Published in final edited form as:

J Clin Neurophysiol. 2017 March ; 34(2): 168–173. doi:10.1097/WNP.0000000000000344.

Characteristics of EEG Interpreters Associated With Higher Interrater Agreement

Jonathan J. Halford^{*}, Amir Arain[†], Giridhar P. Kalamangalam[‡], Suzette M. LaRoche[§], Bonilha Leonardo^{*}, Maysaa Basha^{||}, Nabil J. Azar[¶], Ekrem Kutluay^{*}, Gabriel U. Martz[#], Wolf J. Bethany^{**}, Chad G. Waters^{††}, and Brian C. Dean^{††}

^{*}Department of Neurology, Medical University of South Carolina, Charleston, South Carolina, U.S.A.

[†]Department of Neurology, Vanderbilt University, Nashville, Tennessee, U.S.A.

[‡]Department of Neurology, University of Texas Health Science Center of Houston, Houston, Texas, U.S.A.

[§]Mission Health, Asheville, North Carolina, U.S.A.

^{||}Department of Neurology, Wayne State University and Detroit Medical Center, Detroit, Michigan, U.S.A.

[¶]Academic Health System, Hamad Medical Corporation, Doha, Qatar

[#]Norton Neurology Services, Louisville, Kentucky, U.S.A.

^{**}Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, U.S.A.

^{††}School of Computing, Clemson University, Clemson, South Carolina, U.S.A.

Abstract

Purpose—The goal of the project is to determine characteristics of academic neurophysiologist EEG interpreters (EEGers), which predict good interrater agreement (IRA) and to determine the number of EEGers needed to develop an ideal standardized testing and training data set for epileptiform transient (ET) detection algorithms.

Methods—A three-phase scoring method was used. In phase 1, 19 EEGers marked the location of ETs in two hundred 30-second segments of EEG from 200 different patients. In phase 2, EEG events marked by at least 2 EEGers were annotated by 18 EEGers on a 5-point scale to indicate whether they were ETs. In phase 3, a third opinion was obtained from EEGers on any inconsistencies between phase 1 and phase 2 scoring.

Results—The IRA for the 18 EEGers was only fair. A select group of the EEGers had good IRA and the other EEGers had low IRA. Board certification by the American Board of Clinical

Address correspondence and reprint requests to Jonathan J. Halford, MD, 96 Jonathan Lucas St., Suite 307 CSB, MSC606, Charleston, SC 29425, U.S.A.; halfordj@musc.edu.

The authors have no conflicts of interest to disclose.

Presented as a poster at the American Epilepsy Society Annual Meeting, Philadelphia, PA, December 6, 2015.

Neurophysiology was associated with better IRA performance but other board certifications, years of fellowship training, and years of practice were not. As the number of EEGers used for scoring is increased, the amount of change in the consensus opinion decreases steadily and is quite low as the group size approaches 10.

Conclusions—The IRA among EEGers varies considerably. The EEGers must be tested before use as scorers for ET annotation research projects. The American Board of Clinical Neurophysiology certification is associated with improved performance. The optimal size for a group of experts scoring ETs in EEG is probably in the 6 to 10 range.

Keywords

EEG; Electroencephalography; Spike detection; Epileptiform transient; Interrater agreement

The routine scalp electroencephalogram (rEEG) is the most common clinical neurophysiology procedure. It consists of a 20- to 30-minute recording from approximately 20 scalp electrodes. Approximately 350,000 outpatient rEEGs were performed in the United States for Medicare alone in 2013, according to the Centers for Medicare and Medicaid Services, and more than 1 million were probably performed in total.

The EEG of a patient with epilepsy is characterized by occasional epileptiform transients (ETs) consisting of spikes of 20 to 70 milliseconds and sharp waves of 70 to 200 milliseconds in duration. Detecting ETs in EEG is important because their presence is predictive of seizure recurrence in patients following a first seizure,^{1,2} and is useful in supporting the diagnosis of epilepsy.³ However, because of the wide variety of morphologies of ETs and their similarity to waves that are part of the normal background activity and to artifacts (extracerebral potentials from muscle, eyes, heart, electrodes, etc.), the detection of ETs is far from straightforward.⁴

It is well known that rEEGs are frequently misinterpreted by neurologists without neurophysiology fellowship training (who are the majority of neurologists interpreting rEEGs).⁵ Misinterpretation of the rEEG can adversely affect patients, leading to the misdiagnosis of epilepsy and the inappropriate use of antiepileptic medications, as well as delay in the treatment of the true underlying cause of the seizure-like events, which can be cardiac arrhythmias, psychogenic events, or other types.⁵ Patients with nonepileptic events endure an average of 7 years of antiepileptic medication exposure before they are correctly diagnosed.⁶

Despite the problem of rEEG misinterpretation, there are no commonly-accepted ET detection software programs or devices available. Few studies have measured the performance of automated ET detection software and almost none have compared the performance between software programs.⁴ Based on the few studies that have been done and based on the commonly held opinion of academic EEG interpreters, current ET detection software has an insufficient sensitivity, requiring a neurologist to look over each EEG to make sure that ETs are not missed.⁷ Current software also has an insufficient specificity, requiring a neurologist to look through a large list of detections, many of which are false positive.^{4,8,9}

Our eventual goal is to create a large database for testing and training automated algorithms for ET detection in scalp EEG. To create this database, opinions from multiple experts across a large number of EEG recordings will need to be collected. There are two important questions about how to choose a group of expert scorers for such a database. The first question is whether there is optimal number of experts to look at each recording. The second question is whether there are scorer characteristics (such as board certification and years of experience), which are helpful in choosing a group of scorers with high interrater agreement (IRA). This study acquired opinion from a large group of academic neurophysiologists using a large number of EEG recordings to attempt to answer these questions.

MATERIALS AND METHODS

EEG Data Set

A library of rEEG segments was created by retrospective review of approximately 1,000 rEEGs performed in the Medical University of South Carolina Neurophysiology Laboratory for clinical purposes during the years 2010 to 2011. Two hundred 30-second rEEG segments from 200 different rEEG studies performed on 200 different patients were selected by 2 experts (J.J.H. and G.U.M.). These two experts were both board certified by the American Board of Clinical Neurophysiology (ABCN). Fifty of the segments were chosen from within EEGs, which were randomly selected from within the group of all EEGs that were interpreted as containing ETs in the clinical report. Fifty of the segments were from randomly selected EEGs read as normal in the clinical report. Fifty segments were selected because they contained ETs from patients with known epilepsy with difficult-to-interpret ETs and 50 were selected because they contained benign paroxysmal activity (exaggerated alpha activity, wicket spikes, and small sharp spikes), which could easily be misinterpreted by an inexperienced reviewer as being epileptiform. The two investigators who selected the EEG data for this study were not included as EEG scorers. The rEEG studies were recorded referentially (with the digital reference electrode placed between Fz and Cz) using digital XLTEK EEG recording equipment with a sampling frequency of 256 Hz with the standard 10–20 electrode placement. Institutional Review Board approval was granted from the Medical University of South Carolina for collection of retrospective deidentified clinical rEEG data and for sharing this rEEG data with researchers at other universities. The institutional review board approval was also granted at Clemson University. The rEEG data were exported from the Medical University of South Carolina clinical XLTEK server as European Data Format¹⁰ files. All segments were randomly ordered and concatenated into one long 100-minute recording.

EEGnet EEG Annotation System

The EEGnet is a software system that allows users to view and annotate EEG data using a web browser with no local software installation required.¹¹ EEGnet software leverages recent improvements in web standards (e.g., HTML5), to run smoothly in most modern web browsers. The web-based “front end” of the system, written in JavaScript, interacts with a secure backend database (MySQL running on CentOS Linux) hosted at the Clemson University School of Computing. Any number of users can simultaneously access the system from any location on the web to view and annotate the same EEG data sets through this

interface. The software displays 10 seconds of rEEG data at a time with channel labels. The ECG channel is also displayed. Channel gain for all the channels can be adjusted together or individually, and the user can scroll forward and backward through the EEG data 1 second or 10 seconds at a time. Scoring is performed using any one of a list of conventional EEG montages including AP bipolar, transverse bipolar, hatband bipolar, average reference, Cz reference, and ipsilateral ear reference.

EEG Scoring Paradigm

Invitations to take part in the study were sent to the members of the Critical Care EEG Monitoring Research Consortium. A total of 19 academic clinical neurophysiologists (EEGers) participated in the project. Before they began annotating EEG recordings, each the EEGnet system prompted each EEGer to answer a series of questions regarding board certification, years of neurophysiology fellowship training, and years in practice. The EEGers were asked if they possessed board certification by the ABCN, the ABPN (American Board of Psychiatry and Neurology) Neurophysiology subspecialty, and ABPN Epilepsy subspecialty.

This study used a novel three-phase method for EEG scoring. Three to 4 months elapsed between each phase of the scoring. During the first phase of the EEG annotation, each of the 19 EEGers was instructed to label all ETs in the recording on the channel on which they thought it was best represented. An algorithm used in previous studies¹¹ was used to eliminate the redundancy within this database of labeled EEG segments by clustering highly correlated segments overlapping in time. Two marked EEG segments were merged into the same cluster if they overlapped temporally by at least 50% of the length of the shortest segment. The EEG segment with the maximum sum of correlations to the others in its cluster was chosen as the “representative” segment for the cluster, and these representative EEG segments were added to a “master list” of events to be scored in the second phase. The master list included a “representative” segment from each cluster, which included at least two members (events that were marked by at least two EEGers.)

Of the 19 EEGers who participated in phase 1 of the study, 18 EEGers participated in the second phase of EEG annotation. (One EEGer did not complete phases 2 and 3 of the scoring task, despite multiple reminder e-mails.) In this phase, each EEGer was presented with the same EEG recording with the master list of 235 events labeled within it, presented as yellow boxes superimposed over single-channel EEG segments. The EEGers were instructed to left mouse-click on each highlighted segment and to label it on a 5-point Likert scale (using keyboard keys 1–5) as either (1) “definitely not an epileptiform discharge,” (2) “not likely an epileptiform discharge,” (3) “not sure, could go either way,” (4) “likely an epileptiform discharge,” and (5) “definitely an epileptiform discharge.”

In multiple cases, EEGers were inconsistent in their marking of ETs between phase 1 and phase 2 scoring. For example, in some cases, an EEGer marked an event as an ET in phase 1 scoring, yet labeled the same event with a score of 1 to 3 in phase 2 scoring, indicating that they did not think that the event was an ET. In the third phase of scoring, each EEGer was presented with a unique list of events for which that scorer was inconsistent between phase 1 and phase 2 scoring. The EEGer was then asked to rescore this list of events in EEGnet

using the same method as in phase 2 scoring. This provided a third “tie break” opinion on these events. In cases that the EEGer did not mark the event as an ET in phase 1 scoring but labeled the event with a 4 or 5 in phase 2 scoring, if they labeled the event with a 4 or 5 in phase 3 scoring, the phase 2 scoring value was used in the analysis. But if the EEGer labeled it with a 1 to 3, the phase 3 scoring value was used in the analysis. In cases that the EEGer marked an event as an ET in phase 1 scoring but labeled it with a 1 to 3 in phase 2 scoring, if the EEGer labeled the event with a 1 to 3 in phase 3 scoring, the phase 2 score was used for analysis. In cases that the EEGer marked the event with a 4 or 5, the phase 3 scoring value was used in the analysis.

Statistical Analysis

All data processing and statistical analysis was performed offline using a commercial software package (MATLAB 8.6 with the Statistical Analysis Toolbox; The Math Works Inc, Natick, MA, 2000). Nonnormal distribution of all variables was assumed and simple statistical comparisons between groups of continuous variables were performed using the nonparametric Wilcoxon rank sum test with Bonferroni–Holm correction for multiple comparisons. Correlations were performed with the Spearman correlation test. Two IRA statistical measures were used to analyze the data: (1) Fleiss kappa coefficient,¹² and (2) linearly weighted Gwet agreement coefficient (AC2).¹³ The Gwet coefficient is considered superior to Fleiss kappa for IRA analysis because it provides adjustment for both chance agreement and misclassification errors. Analysis was performed using all the data in the 5-point Likert scale and also using binary data by considering events labeled as 1 to 3 on the 5-point scale as “non-ET” and events labeled as 4 to 5 as “ET” events. Because this multiphase method of ET scoring enables the definition of true negative detections, sensitivity and specificity were calculated for all EEGers based on consensus opinion of all EEGers. The IRA for all 262,124 possible groups of the 18 EEGers (of size 2–17) were evaluated and groups of each size with the best IRA based on the Fleiss kappa and AC2 statistics were identified.

RESULTS

The performance of the EEGers is reported as a group but how well each individual performed is kept confidential and is known only by Dr. Halford and staff at Clemson University. In the first phase, the 19 EEGers placed annotations to mark a total of 1,115 single-channel EEG segments as containing ETs. The mean number of ETs annotated by the EEGers was 58.7 (range, 16–195). A master list of 235 segments was created by clustering all the segments to remove redundancy and including only events, which were marked by at least two EEGers, as described above. Each EEGer spent approximately 1 to 2 hours to complete phase 1. In the second phase, 18 EEGers were asked to categorize the events in the master list on the 5-point scale described above. A total of 4,320 classification decisions were made by the 18 EEGers in phase 2 scoring. Each EEGer was presented between 30 and 104 events to be rescored in phase 3.

The IRA for both phase 1 and phase 2 to 3 scoring for all EEGers using the 5-point scale and binary scale is given in Table 1. Overall, agreement for the 18 EEGers was good for phase 1

scoring based on analysis with the AC2 coefficient. Agreement for phase 2 to 3 scoring was only fair. IRA for binary scoring was higher than for scoring on the 5-point scale, as would be predicted by chance agreement. When the IRA of all 262,124 possible groups of EEGers (of size 2–17) was analyzed, there were groups of size 2 to 7 with good IRA. See Fig. 1 for the performance of groups with best IRA of all possible group sizes. There does not appear to be an optimal group size based on analysis of IRA, because the performance of the best group for each group size decreases steadily as the group size increases. But if the consensus majority opinions for each group with the best IRA is evaluated, the percentage change in this consensus opinion as the group sizes increase from 2 to 18 drops steadily from group size 2 to 10 and then remains low (Fig. 2).

To assess if there were groups of EEGer who were not members of a group with best IRA who had an IRA among themselves, which was nearly as high as the group with the best IRA, the IRA of second best groups of various sizes is plotted against the best groups for group sizes 2 to 9 in Fig. 3. It does not appear that there are second best groups with estimated IRA, which are close to the best groups, arguing against the existence of multiple “islands” of independent opinion and demonstrating that certain EEGers in our cohort agree well with each other and the remainder of the EEGers do not agree as well with each other.

Characteristics of each EEGer who participated in phase 2 and phase 3 scoring, including board certification status, years of fellowship training, and years in practice are listed in Table 2. Because there is no gold standard “correct” identification for each event, the performance of each EEGer is given as (1) the mean pairwise AC2 IRA between that scorer and all other scorers for binary scoring and (2) the number of memberships for that scorer in the groups of size 2 to 17 with the best IRA. The only EEGer characteristic that was significantly associated with performance is ABCN board certification status ($\rho = 0.50$, $P < 0.04$). Mean AC2 for each EEGer versus all other EEGers was significantly higher at 0.39 for those with ABCN certification than 0.31 without it ($P < 0.02$). There were more memberships in groups with the best IRA for EEGers with ABCN certification (average 12.1 memberships) versus EEGers without ABCN certification (average 7.0 memberships) ($P < 0.04$). The sensitivity and specificity of each scorer as well as the ABCN board certification status are plotted in Fig. 4.

DISCUSSION

This study demonstrates that there is considerable variability in the performance of academic neurophysiologists in ET identification in EEG recordings. Overall, agreement among all 18 EEGers for phase 2 to 3 scoring was only fair, consistent with previous studies.^{4,11,14} A subset of the EEGers had moderate to good IRA (depending on the size of the subset) and the other EEGers did not agree well among themselves. This shows that experts board certified by the ABCN produced IRA that was significantly higher than experts without ABCN certification. It is not clear why EEGers with ABCN certification performed better. The ABCN board examination may attract applicants who are more interested in EEG interpretation or who feel more confident in their EEG interpretation skills. Or perhaps some of the EEGers in this study failed the ABCN examination, so the examination served to select EEGers who are more knowledgeable. Although this is the largest study of its type to

date, the number of EEGers studied is still limited, so this finding may be spurious and needs to be validated by further investigation.

In this study, the length of fellowship training and years of experience as well as board certification in either the Clinical Neurophysiology or Epilepsy subspecialty examination of the ABPN were not associated with good IRA performance. But the limited number of EEGers prevents us from excluding the possibility of weaker associations with performance, which might be detected if we had used more EEGers in this study. For similar projects in the future, we plan on using only ABCN-certified experts who perform well when we test them using this data set, because not all ABCN-certified experts in this study produced good IRA with other experts.

It is not known how many experts are needed to create an ideal gold standard testing and training data set for ET detection. Wilson et al.¹⁵ suggested that only five experts were needed, but no attempt was made to calculate this based on a large group of experts, since their study included only five EEGers. The data we present here suggests that more than 5 is needed but probably not more than 10. As shown in Fig. 4, the change in the consensus opinion from experts declines as experts are added and becomes quite small after the group of experts reaches a size of 7 to 8. The amount of change in consensus expert opinion begins to increase again as the size of the expert group gets to 16 to 18 because the experts who are added in the larger group sizes are experts with low IRA with other experts, causing significant shift of consensus opinion. Since increasing the size of the group of experts who score EEGs causes significant increase in study cost, the optimal size of the expert scoring group is probably around seven.

There are several weaknesses in this study. First, IRA between experts would probably have been higher if the EEG segments from each patient had been longer than 30 seconds. It is known that IRA is considerably higher if experts are interpreting 20-minute routine EEG recordings than for shorter EEG recordings.¹⁶ Second, intra-rater agreement was not assessed in this study, although it was assessed in a previous similar study of ours and found to be low to moderate.¹¹ Third, some of the EEG segments used for this study were selected by only two investigators, which could have led to selection bias. Steps were taken to mitigate this since some of the segments were selected randomly and investigators who selected the segments were not EEG scorers for this study. Fourth, participation in this study was the first time many of the EEGers had used the EEGnet system to view EEG and label ETs and so their performance may not have been as good as if they had more time to get used to viewing EEG using this software. Fifth, the authors acknowledge a potential conflict of interest since the study finds that EEGers who are board certified by the ABCN performed better than those that were not and many of the authors of this manuscript are board certified by the ABCN and involved in organizations closely affiliated with the ABCN such as the Critical Care EEG Monitoring Research Consortium, the American Clinical Neurophysiology Society and the *Journal of Clinical Neurophysiology*.

Our future plans are to leverage the insight provided by this pilot study to create a much larger standardized database for the training and testing of automated ET detection algorithms. We will use a group of 7 to 8 ABCN board-certified EEGers who generated

substantial IRA when scoring the pilot data presented in this report. Although it is not possible to predict the necessary size of such a standardized EEG database, a practical number is as many routine EEGs as a neurophysiology fellow interprets in a year (approximately 500).

In conclusion, this study of IRA for the detection of ETs in rEEG recordings includes the largest number of experts to date. Although IRA for the whole group of EEGs was only fair, a select group of the EEGers had good IRA. Board certification by ABCN was associated with better performance. The optimal size for a group of experts scoring ETs in EEG is shown to be in the 6 to 10 range.

Acknowledgments

The authors acknowledge the support of the Critical Care EEG Monitoring Research Consortium. This project is unfunded. Statistical analysis for this project was supported by the South Carolina Clinical & Translational Research Institute, Medical University of South Carolina's CTSA, NIH/NCRR Grant Number UL1RR029882. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or NCRR.

References

1. Seidel S, Pablik E, Aull-Watschinger S, Seidl B, Patariaia E. Incidental epileptiform discharges in patients of a tertiary centre. *Clin Neurophysiol.* 2016; 127:102–107. [PubMed: 25802204]
2. van Donselaar CA, Schimsheimer RJ, Geerts AT, Declerck AC. Value of the electroencephalogram in adult patients with untreated idiopathic first seizures. *Arch Neurol.* 1992; 49:231–237. [PubMed: 1536624]
3. Fountain NB, Freeman JM. EEG is an essential clinical tool: pro and con. *Epilepsia.* 2006; 47:23–25. [PubMed: 17044821]
4. Halford JJ. Computerized epileptiform transient detection in scalp EEG: obstacles to progress and the example of computerized ECG interpretation. *Clin Neurophysiol.* 2009; 120:1909–1915. [PubMed: 19836303]
5. Benbadis SR. Errors in EEGs and the misdiagnosis of epilepsy: importance, causes, consequences, and proposed remedies. *Epilepsy Behav.* 2007; 11:257–262. [PubMed: 17719853]
6. Reuber M, Fernandez G, Bauer J, Helmstaedter C, Elger CE. Diagnostic delay in psychogenic non-epileptic seizures. *Neurology.* 2002; 58:493–495. [PubMed: 11839862]
7. Ver Hoef L, Elgavish R, Knowlton RC. Effect of detection parameters on automated electroencephalography spike detection sensitivity and false-positive rate. *J Clin Neurophysiol.* 2010; 27:12–16. [PubMed: 20087204]
8. De Lucia M, Fritschy J, Dayan P, Holder DS. A novel method for automated classification of epileptiform activity in the human electroencephalogram based on independent component analysis. *Med Biol Eng Comput.* 2008; 46:263–272. [PubMed: 18071771]
9. Indiradevi KP, Elias E, Sathidevi PS, D NS, Radhakrishnan K. A multilevel wavelet approach for automatic detection of epileptic spikes in the electroencephalogram. *Comput Biol Med.* 2008; 38:805–816. [PubMed: 18550047]
10. Kemp B, Värri A, Rosa AC, Nielsen KD, Gade J. A simple format for exchange of digitized polygraphic recordings. *Electroencephalogr Clin Neurophysiol.* 1992; 82:391–393. [PubMed: 1374708]
11. Halford JJ, Schalkoff RJ, Zhou J, et al. Standardized database development for EEG epileptiform transient detection: EEGnet scoring system and machine learning analysis. *J Neurosci Methods.* 2013; 212:308–316. [PubMed: 23174094]
12. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* 1971; 76:378–382.

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. [PubMed: 18482474]
14. Halford JJ, Pressly WB, Benbadis SR, et al. Web-based collection of expert opinion on routine scalp EEG: software development and inter-rater reliability. *J Clin Neurophysiol.* 2011; 28:178–184. [PubMed: 21399515]
15. Wilson SB, Harner RN, Duffy FH, Tharp BR, Nuwer MR, Sperling MR. Spike detection. I. Correlation and reliability of human experts. *Electroencephalogr Clin Neurophysiol.* 1996; 98:186–198. [PubMed: 8631278]
16. Black MA, Jones RD, Carroll GJ, Dingle AA, Donaldson IM, Parkin PJ. Real-time detection of epileptiform activity in the EEG: a blinded clinical trial. *Clin Electroencephalogr.* 2000; 31:122–130. [PubMed: 10923198]

FIG. 1.
Interrater agreement (IRA) of the groups of EEG interpreters with best IRA for each group size.

FIG. 2.

Percentage change in consensus opinion as group size of best groups changes. For example, when group size changes from 2 to 3, the consensus opinion changes by approximately 7%.

FIG. 3.

Interrater agreement of best group and second best group for all group sizes.

FIG. 4.
Sensitivity and specificity of all EEG interpreters versus consensus opinion based on binary scoring.

TABLE 1

Interrater agreement for EEG interpreters

	Phase 1	Phase 2–3	
		5-Point Likert Scale	2-Point Binary Scale
Fleiss kappa	0.30	0.14	0.33
AC2	0.69	0.31	0.36

AC2, Gwet agreement coefficient.

TABLE 2

Characteristics of EEG Scorers

Scorer Number	Board Certification			Years of Fellowship Training	Years of Practice	Mean AC2 Correlation With All Other Scorers	Group Memberships	No. Best
	ABCN	ABPN Clinical Neurophysiology	ABPN Epilepsy					
1	1	0	1	2	17	0.37		6
2	1	1	0	2	14	0.47		17
3	1	0	0	2	15	0.44		16
4	0	1	0	1.2	8	0.23		3
5	0	1	0	2	11	0.36		7
6	0	1	0	1	4	0.39		15
7	1	0	0	1.5	8	0.42		17
8	1	1	0	1	10	0.36		13
9	0	1	0	2	0	0.34		7
10	0	0	1	1	2	0.32		9
11	1	1	0	2	12	0.28		5
12	0	0	0	2	4	0.24		1
13	0	0	1	2	3	0.32		11
14	0	1	0	1	0	0.23		2
15	1	1	0	1	29	0.37		13
16	1	0	1	1	2	0.38		10
17	0	1	1	2	8	0.36		12
18	0	1	0	1	9	0.31		3

ABCN, American Board of Clinical Neurophysiology; ABPN, American Board of Psychiatry and Neurology; AC2, Gwet agreement coefficient.