

Criteria for defining interictal epileptiform discharges in EEG

A clinical validation study

Mustafa Aykut Kural, MD, Lene Duez, MD, PhD, Vibeke Sejer Hansen, MD, PhD, Pål G. Larsson, MD, DMSc, Stefan Rampp, MD, PhD, Reinhard Schulz, MD, Hatice Tankisi, MD, PhD, Richard Wennberg, MD, PhD, Bo M. Bibby, PhD, Michael Scherg, PhD, and Sándor Beniczky, MD, PhD

Neurology® 2020;94:e2139-e2147. doi:10.1212/WNL.0000000000009439

Correspondence

Dr. Beniczky
sbz@filadelfia.dk

Abstract

Objective

To define and validate criteria for accurate identification of EEG interictal epileptiform discharges (IEDs) using (1) the 6 sensor space criteria proposed by the International Federation of Clinical Neurophysiology (IFCN) and (2) a novel source space method. Criteria yielding high specificity are needed because EEG over-reading is a common cause of epilepsy misdiagnosis.

Methods

Seven raters reviewed EEG sharp transients from 100 patients with and without epilepsy (diagnosed definitively by video-EEG recording of habitual events). Raters reviewed the transients, randomized, and classified them as epileptiform or nonepileptiform in 3 separate rounds: in 2, EEG was reviewed in sensor space (scoring the presence/absence of each IFCN criterion for each transient or classifying unrestricted by criteria [expert scoring]); in the other, review and classification were performed in source space.

Results

Cutoff values of 4 and 5 criteria in sensor space and analysis in source space provided high accuracy (91%, 88%, and 90%, respectively), similar to expert scoring (92%). Two methods had specificity exceeding the desired threshold of 95%: using 5 IFCN criteria as cutoff and analysis in source space (both 95.65%); the sensitivity of these methods was 81.48% and 85.19%, respectively.

Conclusions

The presence of 5 IFCN criteria in sensor space and analysis in source space are optimal for clinical implementation. By extracting these objective features, diagnostic accuracy similar to expert scorings is achieved.

Classification of evidence

This study provides Class III evidence that IFCN criteria in sensor space and analysis in source space have high specificity (>95%) and sensitivity (81%–85%) for identification of IEDs.

RELATED ARTICLE

Editorial

Epileptiform discharges:
Are we still defining them?
Page 862

MORE ONLINE

→ Class of Evidence

Criteria for rating
therapeutic and diagnostic
studies

[NPub.org/coe](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7281111/)

From the Department of Clinical Neurophysiology (M.A.K., L.D., V.S.H., H.T., S.B.), Aarhus University Hospital, Aarhus, Denmark; Department of Neurosurgery (P.G.L.), Rikshospitalet, Oslo University Hospital, Norway; Department of Neurosurgery (S.R.), University Hospital Erlangen, Germany; Department of Neurosurgery (S.R.), University Hospital Halle (Saale), Germany; Epilepsy Center Bethel (R.S.), Mara Hospital, Bielefeld, Germany; Krembil Brain Institute (R.W.), Toronto Western Hospital, University of Toronto, Canada; Department of Biostatistics (B.M.B.), Aarhus University, Denmark; Department of Research (M.S.), BESA GmbH, Gräfelfing, Germany; Department of Clinical Neurophysiology (S.B.), Danish Epilepsy Centre, Dianalund, Denmark; and Department of Clinical Medicine (S.B.), Aarhus University, Denmark.

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.

The Article Processing Charge was funded by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND), which permits downloading and sharing the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Glossary

AC1 = Gwet agreement coefficients; **IED** = interictal epileptiform discharge; **IFCN** = International Federation of Clinical Neurophysiology; **IRA** = interrater agreement; **LTM** = long-term video-EEG monitoring; **ROC** = receiver operating characteristic.

Visual analysis is still the gold standard in clinical EEG.¹ However, most recordings are interpreted by neurologists without expert certification.²

Interictal epileptiform discharges (IEDs) are typically found in patients with epilepsy.^{1,3–5} However, interpreting EEG is based on pattern recognition, and reading EEG is still more an art than a science.⁵ The body of knowledge is passed down by word of mouth from teacher to student, and EEG training is like an extensive process of probabilistic classification learning, where teachers provide reinforcement for the visually identified patterns.⁵

Overinterpretation of EEG is a major contributor to misdiagnosis^{2,6,7}: 25%–30% of patients seen at epilepsy centers for drug-resistant seizures do not have epilepsy.^{8–11} Misdiagnosis has multiple detrimental consequences for patients.^{12,13}

Normal sharp transients are often misread as IEDs.^{2,6,14} There have been several attempts to develop criteria for identifying IEDs,^{3,5,7,15–22} and the International Federation of Clinical Neurophysiology (IFCN) has recently proposed an operational definition of IEDs.²³ However, these criteria have not yet been clinically validated.

Traditionally, EEG is inspected in sensor space, using electrode montages. Using computational methods, signals can be transformed into source space^{24,25} to display activities of different brain regions.

This study aimed at assessing the diagnostic accuracy and interrater agreement (IRA) for identification of IEDs using (1) the IFCN criteria in sensor space and (2) a new standardized source space. Optimally, these methods would reduce over-reading; therefore, a specificity of >95% is needed.

We present infographics and detailed descriptions of these criteria to guide neurologists in the accurate identification of IEDs.

Methods

Study design and EEG recordings

We included EEG recordings from consecutive patients who underwent long-term video-EEG monitoring (LTM) at Aarhus University Hospital (Denmark) and at the Danish Epilepsy Centre (Dianalund, Denmark) between January 2012 and September 2017. Inclusion criteria were patients who had habitual paroxysmal events recorded during LTM (classified as epileptic seizures or as nonepileptic paroxysmal events) and

who had sharp transients (epileptiform or nonepileptiform) in their interictal recordings. For patients with epileptic seizures, there was an additional criterion: the IEDs recorded in the interictal period had to be concordant with the recorded ictal event (i.e., bilateral synchronous IEDs for generalized seizures in patients with generalized epilepsies and focal IEDs in the same lobar localization as the ictal recordings acquired in patients with focal onset seizures). Participants were included if they were aged 1 year or older. We excluded patients with diagnostically inconclusive LTM despite recorded habitual events and patients with both epileptic seizures and nonepileptic paroxysmal events. Because the results of the EEG interpretations were dichotomous, we aimed at including a similar number of patients with epilepsy and nonepileptic paroxysmal events so that the pretest probability was close to 50%. For the expected specificity of 95%, with a significance level of 5% and a power of 0.8, we needed at least 68 patients.²⁶

EEG was recorded with the standardized 25 electrode array of the IFCN²⁷ using a sampling frequency of 256 samples/s. Interictal EEG epochs, 10–20 seconds long, were selected by 2 authors (M.A.K. and S.B.) who did not participate in rating the samples according to the index tests (IFCN criteria and sensor space). Sharp transients previously marked (during the diagnostic evaluation process) were then reviewed by the 2 authors, and the first marking confirmed by both experts as a sharp transient (epileptiform or not) was selected for further evaluation, provided it fulfilled the selection criterion: a transient with pointed peak and phase reversal in a bipolar montage. For patients with epilepsy, there was an additional criterion: the selected transient had to be concordant with a patient's recorded seizure, implying the same seizure type and epilepsy syndrome and, for focal seizures, the same localization as the ictal recording. We chose epochs of 10–20 seconds to provide the raters sufficient EEG background before and after the sharp transient they had to score, corresponding to 1 screen display for each example of transient.

Standard protocol approvals, registrations, and patient consents

The regional ethics committee reviewed and cleared this project, in which only deidentified patient data were collected. According to Danish regulations, this study did not need written informed consent from the patients because it involved a retrospective analysis of anonymized data.

The study was registered in ClinicalTrials.gov (Unique Protocol ID: CIED; identifier: NCT03533374). We report the study according to the Standards for the Reporting of Diagnostic Accuracy Studies guidelines.²⁸

Procedures

Seven raters (H.T., L.D., P.G.L., R.S., R.W., S.R., and V.S.H.) with experience in reading EEG (median 14 years; range: 2–37 years) evaluated the samples, blinded to all other data (including the gold standard). The 2 authors who selected the samples for evaluation (as described above) did not participate in rating the samples. Previous studies suggested that the optimal size for a group of experts scoring IEDs in EEG was between 6 and 10.²⁹ Raters independently evaluated the EEG samples in 3 different rounds (more than 1 month apart), in sensor space and in source space, the samples presented in different randomized order in each round. The raters did not know how many samples were from patients with epilepsy.

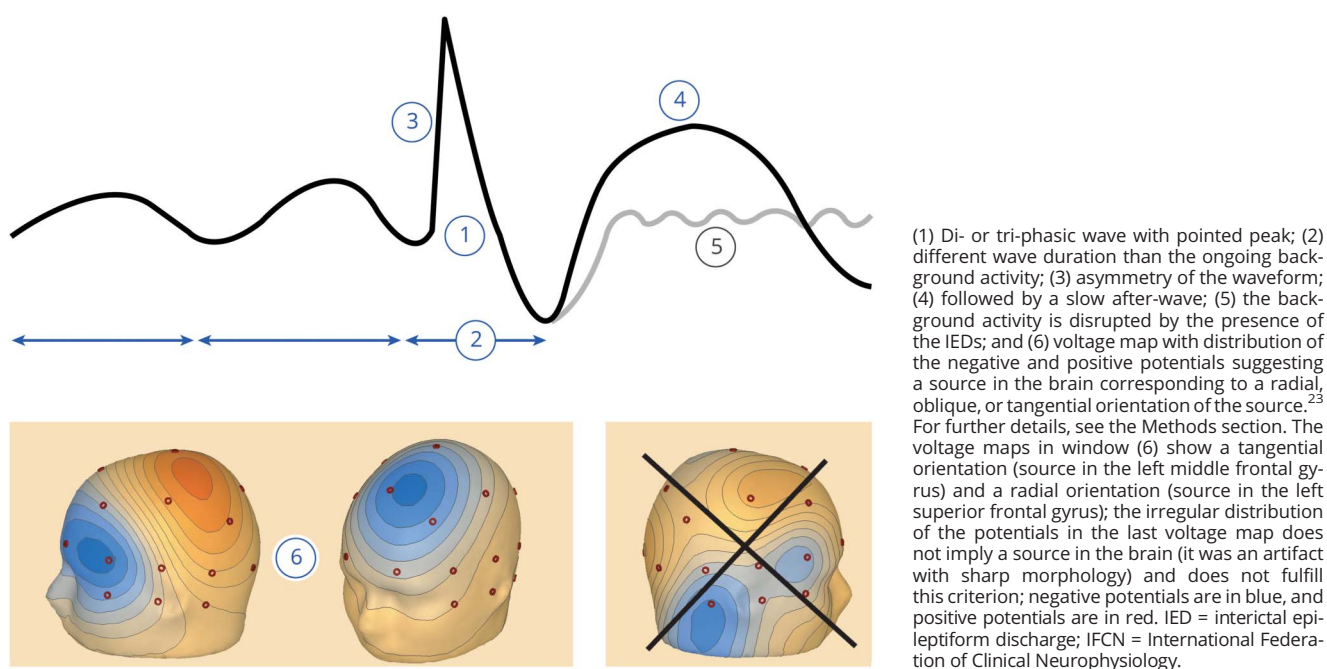
Sensor space refers to the traditional way of inspecting EEG, using montages of scalp EEG electrodes (sensors), which is a 2-dimensional representation. In the first round, raters visually extracted the EEG features corresponding to each IFCN criterion in sensor space²³ and scored the presence or absence of each criterion, without concluding whether the waveform was epileptiform. These criteria were as follows: (1) di- or tri-phasic 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 after-wave; (5) the background activity surrounding IEDs is disrupted by the presence of the IEDs; and (6) 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.²¹

Figure 1 is an infographic summarizing the IFCN criteria for IEDs in sensor space. The raters inspected the selected epochs containing the marked sharp transients using a commercially available EEG reader (BESA Research 6.1); they were allowed to switch between montages in sensor space (longitudinal and transversal bipolar and common average), change gain, time scale (starting at 30 mm/s), and digital filters. Single clicking on the peak of the sharp transients displayed a 3-dimensional voltage map, including isoelectric lines.

In the second round, raters inspected the EEG samples in source space in a different randomized order. Using a spatial filtering method,^{24,25} the scalp EEG samples were transformed into a standardized source space corresponding to 25 regions of interest in the brain. In this process, a generalized linear operator combines the scalp EEG signals in a specific way to enhance the activity from one region while reducing the volume-conducted overlap from other areas.^{24,25} An icon at the left side of each displayed source trace symbolizes the corresponding brain region. Both superficial cortical (radially oriented) and deep (tangentially oriented) sources (in sulci) are included into the brain regions to localize the irritative zone at a sublobar level. A detailed technical description of this is available from Dryad (Supplementary material 1, available on Dryad, doi.org/10.5061/dryad.xsj3tx99w).

For the identification of IEDs in source space, the following operational definition was used by the raters: Epileptiform discharges are abnormal sharp transients, with characteristic (1) changes in time and (2) distribution in space, (3) excluding artifacts and normal variants. To qualify as an IED, all

Figure 1 Infographic summarizing the 6 IFCN criteria for identifying IEDs



of these features had to be present. Raters were provided with the following instructions: “Changes in time: the sharp transient is clearly distinguishable from the ongoing (background) activity, based on its amplitude, duration and morphology. Check this in the source channel(s) where the discharge is visible. Changes in space: the location of the signal should make sense anatomically. Check this across all channels: Regions with the same polarity should be close together. Opposite polarity may be observed in the other hemisphere or more remote regions. Caveat 1: temporal lobe traces may show opposite polarities. Caveat 2: Tangential IEDs in fissures can show polarity reversal between neighboring regions. If unclear, compare time-lag between the peaks in different channels to observe presence of propagation as an additional indicator of epileptiform transients. Presence of propagation is optional (i.e., not compulsory for all spikes).” Supplementary material 2 (available on Dryad, doi.org/10.5061/dryad.xsj3tx99w) contains an infographic summarizing how IEDs are identified in source space.

Finally, the experts scored the marked waveforms as IEDs or nonepileptiform sharp transients, unrestricted by any criteria, and solely based on their expertise, as if evaluated in their clinical practice (expert scorings).

Before starting scoring of the test samples, the raters went through open-label training samples (5 recordings with IEDs and 5 with nonepileptiform sharp transients). The training samples and the written instructions given to the raters in the training period are available on Dryad (Supplementary material 2, doi.org/10.5061/dryad.xsj3tx99w). The whole data set can be downloaded using this link: datadryad.org/stash/share/Fj1923GZtwudCKU2_ltcFinVpcVOHhYudHWIkWzQ538.

Evaluation of the extracted features

For each extracted feature, the consensus majority scorings were used for determining diagnostic accuracy. For the IFCN criteria in sensor space, we used cutoff values of 2–6 fulfilled criteria for qualifying a waveform as an IED. Analysis in source space and expert scoring resulted in dichotomous scores (IEDs or nonepileptiform sharp transients).

The gold standard (clinical diagnostic reference standard) for classification of patients as having epilepsy or having nonepileptic paroxysmal events was derived from the video-EEG recordings of the patients’ habitual clinical events.

We used the standard formulas: sensitivity = $TP / (TP + FN)$, specificity = $TN / (TN + FP)$, and accuracy = $(TP + TN) / \text{all patients}$ (where FN = false negative, FP = false positive, TN = true negative, and TP = true positive).

Statistical analysis

Based on the sensitivities and specificities obtained at cutoff values of 2–6 IFCN criteria in sensor space, we calculated the estimated receiver operating characteristic (ROC) curve.³⁰ We calculated 95% confidence intervals for sensitivity, specificity,

and accuracy using the Wilson method,³¹ and we compared them between the IED identification methods using the McNemar test based on the true positive, the true negative, and both.³²

We calculated IRA using Gwet³³ agreement coefficient (AC1) to avoid the paradoxes of kappa. IRA was interpreted according to the conventional criteria: poor (<0.02), fair (0.2–0.4), moderate (0.4–0.6), substantial (0.6–0.8), and almost perfect agreement (>0.8). We compared IRAs of the IED identification methods using bootstrap.³⁴ The analyses were performed using Stata version 15.1.

Classification of evidence

This diagnostic accuracy study was designed to answer the following question: What is the accuracy of the IFCN criteria in sensor space, and of analysis in source space, for identifying epileptiform EEG discharges?

This study provides Class III evidence that IFCN criteria in sensor space and analysis in source space have high specificity (>95%) and sensitivity (81%–85%) for identification of IEDs. The study was retrospective and blinded. All EEG samples, selected according to the criteria described in detail in the Methods section, were analyzed and evaluated. All patients had the diagnostic reference standard inferred from an independent data set and method (recordings of the patients’ habitual clinical episode) that is the most reliable diagnostic test for distinguishing epileptic from nonepileptic paroxysmal events. The included patients had a broad spectrum of diagnoses, with and without epilepsy. The following 3 procedures were performed by different experts, using different methods and blinded for each other: selection of the EEG samples, evaluation of the index test, and evaluation of the diagnostic reference standard (gold standard).

Data availability

Individual deidentified EEG data in EDF format (European Data Format) with the marked sharp transients of all patients included in this study will be shared, along with the following related documents: demographic data of the patients, gold standard of each patient, detailed technical description of signal reconstruction in source space, and STARD flow diagrams. Unrestricted access to these data will be made available from the day of the online publication of the article, until 2030, using a publicly accessible repository on Dryad (doi.org/10.5061/dryad.xsj3tx99w).

Results

We assessed EEG samples from 100 patients (60 female; age: 2–89 years, median: 31 years) who fulfilled the inclusion criteria. According to the reference standard, 54 patients had epilepsy (focal: 45; generalized: 7; both focal and generalized: 2), and 46 patients had nonepileptic paroxysmal events (psychogenic nonepileptic seizures: 31; parasomnia: 7; paroxysmal movement disorders: 4; syncope: 4).

Examples of IEDs and nonepileptiform sharp transients are shown in figure 2. The STARD diagrams and additional examples are available from Dryad (Supplementary materials 3 and 4, available on Dryad, doi.org/10.5061/dryad.xsj3tx99w).

The area under the ROC curve for IFCN criteria in sensor space was 0.977 for the consensus majority scorings (figure 3). The area under ROC curves for the individual raters was between 0.916 and 0.965 (median: 0.941) (Supplementary material 5, available on Dryad, doi.org/10.5061/dryad.xsj3tx99w). Table 1 shows the sensitivity, specificity, and accuracy of identifying IEDs using the IFCN criteria in sensor space with cutoff values between 2 and 6, the analysis in source space, and the expert scorings. The ROC curve showed that 3 methods (IFCN criteria with cutoff values of 4 and 5 and analysis in source space) achieved high diagnostic performance, similar to expert scorings (figure 3), and there was no statistically significant difference in the accuracy of these methods ($p > 0.157$). Using a cutoff value of 5 IFCN criteria resulted in significantly higher specificity ($p = 0.025$) but significantly lower sensitivity ($p = 0.005$) compared with the cutoff value of 4 IFCN criteria. All other comparisons in sensitivity and specificity between the 3 methods with high diagnostic performance and the expert scorings remained below the level of statistical significance.

Two methods achieved specificity higher than the expected threshold of 95%: IFCN criteria with a cutoff value of 5 and analysis in source space (figure 3 and table 1).

Agreement coefficients of the methods of identifying IEDs were between 0.490 and 0.830 (table 1). IRA of the 3 methods with high diagnostic performance was moderate to substantial

(AC1: 0.490–0.608), and there was no significant difference in IRA between them ($p > 0.900$).

On their own, none of the IFCN criteria qualifies a transient as epileptiform or not (it is the number of the fulfilled criteria that are diagnostic). Nevertheless, to assess how reliable visual analysis of the individual criteria was, we calculated IRA of each criterion (table 2). IRA of criteria numbers 1, 2, and 4 was substantial; it was moderate for criterion 6 and fair for criteria numbers 3 and 5.

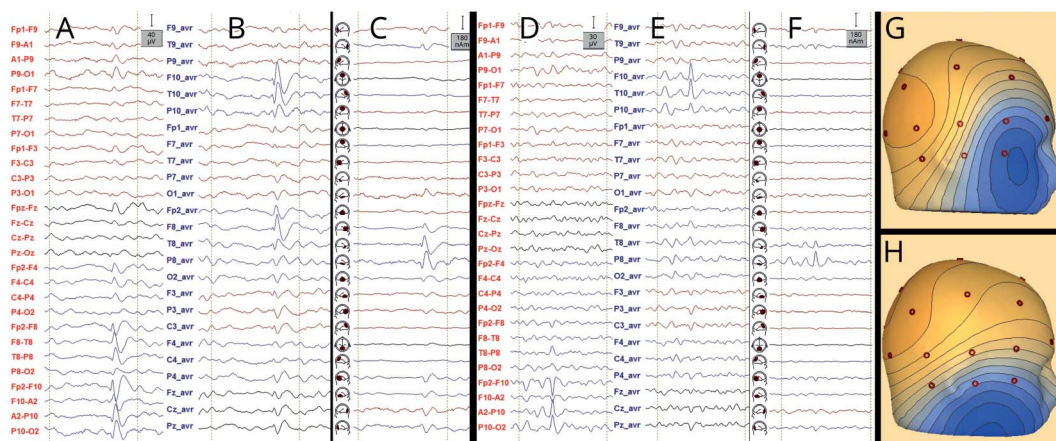
Using the consensus expert scorings as a reference standard (rather than the clinical diagnostic gold standard), the highest accuracies were achieved using a threshold of 5 IFCN criteria in sensor space and evaluating the signals in source space—both methods had an accuracy of 93%. A table showing these results is available from Dryad (Supplementary material 6, available on Dryad, doi.org/10.5061/dryad.xsj3tx99w).

Discussion

We found that the set of 6 IFCN criteria for identifying IEDs had high value of the area under the ROC curve (close to 1), demonstrating the high diagnostic ability of these criteria to discriminate between IEDs and nonepileptiform sharp transients. Three methods achieved high diagnostic performance, with accuracy similar to expert scorings: analysis in sensor space using cutoff values of 4 or 5 IFCN criteria and analysis in source space. The IRA of these methods was moderate to substantial.

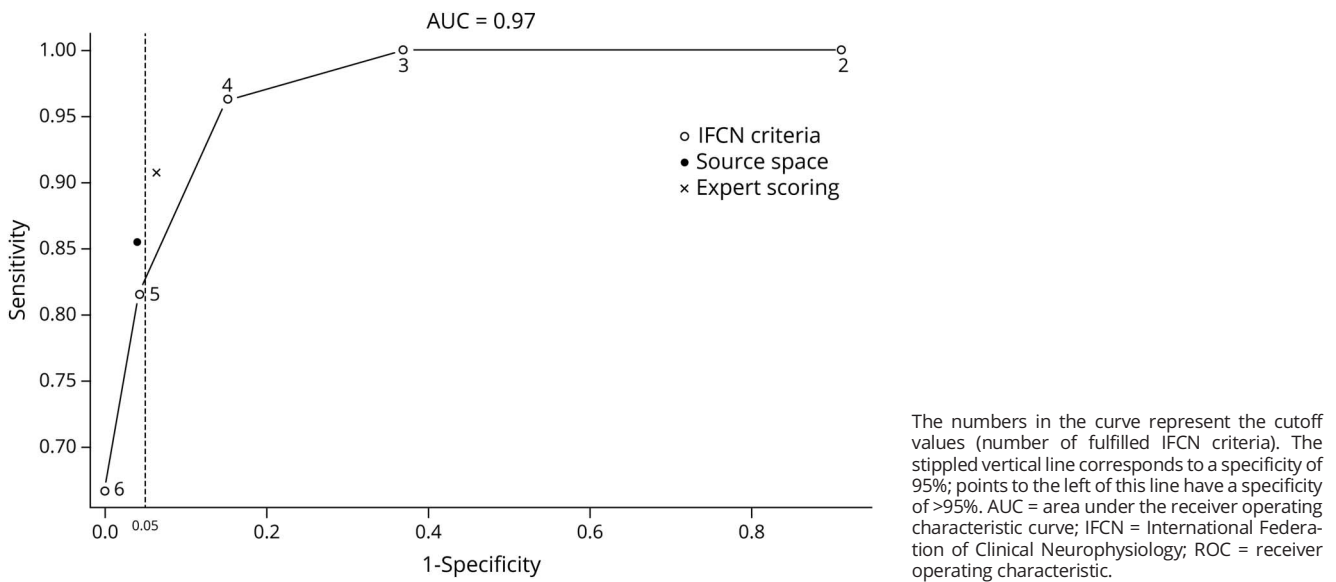
It has been suggested that 4 IFCN criteria should be fulfilled to qualify a transient as an IED.²³ Although the cutoff value of

Figure 2 Examples of IED (A–C) and nonepileptiform sharp transient (D–F)



Montages: longitudinal bipolar (A and D), common average (B and E), and source space (C and F). Voltage maps are shown in G and H. The transient shown in A–C with voltage map in G (oblique distribution corresponding to a source in the right temporal pole) fulfills all IFCN criteria in sensor space, except for criterion 5, and hence, it qualifies as an IED; in source space, it also qualifies as an IED, and the propagation from the temporal pole to the basal temporal region can be observed (C). The transient shown in D–F with voltage map in H only fulfills 2 IFCN criteria in sensor space (1 and 6), and hence, it does not qualify as an IED; in source space, one can observe that the transient belongs to the background activity from the right basal temporal region (fragmented during drowsiness); the orientation of the voltage map corresponds to a source in the right basal temporal region (fulfilling criterion 6). Voltage maps are useful in distinguishing IEDs from artifacts (figure 1); however, nonepileptiform transients originating from the brain can show voltage distributions similar to IEDs. IED = interictal epileptiform discharge; IFCN = International Federation of Clinical Neurophysiology.

Figure 3 Receiver operating characteristic (ROC) curve



4 criteria had high accuracy, the specificity was only 85%, which is below the threshold of 95% expected for clinical implementation. Only 2 methods had specificity over 95%: analysis in sensor space using 5 IFCN criteria as cutoff and analysis in source space. The sensitivity of these 2 methods was 81% and 85%, respectively.

There was substantial IRA on extracting 3 EEG features corresponding to IFCN criteria 1, 2, and 4: the spiky morphology, wave duration distinct from the background activity, and the presence of the slow wave following the transient. Assessment of the voltage maps (criterion 6) had moderate IRA, whereas assessment of the wave asymmetry and the

disruption of the background activity (criteria 3 and 5) had only fair IRA. According to the IFCN operational definition, for qualifying as an IED, a transient has to fulfill a number of these criteria, regardless which criteria are fulfilled. When analyzing the IRA of qualifying a transient as an IED using 4 or 5 fulfilled criteria as cutoff, the agreement coefficient was 0.490 (moderate) and 0.608 (substantial), respectively.

In addition to the traditional evaluation in sensor space, we have assessed a novel approach of inspecting EEG signals in source space. In our opinion, it provides a clearer view onto the signals of the underlying brain areas compared with the scalp signals in sensor space (figure 2), which are more blurred by

Table 1 Sensitivity, specificity, accuracy, and IRA of the methods for identifying IEDs in sensor space and source space compared with expert scorings

Method	Sensitivity % (95% CI)	Specificity % (95% CI)	Accuracy % (95% CI)	AC1 (95% CI)
IFCN criteria in sensor space				
Cutoff (no. of fulfilled criteria)				
≥2	100 (93.40–100)	8.70 (2.42–20.79)	58 (47.71–67.80)	0.83 (0.76–0.88)
≥3	100 (93.40–100)	63.04 (47.55–76.79)	83 (74.18–89.77)	0.64 (0.53–0.74)
≥4	96.30 (87.25–99.55)	84.78 (71.13–93.66)	91 (83.60–95.80)	0.60 (0.53–0.69)
≥5	81.48 (68.57–90.75)	95.65 (85.16–99.47)	88 (79.98–93.64)	0.49 (0.38–0.60)
=6	66.67 (52.53–78.91)	100 (92.29–100)	82 (73.05–88.97)	0.60 (0.50–0.70)
Source space	85.19 (72.88–93.38)	95.65 (85.16–99.47)	90 (82.38–95.10)	0.59 (0.51–0.68)
Expert scorings	90.74 (79.70–96.92)	93.48 (82.10–98.63)	92 (84.84–96.48)	0.58 (0.50–0.67)

Abbreviations: AC1 = Gwet agreement coefficient; CI = confidence interval; IED = interictal epileptiform discharge; IFCN = International Federation of Clinical Neurophysiology; IRA = interrater agreement.

Table 2 IRA of the 6 IFCN criteria, scored individually

IFCN criterion	AC1 (95% CI)
1	0.75 (0.68–0.81)
2	0.63 (0.53–0.72)
3	0.30 (0.22–0.38)
4	0.65 (0.57–0.73)
5	0.36 (0.30–0.45)
6	0.51 (0.42–0.60)

Abbreviations: AC1 = Gwet agreement coefficient; CI = confidence interval; IFCN = International Federation of Clinical Neurophysiology; IRA = interrater agreement.

overlap from remote regions. This property might be the reason for the slightly higher sensitivity and the high specificity, comparable to the best performing method in sensor space. Besides identification of IEDs, source space montages also help in localization, because they show the brain regions generating the discharges (figure 2), and aid in the separation of sublobar activities, for example, involving different parts of the temporal lobe, as documented previously.^{24,25} Thus, interpreting EEG in source space may be more intuitive and easier to explain to young trainees.²⁵

The previous IFCN definitions of IEDs were conceptual,^{20,21} lacking specific rules for clinical implementation. Mulsby²² proposed a set of practical guidelines for assessment of IEDs. Many of them translate to the current criteria: “Every spiky-looking wave is an artifact unless there are one or more good reasons for suspecting otherwise” (similar to criterion 3 in source space); IEDs “occupy a definable electrical field on the scalp” (similar to criterion 6 in sensor space); IEDs “are followed by a slow wave or series of slow deflections” (criterion 4 in sensor space); ignore sharp transients “which can be logically explained by simple alterations in voltage of the existing background rhythm” (criterion 2 in sensor space). Most of the current criteria can be traced back to Mulsby proposals²² and to those of Pierre Gloor.¹⁵

Because of the need for a reliable reference standard, we analyzed interictal EEG epochs extracted from LTM recordings. The odds for recording sharp transients (either epileptiform or not) are higher in long-term recordings compared with short, routine EEG recordings. Nevertheless, the specific aspect of EEG interpretation addressed in this study was the distinction between epileptiform and nonepileptiform sharp transients, and therefore, our results can be generalized to all instances where sharp transients are observed in EEG recordings of any duration.

A limitation in this study was that we did not assess the pattern of repetition (similarity of the wave morphology and frequency of occurrence) throughout the recordings. Because the IFCN criteria address features related to a single discharge and do not include a criterion on the repetition of the

waveforms, we designed the study to reflect that. Moreover, in clinical practice, one needs to classify a transient as epileptiform or not, even when only one or few discharges are observed in a routine EEG recording. Another limitation is that patients with generalized epilepsies are underrepresented in our series, which is a reflection of the general pattern of referring patients for diagnostic LTM. However, distinguishing generalized IEDs from nonepileptiform sharp transients is a lesser problem in clinical practice.^{2,6}

In this study, we opted for an unequivocal diagnostic gold standard: classification of the patients’ habitual paroxysmal clinical episodes, based on the video-EEG recordings. This resulted in a potential risk of spectrum bias because patients without habitual episodes during the monitoring and patients with inconclusive LTM were excluded. We considered it more important to have an unequivocal gold standard than to include all patients.

Raters evaluated the data set 3 different times, and repeated evaluation is a potential source of bias. We attempted to decrease this bias by presenting the samples in different, randomized order, by separating the evaluation rounds by at least 1 month, and by keeping the raters blinded to the gold standard until all 3 rounds of evaluation were completed.

The rate of IEDs in routine EEG recordings obtained in nonepileptic individuals without progressive neurologic disorder has been reported as 0.5%–2.6% in adults^{4,35,36} and 2.2%–3.5% in children.^{4,35} A much higher rate of 12% was found in a patient population where most patients had progressive neurologic disorders (known to have a higher rate of IEDs) but no history of seizure.³⁷ All our patients had a history of paroxysmal episodes, but none of our patients without epilepsy had progressive neurologic disorder. The sensitivity of single, routine EEG recording is relatively low (32%–59%).^{4,38,39} However, our data set was derived from LTM. Repeated routine recordings and long-term video-EEG recordings increase the sensitivity of EEG for IEDs to 80%–90%.^{1,4,39}

Although at first glance, the set of IFCN criteria might seem too time consuming to apply in routine clinical practice, with some experience, they are actually rapidly and naturally applied during visual EEG assessment. We provide here infographics (figure 1) to facilitate this. Systematic application of these criteria is especially important in difficult cases, and it provides guidance when disagreement occurs between readers.

In our experience, assessing EEG in source space is even more rapid than the traditional way of inspecting EEG in sensor space. Although the computation is complex, for the EEG reader, switching to source space is not time consuming (it takes the same time as switching from one montage to the other).

The currently available software tools for automated spike detection typically have low specificity (many false detections) but relatively high sensitivity.⁴⁰ Therefore, there is

a need for validating the automatically detected spikes by experts, who provide the high specificity by rejecting the false detections. However, interrater reliability among experts is variable. The criteria described in this article will be useful for guiding neurologists when validating the automatically detected spikes. By grouping similar types of spikes into clusters, the software tools can significantly reduce the time burden for reviewing the detected spikes.⁴⁰

There is a need for more precise definition of the 2 EEG features that had only fair IRA (asymmetry of the waveform and disruption of the background activity). Further improvement of the methods could be achieved by adding to the criteria features related to the repetition of the discharge in the recording. The features described in this article are potential targets for development of automated algorithms for spike detection. We grant unrestricted access to the deidentified EEG samples collected for this study to facilitate further development in this field.

In conclusion, the presence of EEG features corresponding to 5 fulfilled IFCN criteria in sensor space and analysis in source space have high diagnostic accuracy. Their implementation in clinical practice will contribute to decreasing overinterpretation of EEG and misdiagnosis of epilepsy.

Acknowledgment

The authors thank all medical staff of involved centers for taking best care of their patients.

Study funding

The study was supported by a grant from Henry and Karla Hansen Foundation.

Disclosure

M.A. Kural, L. Duez, and V.S. Hansen report no relevant disclosures. P.G. Larsson reports personal fees from Epilog, outside the submitted work. S. Rampp reports personal fees from Elekta Oy, outside the submitted work. R. Schulz, H. Tankisi, R. Wennberg, and B.M. Bibby report no relevant disclosures. M. Scherg is employee and shareholder of BESA GmbH. S. Beniczky reports personal fees from UCB Pharma, Eisai, Brain Sentinel, and UNEEG, outside the submitted work. Go to Neurology.org/N for full disclosures.

Publication history

Received by *Neurology* August 8, 2019. Accepted in final form December 10, 2019.

Appendix Authors

Name	Location	Contribution
Mustafa Aykut Kural, MD	Aarhus University Hospital, Denmark	Designed and conceptualized the study; analyzed the data; interpreted the results; and drafted the manuscript for intellectual content.

Appendix (continued)

Name	Location	Contribution
Lene Duez, MD, PhD	Aarhus University Hospital, Denmark	Analyzed the data and revised the manuscript for intellectual content.
Vibeke Sejer Hansen, MD, PhD	Aarhus University Hospital, Denmark	Analyzed the data and revised the manuscript for intellectual content.
Pål G. Larsson, MD, DMSc	Oslo University Hospital, Norway	Analyzed the data and revised the manuscript for intellectual content.
Stefan Rampp, MD, PhD	University Hospital Erlangen and University Hospital Halle, Germany	Analyzed the data and revised the manuscript for intellectual content.
Reinhard Schulz, MD	Epilepsy Center Bethel, Mara Hospital, Bielefeld, Germany	Analyzed the data and revised the manuscript for intellectual content.
Hatice Tankisi, MD, PhD	Aarhus University Hospital, Denmark	Analyzed the data and revised the manuscript for intellectual content.
Richard Wennberg, MD, PhD	Krembil Brain Institute, Toronto Western Hospital, University of Toronto, Canada	Analyzed the data and revised the manuscript for intellectual content.
Bo M. Bibby, PhD	Aarhus University, Denmark	Performed biostatistical analysis; interpreted the results; and revised the manuscript for intellectual content.
Michael Scherg, PhD	Department of Research, BESA GmbH, Gräfelting, Germany	Constructed source space montages; interpreted the results; and revised the manuscript for intellectual content.
Sándor Beniczky, MD, PhD	Aarhus University Hospital, Danish Epilepsy Centre, Dianalund, and Aarhus University, Denmark	Designed and conceptualized the study; analyzed the data; interpreted the results; and drafted the manuscript for intellectual content.

References

- Tatum WO, Rubboli G, Kaplan PW, et al. Clinical utility of EEG in diagnosing and monitoring epilepsy in adults. *Clin Neurophysiol* 2018;129:1056–1082.
- Benbadis SR, Lin K. Errors in EEG interpretation and misdiagnosis of epilepsy. Which EEG patterns are overread? *Eur Neurol* 2008;59:267–271.
- Pillai J, Sperling MR. Interictal EEG and the diagnosis of epilepsy. *Epilepsia* 2006;47(suppl 1):14–22.
- Hoppe M, Wennberg R, Tai P, Pohlmann-Eden B. EEG in epilepsy; in Lozano AM, Gildenberg PL, Tasker RR, eds. *Textbook of Stereotactic and Functional Neurosurgery*. Berlin Heidelberg: Springer, 2009:2575–2585.
- Engel J Jr. A practical guide for routine EEG studies in epilepsy. *J Clin Neurophysiol* 1984;1:109–142.
- Benbadis SR, Tatum WO. Overinterpretation of EEGs and misdiagnosis of epilepsy. *J Clin Neurophysiol* 2003;20:42–44.
- Benbadis SR. Errors in EEGs and the misdiagnosis of epilepsy: importance, causes, consequences, and proposed remedies. *Epilepsy Behav* 2007;11:257–262.
- Uldall P, Alving J, Hansen LK, Kibaek M, Buchholt J. The misdiagnosis of epilepsy in children admitted to a tertiary epilepsy centre with paroxysmal events. *Arch Dis Child* 2006;91:219–221.
- Asano E, Pawlak C, Shah A, et al. The diagnostic value of initial video-EEG monitoring in children—review of 1,000 cases. *Epilepsy Res* 2005;66:129–135.
- McBride AE, Shih TT, Hirsch LJ. Video-EEG monitoring in the elderly: a review of 94 patients. *Epilepsia* 2002;43:165–169.
- Chadwick D, Smith D. The misdiagnosis of epilepsy. *BMJ* 2002;324:495–496.
- Ferrie CD. Preventing misdiagnosis of epilepsy. *Arch Dis Child* 2006;91:206–209.
- LaFrance WC Jr, Benbadis SR. Avoiding the costs of unrecognized psychological nonepileptic seizures. *Neurology* 2006;66:1620–1621.

14. Chadwick D. Diagnosis of epilepsy. *Lancet* 1990;336:291–295.
15. Gloor P. The EEG in the differential diagnosis of epilepsy; in Van Duyn H, Donker DNJ, Van Huffelen AC, editors. *Current Concepts in Clinical Neurophysiology*. The Hague: Drukker Trio, 1977:9–21.
16. Drury I, Beydoun A. Pitfalls of EEG interpretation in epilepsy. *Neurol Clin* 1993;11:857–881.
17. Tatum WO, Husain AM, Benbadis SR, Kaplan PW. Normal adult EEG and patterns of uncertain significance. *J Clin Neurophysiol* 2006;23:194–207.
18. Mizrahi EM. Avoiding the pitfalls of EEG interpretation in childhood epilepsy. *Epilepsia* 1996;37(suppl 1):S41–S51.
19. Pedley TA. Interictal epileptiform discharges: discriminating characteristics and clinical correlations. *Am J EEG Technol* 1980;20:101–119.
20. Noachtar S, Binnie C, Ebersole J, Mauguière F, Sakamoto A, Westmoreland B. A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings. *The International Federation of Clinical Neurophysiology. Electroencephalogr Clin Neurophysiol Suppl* 1999;52:21–41.
21. Storm van Leeuwen W, Bickford R, Brazier M, et al. Proposal for an EEG terminology by the Terminology Committee of the International Federation for Electroencephalography and Clinical Neurophysiology. *Electroenceph Clin Neurophysiol* 1966;20:293–320.
22. Mulsby RL. Some guidelines for assessment of spikes and sharp waves in EEG tracings. *Amer J EEG Technol* 1971;11:3–16.
23. 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.
24. Scherg M, Ille N, Bornfleth H, Berg P. Advanced tools for digital EEG review: virtual source montages, whole-head mapping, correlation, and phase analysis. *J Clin Neurophysiol* 2002;19:91–112.
25. Scherg M, Berg P, Nakasato N, Beniczky S. Taking the EEG back into the brain: the power of multiple discrete sources. *Front Neurol* 2019;10:855.
26. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005;58:859–862.
27. Seeck M, Koessler L, Bast T, et al. The standardized EEG electrode array of the IFCN. *Clin Neurophysiol* 2017;128:2070–2077.
28. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41–44.
29. Halford JJ, Arain A, Kalamangalam GP, et al. Characteristics of EEG interpreters associated with higher Interrater agreement. *J Clin Neurophysiol* 2017;34:168–173.
30. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Lett* 2006;27:861–874.
31. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927;22:209–212.
32. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947;12:153–157.
33. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol* 2008;61:29–48.
34. Vanbelle S, Albert A. A bootstrap method for comparing correlated kappa coefficients. *J Stat Comput Simul* 2008;78:1009–1015.
35. Walczak TS, Jayakar P. Interictal EEG. In: Engel JJ, Pedley TA, editors *Epilepsy: A Comprehensive Textbook*. Philadelphia: Lippincott-Raven, 1997:831–848.
36. Gregory RP, Oates T, Merry RT. Electroencephalogram epileptiform abnormalities in candidates for aircrew training. *Electroencephalogr Clin Neurophysiol* 1993;86:75–77.
37. Sam MC, So EL. Significance of epileptiform discharges in patients without epilepsy in the community. *Epilepsia* 2001;42:1273–1278.
38. Goodin DS, Aminoff MJ. Does the interictal EEG have a role in the diagnosis of epilepsy? *Lancet* 1984;14:837–839.
39. Salinsky M, Kanter R, Dasheiff RM. Effectiveness of multiple EEGs in supporting the diagnosis of epilepsy: an operational curve. *Epilepsia* 1987;28:331–334.
40. Scherg M, Ille N, Weckesser D, et al. Fast evaluation of interictal spikes in long-term EEG by hyper-clustering. *Epilepsia* 2012;53:1196–1204.