Original article

Epileptic Disord 2022; 24 (2): 353-358

Epileptic Disorders

The operational definition of epileptiform discharges significantly improves diagnostic accuracy and inter-rater agreement of trainees in EEG reading

Mustafa Aykut Kural[1], Sabiha Tezcan Aydemir[2], Hafize Cotur Levent[2], Büsra Ölmez[2], Inci Sule Özer[2], Maria Vlachou[3], Agnes Hauschultz Witt[3], Abdullah Yasir Yilmaz[2], Sándor Beniczky[1,4]

[1] Department of Clinical Neurophysiology, Aarhus University Hospital, Aarhus, Denmark
[2] Department of Clinical Neurology, Ankara University, Ankara, Turkey
[3] Department of Neurology, Viborg Regional Hospital, Viborg, Denmark
[4] Department of Clinical Neurophysiology, Danish Epilepsy Centre, Dianalund, Denmark and Department of Clinical Medicine, Aarhus University, Member of the ERN EpiCARE

Received June 30, 2021; Accepted October 18, 2021

VIDEO ONLINE

Correspondence:
Sándor Beniczky and Mustafa Aykut Kural
Aarhus University Hospital and Danish Epilepsy Centre,
Visbý Allé 5, 4293 Dianalund, Denmark
<mustafaaykutkural@gmail.com>
<sbz@filadelfia.dk>
<sandor.beniczky@clin.au.dk>

doi:10.1684/epd.2021.1395

ABSTRACT

Objective. To assess whether trainees can learn and implement the operational definition of interictal epileptiform discharges (IEDs) of the International Federation of Clinical Neurophysiology (IFCN), based on six morphological criteria, and whether its implementation improves their diagnostic performance and inter-rater agreement (IRA).

Methods. Seven trainees evaluated a balanced dataset of 70 EEG samples containing sharp transients (35 from patients with epilepsy and 35 from patients with non-epileptic paroxysmal events). The gold standard was derived from video-EEG recordings of the habitual clinical episodes. The trainees individually reviewed the EEGs, blinded to all other data, in two successive training sessions, three months apart. The second session was preceded by a teaching module about the IFCN criteria, and the trainees implemented them during the second reading session.

Results. By implementing the IFCN criteria, trainees significantly improved their specificity (94.29% vs. 77.14%; p=0.01) and overall accuracy (81.43% vs. 64.29%; p=0.01) for identifying IEDs. Sensitivity also improved but did not reach the level of statistical significance (77.14% vs. 60%; p=0.07). IRA improved significantly from fair (k=0.31; 95% CI: 0.22-0.40) to high-moderate (k=0.56; 95% CI:0.46-0.67) beyond-chance agreement.

Significance. Implementing the IFCN criteria significantly improves the diagnostic performance and IRA of trainees in identifying IEDs. Teaching the IFCN criteria for IEDs will increase specificity in clinical EEG and avoid over-reading, the most common cause of misdiagnosing epilepsy.

Key words: diagnostic accuracy, definition, criteria, EEG, interictal epileptiform discharges, teaching

Electroencephalography (EEG) is the most commonly used investigative method in patients with epilepsy [1]. Interictal epileptiform discharges (IEDs) confirm the diagnosis [2, 3], predict recurrence of seizures [4], and help classifying epilepsy [5]. In spite of advances in automated signal analysis

[6, 7], algorithms have not yet achieved sufficient accuracy and reliability for clinical implementation, and visual assessment remains the clinical method of EEG reading [8]. However, this requires expertise, which is not available everywhere. Most EEGs in the United States are read by neurologists who have not had fellowship training in EEG, and a recent study identified a lack of consistency in EEG training of neurology residents [9].

Routine interictal EEG recording is one of the most abused investigations in clinical practice: over-reading of EEG is the most frequent cause of epilepsy misdiagnosis [10-13]. Approximately 30% of patients seen at epilepsy centres for drug-resistant seizures do not have epilepsy [13-16]. Epilepsy misdiagnosis is detrimental to the patients, as it unnecessarily exposes them to antiepileptic medications, prevents them from driving, and limits their career choices [17, 18]. Therefore, "conservative" reading of the EEG has been recommended [8] in order to achieve high specificity and avoid over-diagnosing epilepsy [11, 13, 19].

Sharp transients are often misinterpreted as IEDs when presenting with sharp morphology and higher amplitude on EEG during early drowsiness [10-12]. Several attempts have been made to define the characteristic morphology of IEDs [1, 10, 11, 19-26]. Recently, The International Federation of Clinical Neurophysiology (IFCN) has proposed an operational definition, consisting of six criteria, describing the morphological features of IEDs:

- di- or tri-phasic waves with sharp or spiky morphology (*i.e.*, pointed peak);
- different wave duration relative to ongoing background activity, either shorter or longer;
- asymmetry of the waveform- a sharply rising ascending phase and a more slowly decaying descending phase, or vice versa;
- the transient is followed by an associated slow after-wave;
- the background activity surrounding epileptiform discharges is disrupted by the presence of the epileptiform discharges;
- and 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 [27] (*figure 1*).

We have previously demonstrated that systematic implementation of the IFCN operational criteria yielded high specificity (over 95%) and reasonable sensitivity (over 80%) when at least five of the IFCN criteria were present (in any combination), or the specific combination of Criteria 1, 4 and 6 were present [28, 29]. However, in these studies, all raters were experienced EEG readers.

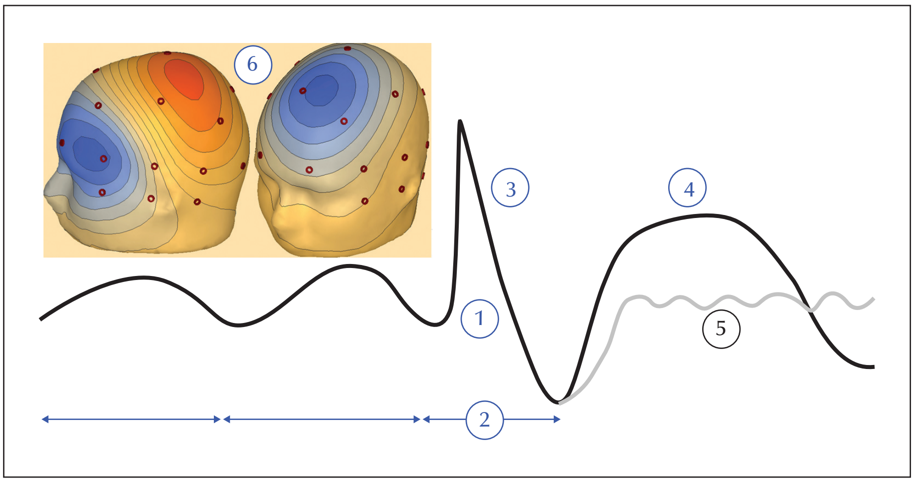

Figure 1. Infographic summarizing the six spike criteria of the International Federation of Clinical Neurophysiology: (1) di- or tri-phasic waves with sharp or spiky morphology (*i.e.*, pointed peak); (2) different wave duration relative to ongoing background activity, either shorter or longer; (3) asymmetry 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 epileptiform discharges is disrupted by the presence of the epileptiform discharges; (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 (modified with permission from: Kural *et al.*, 2020 [29]).

The goal of this study was to assess whether trainees can learn and apply, in practice, the operational definition of IED. We have evaluated the diagnostic accuracy of the trainees before and after a teaching module consisting of one webinar and two hands-on tutorials.

Methods

EEG recordings

The trainees reviewed a test dataset of 70 EEG samples. Each sample had a sharp transient -either an IED or non-epileptiform sharp transient- with a pointed peak and negative phase-reversal. The dataset was balanced: 35 consecutive patients had epilepsy, and 35 consecutive patients had non-epileptic paroxysmal events. The diagnostic gold standard was derived from the long-term video-EEG monitoring showing the patients habitual clinical episodes. We excluded patients with diagnostically inconclusive long-term video-EEG recordings, patients with both epileptic seizures and non-epileptic paroxysmal events, and patients younger than one year.

From the included patients, interictal EEG samples of 10 seconds, showing the first sharp transient, with pointed peak and negative phase-reversal (epileptic or not), were selected by two of the authors (MAK and SB). For patients with epilepsy, the sharp transients had to fulfil one additional criterion: they had to be congruent with the ictal EEG pattern, *i.e.*, in the same area as the seizure onset for patients with focal epilepsy and bilateral-synchronous for patients with generalized epilepsy. Hence, the diagnostic reference standard was robust and derived from an independent feature: the habitual clinical episode of the patient, which is considered the diagnostic gold standard [8]. All EEGs were recorded using the IFCN electrode array of 25 electrodes [30], including the electrodes in the inferior temporal chains. The regional ethics committee reviewed this project, in which only de-identified patient data were collected. According to Danish regulations, this study did not require written informed consent from the patients since it involved a retrospective analysis of anonymized data. We report the study according to the standards for reporting diagnostic accuracy studies (STARD) guidelines [31].

Study design

Seven trainees (MV, STA, HCL, BÖ, AYY, AHW, and ISÖ) participated in the study as EEG raters. Three of them were neurology residents and four of them had recently completed their neurology residency. They were educated in EEG as a part of their neurophysiology rotation, with at least three months but not more than six months of experience / previous training in reading EEGs. Five of them were working in Turkey and two of them in Denmark. The trainees were not familiar with the operational definition of IEDs before this study. The trainees independently reviewed the test EEG dataset, blinded to all other data. They were not informed that the dataset was balanced. The trainees were allowed to change montages, filters, gain, time resolution, and generate voltage maps.

In the first reading session, the trainees reviewed the EEG samples and concluded whether the sharp transients were epileptiform or not, solely based on their previous training and experience. The trainees were not given feedback after the first session, and they remained blinded to the gold standard. Next, the trainees participated in a teaching module about the operational definition and the six IFCN criteria of IEDs (*figure 1*). The teaching module consisted of a 50-minute webinar (*supplementary video*) with theoretical background and two hands-on sessions, during which they read 25 training samples, together with one of the authors (MAK), and applied the IFCN criteria to these samples. The EEGs in the training dataset were extracted from different patients to those for the test dataset.

To compare the performance of the trainees before and after the teaching module, we had to use the same EEG samples. To avoid potential bias (due to repetition), we presented the samples in a different, randomized order, the trainees did not receive feedback on their choices in the first round, they remained blinded to the gold standard, and the two reading sessions were separated by an interval of three months. The trainees did not receive any additional education between the sessions (only the teaching module described above). In the second reading session, the trainees systematically applied the six IFCN criteria, and specified for each sample, which criteria were present. Then, based on previous studies [28, 29], a sharp transient was considered an IED if it fulfilled at least five criteria (any combination) or the specific combination of the three optimal criteria (1, 4 and 6). Otherwise, it was labelled as a non-epileptiform sharp transient.

Outcome measures and statistics

For each EEG sample, we compared the scoring of the trainees (IED or non-epileptiform sharp transient) with the diagnostic gold standard, as described above. Then, we calculated the sensitivity (rate of true positives), specificity (rate of true negatives), and overall diagnostic accuracy (rate of true positives and true negatives) for each trainee before and after the teaching module, and we calculated the median performance of the group of trainees, before and after the teaching module. To compare sensitivity, specificity, and accuracy between the two reading sessions, we used the Wilcoxon rank-sum test. To calculate inter-rater agreement (IRA) we used Cohen's Kappa (k).

Results

EEG samples from 70 consecutive patients (43 female; age: 2-80 years, median: 34) who met the inclusion criteria were analysed in the test dataset. Thirty-five patients had epilepsy (31 focal and four generalized). Fifteen patients had spikes, 13 patients had sharp-waves and seven patients had polyspikes. Thirty-five patients had non-epileptic paroxysmal events (18 patients had psychogenic non-epileptic seizures, seven patients had sleep disorders, five patients had paroxysmal movement disorders, and five patients had syncope).

After the teaching module, applying the IFCN criteria for IEDs significantly increased the specificity and accuracy for the group of trainees (*table 1*). There was a trend for increased sensitivity, but this failed to reach statistical significance (*table 1*). Specificity and accuracy improved for all trainees (*table 2*), and sensitivity also improved for five of the seven trainees (*table 2*). Both positive predictive value (PPV) and negative predictive value (NPV) improved following the teaching module (from 71.88% to 90.62% and from 68.42% to 84.21%, respectively). The IRA increased significantly from fair (k=0.31; 95% CI: 0.22-0.40) to high-moderate (k=0.56; 95% CI:0.46-0.67) beyond-chance agreement.

Supplementary figures show examples of IEDs and non-epileptiform sharp transients from the dataset.

Discussion

We found that trainees can learn the IFCN criteria for IEDs using a teaching module consisting of a webinar and two supervised reading sessions of 25 EEG training samples. Implementing the IFCN criteria led to a significant increase in the specificity, overall accuracy and IRA of the trainees. Before the teaching module, the specificity of the trainees (77.14%) was lower than that previously reported for EEG experts (85.71-93.48%) [28, 29]. However, when the trainees implemented the IFCN criteria, they reached a level of specificity similar to that of experts (94.29%). It is important to improve the specificity of trainees in recognizing IEDs because EEG over-reading is the most common cause of epilepsy misdiagnosis [10-12] with significant detrimental consequences to the patients [17, 18]. There is a broad consensus that EEG over-reading (low specificity) is potentially more harmful than under-reading (low sensitivity). Implementing the IFCN criteria significantly improved IRA among the trainees (from fair to high-moderate beyond-chance agreement).

The trainees successfully implemented the six criteria for the IFCN operational definition of IEDs. For each sharp transient (epileptic or not), in each EEG sample, they noted the presence or absence of each criterion. The threshold based on previous studies was then applied (at least five criteria in any combination, or the specific combination of Criteria 1, 4 and 6) to define a sharp transient as an IED. Previously, this approach has been demonstrated to yield high specificity and sensitivity in the hands of EEG experts [28, 29], and we now demonstrate that trainees can learn and apply this, achieving high specificity and significantly improving diagnostic accuracy.

The accuracy of the trainees before the IFCN criteria teaching module (64.29%) was slightly higher than that previously reported after formal EEG training of neurology residents (44-50%) [32, 33]. Yet, after the teaching module of IFCN criteria, this increased further (81.43%), and the increase was statistically significant (p=0.017). Weber *et al*. examined the efficiency of a novel automated program, educating neurology residents in EEG evaluation. The performance was evaluated before and after the automated EEG teaching program. All residents increased their score, from a mean of 42.7% pre-test to 75.4% post-test (p<0.001) [34].

Selection of the sharp transients (both IEDs and non-epileptiform sharp transients) was based on the consensus of two experts (authors MAK and SB). However, whether they were epileptiform or not was decided based on an independent factor (the diagnostic gold standard). The same EEG dataset was independently reviewed by three other raters, who were EEG

▼ **Table 1.** Sensitivity, specificity and accuracy of the group of trainees identifying interictal epileptiform discharges, before and after the teaching module and implementation of the IFCN criteria. Median values (interquartile range in parenthesis).

	Before	After	*p*
Sensitivity	60.00% (48.57-62.86%)	77.14% (65.71-80.00%)	0.074
Specificity	77.14% (68.57-80.00%)	94.29% (85.71-94.29%)	0.017
Accuracy	64.29% (58.57-68.57%)	81.43% (77.14-87.14%)	0.017

▼ **Table 2.** Sensitivity, specificity and accuracy of the group of seven trainees, before and after the teaching module and implementation of the IFCN criteria.

	Sensitivity		Specificity		Accuracy	
	Before	After	Before	After	Before	After
Rater-1	60.00%	80.00%	77.14%	94.29%	68.57%	87.14%
Rater-2	77.14%	77.14%	70.00%	85.71%	57.14%	81.43%
Rater-3	48.57%	65.71%	80.00%	94.29%	64.29%	80.00%
Rater-4	60.00%	51.43%	68.57%	97.14%	64.29%	74.29%
Rater-5	48.57%	91.43%	68.57%	85.71%	58.57%	88.57%
Rater-6	62.86%	65.71%	77.14%	88.57%	70.00%	77.14%
Rater-7	42.86%	80.00%	82.86%	94.29%	62.86%	87.14%

experts; their agreement with the gold standard used in this study was 88.6%, with an agreement coefficient of 0.68, corresponding to a substantial beyond-chance agreement [29].

A possible limitation of our study is the relatively small sample size (balanced dataset including 70 patients). Epilepsy has many different types, yet IEDs are not specific to each epilepsy type. There are three categories of IEDs: spikes, sharp waves and polyspikes. Our test dataset included all these categories, and the age range was between 2 and 80 years. The question addressed in this study was not related to the sub-classification of IED types, but the distinction between epileptiform and non-epileptiform sharp transients. The sample size calculation showed that for the dichotomous classification, with an expected specificity of 95% and a significance level of 5% and power of 0.8, we required 68 patients in total [35].

Because we opted for an unequivocal diagnostic gold standard, derived from video-EEG recordings of the patients' habitual clinical episodes, all our patients had been referred for long-term monitoring. This might be a selection bias. However, indications for video-EEG monitoring include differential diagnosis and classification, *i.e.* the same indications as for "routine" recordings. Moreover, we were unable to find any published reports showing that IEDs of patients who are diagnostically more challenging are different from those of "easy cases". We argue that if this were true, it would be a positive bias, since one requires EEG for diagnosing cases that are not obvious based on other clinical data.

In conclusion, teaching the IFCN operational criteria for IEDs improves diagnostic accuracy of trainees, increases IRA and helps avoiding over-reading of EEG. This should be included in all training programs for neurology residents and clinical neurophysiology fellows. ■

Supplementary material.

Supplementary data accompanying the manuscript are available at www.epilepticdisorders.com.

Acknowledgements and disclosures.

We would like to thank Bo Martin Bibby (Department of Biostatistics, Aarhus University, Denmark) for help with the statistical analysis.

The authors have no conflict of interest to declare.

References

1. Pedley TA. Interictal epileptiform discharges: discriminating characteristics and clinical correlations. *Am J EEG Technol* 1980; 20: 101-19.

2. Fountain NB. Freeman JM EEG is an essential clinical tool: pro and con. *Epilepsia* 2006; 47(Suppl 1): 23-5.

3. Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, *et al*. Operational classification of seizure types by the International League Against Epilepsy: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia* 2017; 58(4): 522-30.

4. Van Donselaar CA, Schimsheimer RJ, Geerts AT, Declerck AC. Value of the electroencephalogram in adult patients with untreated idiopathic first seizures. *Arch Neurol* 1992; 49: 231-7.

5. Pillai J, Sperling MR. Interictal EEG and the diagnosis of epilepsy. *Epilepsia* 2006; 47(Suppl 1): 14-22.

6. Fürbass F, Kural MA, Gritsch G, Hartmann M, Kluge T, Beniczky S. An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: validation against the diagnostic gold standard. *Clin Neurophysiol* 2020; 131: 1174-9.

7. Da Silva Lourenço C, Tjepkema-Cloostermans MC, Van Putten MJAM. Machine learning for detection of interictal epileptiform discharges. *Clin Neurophysiol* 2021; 132(7): 1433-43.

8. Tatum WO, Rubboli G, Kaplan PW, Mirsatari SM, Radhakrishnan K, Gloss D, *et al*. Clinical utility of EEG in diagnosing and monitoring epilepsy in adults. *Clin Neurophysiol* 2018; 129: 1056-82.

9. Nascimento FA, Gavvala JR. Education research: neurology resident EEG education: a survey of US neurology residency program directors. *Neurology* 2021 27; 96(17): 821-4.

10. Benbadis SR, Tatum WO. Overintepretation of EEGs and misdiagnosis of epilepsy. *J Clin Neurophysiol* 2003; 20: 42-4.

11. Benbadis SR. Errors in EEGs and the misdiagnosis of epilepsy: importance, causes, consequences, and proposed remedies. *Epilepsy Behav* 2007; 11: 257-62.

12. Benbadis SR, Lin K. Errors in EEG interpretation and misdiagnosis of epilepsy. Which EEG patterns are overread? *Eur Neurol* 2008; 59: 267-71.

13. Chadwick D. Diagnosis of epilepsy. *Lancet* 1990; 336: 291-5.

14. Uldall P, Alving J, Hansen LK, Kibaek M, Buchholt J. The misdiagnosis of epilepsy in children admitted to a tertiary epilepsy center with paroxysmal events. *Arch Dis Child* 2006; 91: 219-21.

15. Asano E, Pawlak C, Shah A, Shah J, Luat AF, Ahn-Ewing J, *et al*. The diagnostic value of initial video-EEG monitoring in children - review of 1000 cases. *Epilepsy Res* 2005; 66: 129-35.

16. McBride AE, Shih TT, Hirsch LJ. Video-EEG monitoring in the elderly: a review of 94 patients. *Epilepsia* 2002; 43: 165-9.

17. Ferrie CD. Preventing misdiagnosis of epilepsy. *Arch Dis Child* 2006; 91: 206-9.

18. Lafrance Jr WC, Benbadis SR. Avoiding the costs of unrecognized psychological non-epileptic seizures. *Neurology* 2006; 66(11): 1620-1.

19. Engel Jr J. A practical guide for routine EEG studies in epilepsy. *J Clin Neurophysiol* 1984; 1: 109-42.

20. Gloor P. The EEG in the differential diagnosis of epilepsy. In : Van Duyn H, Donker DNJ, Van Huffelen J-P, (eds). *Current concepts in clinical neurophysiology*. The Hague, NV: Drukker Trio.

21. Drury I, Beydoun A. Pitfalls of EEG interpretation in epilepsy. *Neurol Clin* 1993; 11: 857-81.

22. Tatum WO, Husain AM, Benbadis SR, Kaplan PW. Normal adult EEG and patterns of uncertain significance. *J Clin Neurophysiol* 2006; 23: 194-207.

23. Mizrahi EM. Avoiding the pitfalls of EEG interpretation in childhood epilepsy. *Epilepsia* 1996; 37(suppl 1): S41-51.

24. 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.

25. Brazier MA. Preliminary proposal for an EEG terminology by the Terminology Committee of the International Federation for Electroencephalography and Clinical Neurophysiology. *Electroencephalogr Clin Neurophysiol* 1961; 13: 646-50.

26. Maulsby RL. Some guidelines for assessment of spikes and sharp waves in EEG Tracings. *Am J EEG Technol* 1971; 11: 3-16.

27. Kane N, Acharya J, Beniczky S, Caboclo L, Finnigan S, Kaplan PW, *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: 170-85.

28. Kural MA, Duez L, Sejer Hansen V, Larsson PG, Rampp S, Schulz R, *et al*. Criteria for defining interictal epileptiform discharges in EEG: a clinical validation study. *Neurology* 2020; 94: 1-9.

29. Kural MA, Tankisi H, Duez L, Hansen VS, Udupi A, Wennberg R, *et al*. Optimized set of criteria for defining interictal epileptiform EEG discharges. *Clin Neurophysiol* 2020; 131: 2250-4.

30. Seeck M, Koessler L, Bast T, Leijten F, Michel C, Baumgartner C, *et al*. The standardized EEG electrode array of the IFCN. *Clin Neurophysiol* 2017; 128: 2070-7.

31. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *AJR Am J Roentgenol* 2003; 181(1): 51-5.

32. Nascimento FA, Maheshwari A, Chu J, Gavvala JR. EEG education in neurology residency: background knowledge and focal challenges. *Epileptic Disord* 2020; 22(6): 769-74.

33. Dericioglu N, Ozdemir P. The success rate of neurology residents in eeg interpretation after formal training. *Clin EEG Neurosci* 2018; 49(2): 136-40.

34. Weber D, McCarthy D, Pathmanathan J. An effective automated method for teaching EEG interpretation to neurology residents. *Seizure* 2016; 40: 10-2.

35. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005; 58: 859-62.

Legend for video sequence

Supplementary video

Teaching video recording of the webinar about identifying IEDs and the IFCN criteria.