I, et al. Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms. *JAMA Neurol*. 2020;77(1):49-57.
- Nascimento FA, Jing J, Beniczky S, et al. One EEG, one read—a manifesto towards reducing interrater variability among experts. *Clin Neurophysiol*. 2022;133:68-70.
- Kural MA, Tankisi H, Duez L, et al. Optimized set of criteria for defining interictal epileptiform EEG discharges. *Clin Neurophysiol*. 2020;131(9):2250-2254.
- Kural MA, Qerama E, Johnsen B, Fuchs S, Beniczky S. The influence of the abundance and morphology of epileptiform discharges on diagnostic accuracy: how many spikes you need to spot in an EEG. *Clin Neurophysiol*. 2021;132(7):1543-1549.
- Kural MA, Aydemir ST, Levent HC, et al. The operational definition of epileptiform discharges significantly improves diagnostic accuracy and inter-rater agreement of trainees in EEG reading. *Epileptic Disord*. 2022;24(2):353-358.
- Nascimento FA, Gavvala JR. Education research: neurology resident EEG education: a survey of US neurology residency program directors. *Neurology*. 2021;96(17):821-824.