Interrater agreement of classification of photoparoxysmal EEG response

Sándor Beniczky[1], Harald Aurlien[2], Silvana Franceschetti[3], Antonio Martins da Silva[4], Francesca Bisulli[5], Carla Bentes[6], Laura Canafoglia[3], Lorenzo Ferri[7], David Krýsl[8], Ana Rita Peralta[6], Attila Rácz[9], J Helen Cross[10*], Alexis Arzimanoglou[11*]

1. Department of Clinical Neurophysiology, Danish Epilepsy Center, Dianalund and Aarhus University Hospital; Department of Clinical Medicine, Aarhus University.

2. Section for Clinical Neurophysiology, Department of Neurology, Haukeland University Hospital, Bergen, Norway.

3. Neurophysiology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy.

4. Neurophysiology Section, Department of Neurosciences, Hospital Santo António, CHU Porto and UMIB/ICBAS - University of Porto, Portugal.

5. IRCCS Bologna Institute of Neurological Sciences, Bologna, Italy; Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy.

6. Neurology Department, Hospital de Santa Maria- CHULN, Lisbon, Portugal and Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal.

7. IRCCS Bologna Institute of Neurological Sciences, Bologna, Italy; Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy.

8. Department of Clinical Neurophysiology, Sahlgrenska University Hospital; Institute of Neuroscience and Physiology at Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden.

9. Department of Epileptology, University Hospital Bonn, Bonn, Germany.

10. UCL- Great Ormond Street NIHR BRC Institute of Child Health, Great Ormond Street Hospital for Children NHS Trust, London, UK.

11. Department of Paediatric Clinical Epileptology, Sleep Disorders and Functional Neurology, University Hospitals of Lyon (HCL), Lyon, France, Epilepsy Unit Hospital San Juan de Dios, Universitat de Barcelona, Barcelona, Spain.

* Contributed equally

Corresponding author: Professor Sándor Beniczky, Danish Epilepsy Centre, Visby Allé 5, 4293 Dianalund, Denmark; email: sbz@filadelfia.dk; Phone: +4526981536

Keywords: electroencephalography, epilepsy, epileptiform discharges, intermittent photic stimulation, IRA, SCORE

Number of text pages: 13; Number of words (summary): 152; Number of words (main text): 1332; number of references: 15; number of figures: 1; number of tables: 1; supplementary document: 1.

Summary

Our goal was to assess the inter-rater agreement (IRA) of photoparoxysmal response (PPR) using the classification proposed by a taskforce of the International League Against Epilepsy (ILAE), and a simplified classification system proposed by our group. In addition, we evaluated IRA of epileptiform discharges (EDs) and the diagnostic significance of the EEG abnormalities. We used EEG recordings from the European Reference Network (EpiCARE) and the standardized computer-based organized reporting of EEG (SCORE). Six raters independently scored EEG recordings from 30 patients. We calculated the agreement coefficient (AC) for each feature. IRA of PPR using the classification proposed by the ILAE taskforce was only fair (AC=0.38). This improved to a moderate agreement by using the simplified classification (AC=0.56; p=0.004). IRA of EDs was almost perfect (AC=0.98) and IRA of scoring the diagnostic significance was moderate (AC=0.51). Our results suggest that the simplified classification of the PPR is suitable for implementation in clinical practice.

Keywords: electroencephalography, epilepsy, epileptiform discharges, intermittent photic stimulation, IRA, SCORE

Key point box:

- We used the SCORE system to investigate inter-rater agreement of EEGs from the European Reference Network (EpiCARE).
- Inter-rate agreement of assessing the photoparoxysmal response using the classification of the ILAE taskforce was only fair.
- This improved to moderate agreement using the simplified classification proposed in this study.
- Inter-rate agreement of assessing Epileptiform Discharges was almost perfect.
- Inter-rate agreement of assessing diagnostic significance was moderate.

Introduction

Intermittent photic stimulation (IPS) is one of the best known activation procedure in standard clinical EEG recordings[1]. In photosensitive patients, photoparoxysmal response (PPR) is observed in EEG: IPS triggers epileptiform discharges (EDs) in patients with otherwise normal recordings, or accentuates the incidence of EDs in patients with sporadic EDs[1]. About 5% of patients with epilepsy (range: 0.6 - 5.5%) have PPR[1]. It is age-dependent, with highest incidence in late childhood and early adolescence, and is more common in females[1-3]. PPR is most often seen in patients with idiopathic / genetic generalized epilepsy (GGE), Developmental and Epileptic Encephalopathies (DEE) and rarely in patients with focal epilepsy, with the exception of idiopathic occipital lobe epilepsy[1-4]. PPR may exist as a genetic trait in seizure-free siblings of patients with genetic epilepsies[5,6]. At the age of maximum penetrance (between 5 and 15 years)[3], PPR was recorded in 50% of children of patients with photosensitive epilepsy[6]. PPR provides important information for diagnosing epilepsy syndromes[1-4]. PPR can elicit EDs in patients with otherwise normal EEG, contributing to diagnosis. Furthermore, since PPR is often recorded in patients with GGE, especially Juvenile Myoclonic Epilepsy and GGE with generalized tonic-clonic seizures alone, and it is almost invariably found in untreated patients with Jeavons syndrome, the presence of PPR in the clinical context can help confirming the diagnosis of these syndromes[1,2].

PPR has been classified according to its location and extent[7,8]. A taskforce of the International League Against Epilepsy (ILAE) proposed a 4-grade classification of the PPR[9] and this has been implemented into the Standardized Computer-based Organized Reporting (SCORE) system[10,11], endorsed by the International Federation of Clinical Neurophysiology (IFCN). In spite of attempts at standardization, the interpretation and classification of PPR remains largely subjective, which questions its clinical utility. However, the inter-rater agreement of PPR has never been systematically investigated before.

A working group of the European Reference Network, EpiCARE, has collected EEG samples with PPR, using the SCORE database. We have investigated the inter-rater agreement on EDs, PPR and their diagnostic significance, using the standardized terminology of the SCORE system.

Methods

EEGs were recorded using the standardized IFCN electrode array[12]. These routine EEGs were recorded as part of the diagnostic workup of the patients, at the Danish Epilepsy Center (Dianalund, Denmark) and at the Carlo Besta Institute (Milan, Italy). Besides the wake and sleep epochs, all recordings contained IPS, as described in detail elsewhere[5].

De-identified EEG recordings were uploaded to the SCORE-EpiCARE database. Recordings with PPR were extracted for further analysis. In addition, we included recordings with normal (physiologic) photic drive response (as distractors). Six raters, with experience in the clinical interpretation of the EEG recordings in patients with epilepsy, independently reviewed the recordings, blinded to all clinical data. The only information provided to the raters (besides the EEG) were the age and the gender of the patients.

Raters reviewed the EEGs using a web-based tool, consisting of a digital EEG reader (NicoletOne) linked with the SCORE EEG system[10,11]. Raters were allowed to switch freely between montages (longitudinal and transversal bipolar, common average) and to adjust gain, digital filter settings and temporal resolution. The timing of the IPS was displayed as an event-bar (indicating the frequency of the stimulation) and in the channel at the bottom of the screen, showing the timing of each flash (Figure 1). Raters inserted annotations into the recordings, where they observed abnormalities. These annotations were automatically linked to the SCORE EEG system and were subsequently characterized ("scored") by the raters, using the standardized terminology[10,11] (Figure 1). Each rater,

for each annotated abnormality scored the following features: EDs, type of PPR and (for the whole recording) its diagnostic significance.

PPR were scored first according to the classification proposed by the ILAE taskforce[9]. The following choices were available in SCORE:

- Unmodified
- Posterior stimulus-dependent response
- Posterior stimulus-independent response, limited to the stimulus-train
- Posterior stimulus-independent response, self-sustained
- Generalized photoparoxysmal response, limited to the stimulus-train
- Generalized photoparoxysmal response, self-sustained
- Activation of pre-existing epileptogenic area

Subsequently, each rater scored the same recordings, using a simplified classification:

- No PPR
- Posterior PPR (abnormality present only during IPS)
- Generalized PPR (abnormality present only during IPS)
- Accentuation during IPS of spontaneous epileptiform activity. Accentuation was defined as incidence of EDs or duration of EDs at least twice as much / twice as long as unprovoked (spontaneous).

The diagnostic significance of the EEG recording was scored using the standardized categories available in SCORE[10,11]: Normal; No definite abnormality; Epilepsy not further specified; Focal epilepsy; Multifocal epilepsy; Generalized epilepsy; Status epilepticus; CSWS / ESES; Psychogenic non-epileptic seizures (PNES); Other non-epileptic clinical episode; Focal dysfunction of the

central nervous system; Diffuse dysfunction of the central nervous system; Coma; Brain death; EEG abnormality of uncertain clinical significance; Combined generalized and focal epilepsy.

All choices of all raters were automatically saved in SCORE EEG (Figure 1) and then extracted for statistical analysis. We evaluated IRA of three features: EDs, PPR and diagnostic significance. IRA of PPR was evaluated both for the original and for the simplified classification. We calculated IRA using Gwet's agreement coefficient AC1 to avoid the "paradoxes of kappa"[13]. Inter-rater agreement 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)[14]. We compared IRAs of the two PPR classifications using bootstrap method[15]. The analyses were performed using Stata version 15.1.

Results

EEG recordings from 30 patients (20 female) were analyzed by the six raters. The median age of the patients was 14.5 years (range 1-60 years). Eighteen patients had GGE (ten patients had Juvenile Myoclonic Epilepsy, three patients had Juvenile Absence Epilepsy, three patients had Childhood Absence Epilepsy, two patients had Jeavons syndrome).

Eight patients had DEE (mitochondrial encephalopathy, Rett syndrome, Neuronal Ceroid Lipofuscinosis, Lafora disease, Dravet syndrome) and one patient had Neurofibromatosis type-1. Recordings of normal photic drive responses were included from a patient with Mesial Temporal Sclerosis and from two healthy control subjects.

Although the IRA of EDs was almost perfect, IRA of PPR using the classification proposed by the ILAE taskforce[9] was only fair (Table 1). Using the simplified PPR-classification, resulted in improvement of the agreement ($p<0.004$), and IRA became moderate (Table 1). IRA was almost

perfect on presence / absence of any form of PPR (AC1=0.88; 95% CI: 0.71 - 1.00). IRA on diagnostic significance was moderate (Table 1).

Discussion

We found that IRA of PPR, using the classification proposed by the ILAE taskforce[9] was only fair. This limits its clinical utility. Although theoretically the detailed classification provides more precise characterization of the EEG changes, the low IRA questions its feasibility in clinical practice. The simplified version had a better (moderate) IRA, which corresponds to what generally has been reported on IRA in EEG. The simplified classification of the PPR distinguishes conditions when the abnormalities are recorded only during IPS from accentuation during IPS of abnormalities that occur spontaneously. In addition, it distinguishes between PPR limited to the posterior regions from generalized PPR. Although accentuation has been more precisely defined (increase by a factor of 2 during IPS compared to spontaneous occurrence) this remained a significant source of disagreement among raters: when this category was eliminated, IRA became almost perfect.

Fifteen patients had generalized PPR: nine limited to the stimulus train and six self-sustained. Discordance in scoring occurred when the EDs exceeded the last stimulus with a short period (one or two spike-waves after the last stimulus). Merging the two categories contributed to the increased IRA.

The IRA on EDs was unusually high in our study. Since we included patients with PPR, this resulted in a high incidence of patients with generalized EDs, which might be a more obvious (unequivocal) EEG abnormality compared to focal EDs in previous studies reporting moderate IRA. Scoring of the diagnostic significance was moderate, which is not surprising, given the wide variety of possible clinical interpretations of the EEG abnormalities.

Our findings suggest that the simplified version of the PPR classification, proposed in this study has an acceptable IRA, making it suitable for implementation in clinical practice.

Acknowledgments

We would like to express our gratitude to the members of the Clinical Neurophysiology work-package of the ERN-EpiCARE, especially Paolo Tinuper, Dorota Domańska-Pakieła and Magdalena Kaczorowska-Frontczak. We thank Bo M. Bibby (Department of Biostatistics, Aarhus University, Denmark) for the assistance with the statistical calculations.

Disclosure of Conflicts of Interest

Harald Aurlien is CMO and minority share holder in Holberg EEG, the company behind the SCORE EEG software. The rest of the authors do not have any conflict of interest to disclose, related to this paper.

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.

References

1. 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. doi: 10.1016/j.clinph.2018.01.019.
2. Kasteleijn-Nolst Trenite DG. Photosensitivity in epilepsy. Electrophysiological and clinical correlates. Acta Neurol Scand Suppl 1989;125:3–149.
3. Aurlien H, Gjerde IO, Eide GE, et al. Characteristics of generalised epileptiform activity. Clin Neurophysiol 2009;120:3-10.
4. Guerrini R, Dravet C, Genton P, et al. Idiopathic photosensitive occipital lobe epilepsy. Epilepsia 1995;36:883-91.
5. Kasteleijn-Nolst Trenite D, Rubboli G, Hirsch E, et al. Methodology of photic stimulation revisited: updated European algorithm for visual stimulation in the EEG laboratory. Epilepsia 2012;53:16–24.
6. Waltz S, Stephani U. Inheritance of photosensitivity. Neuropediatrics 2000;31:82–5.
7. Waltz S, Christen HJ, Doose H. The different patterns of the photoparoxysmal response--a genetic study. Electroencephalogr Clin Neurophysiol 1992;83:138-45.
8. Trenité DG, Binnie CD, Harding GF, et al. Medical technology assessment photic stimulation--standardization of screening methods. Neurophysiol Clin 1999;29:318-24.
9. Kasteleijn-Nolst Trenité DG, Guerrini R, Binnie CD, Genton P. Visual sensitivity and epilepsy: a proposed terminology and classification for clinical and EEG phenomenology. Epilepsia 2001;42:692-701.
10. Beniczky S, Aurlien H, Brøgger JC, et al. Standardized computer-based organized reporting of EEG: SCORE. Epilepsia 2013;54:1112-24. doi: 10.1111/epi.12135.

11. Beniczky S, Aurlien H, Brøgger JC, et al. Standardized computer-based organized reporting of EEG: SCORE - Second version. Clin Neurophysiol 2017;128:2334-2346. doi: 10.1016/j.clinph.2017.07.418.
12. Seeck M, Koessler L, Bast T, et al. The standardized EEG electrode array of the IFCN. Clin Neurophysiol 2017;128:2070-2077. doi: 10.1016/j.clinph.2017.06.254.
13. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol 2008; 61: 29–48.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–74.
15. Vanbelle S, Albert A. A bootstrap method for comparing correlated kappa coefficients. J Stat Comput Simul 2008;78:1009–1015.

Table 1. Percent agreement and agreement coefficient for assessing the EEG features in the recordings. (CI: confidence interval).

Feature	Percent agreement (95% CI)	Agreement coefficient (95% CI)
Epileptiform Discharges	0.98 (0.94 - 1.00)	0.98 (0.94 - 1.00)
PPR (ILAE taskforce)	0.47 (0.40 - 0.54)	0.38 (0.31 - 0.47)
PPR (simplified)	0.66 (0.56 - 0.76)	0.56 (0.43 - 0.69)
Diagnostic significance	0.56 (0.46 - 0.66)	0.51 (0.39 - 0.63)

Figure 1. Feature extraction and documentation using the SCORE EEG system.

EEG showing PPR. Marking an abnormality in the EEG (red label "Graphoelement" on the top of the screen), automatically inserts an entry to be characterized in SCORE EEG (blue window). EEG features (including classification of PPR shown here) are scored by clicking on the pre-defined lists in the software. All scored features are automatically saved in the database. The content in the text box "Finding summary" is generated automatically, as the user scores the observed features.