ORIGINAL RESEARCH

A Multicenter Training and Interrater Reliability Study of the BASED Score for Infantile Epileptic Spasms Syndrome

OPEN

John R. Mytinger,* Dara V. F. Albert,* Shawn C. Aylward,* Christopher W. Beatty,* Sonam Bhalla,†‡ Sonal Bhatia,§ Guy N. Brock,‖¶ Micheal A. Ciliberto,# Purva R. Choudhari,** Daniel J. Clark,* Jennifer Madan Cohen,†† Theresa M. Czech,# Megan M. Fredwall,‡‡ Ernesto Gonzalez-Giraldo,§§ Chellamani Harini,‖‖ Senyene E. Hunter,¶¶ Amanda G. Sandoval Karamian,## Akshat Katyayan,*** Isaac Kistler,†††‡‡‡ Neil Kulkarni,* Virginia B. Liu,§§§ Corinne McCabe,‡‡ Thomas Murray,* Kerri Neville,‖‖‖ Shital H. Patel,¶¶¶ Spriha Pavuluri,### Donald J. Phillips,§§§ Debopam Samanta,**** Deepa Sirsi,†††† Emily M. Spelbrink,‡‡‡‡ Carl E. Stafstrom,§§§§ Maija Steenari,§§§ Danielle S. Takacs,*** Tyler Terrill,†††† Linh Tran,‖‖‖‖ Jorge Vidaurre,* and Daniel W. Shrey§§§

*Department of Pediatrics, Division of Pediatric Neurology, Nationwide Children's Hospital, The Ohio State University, Columbus, Ohio, U.S.A.; †Division of Neurology, Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia, U.S.A.; ‡Division of Pediatric Neurology, Children's Healthcare of Atlanta, Atlanta, Georgia, U.S.A.; §Department of Pediatrics, Division of Pediatric Neurology, Medical University of South Carolina and Shawn Jenkins Children's Hospital, Charleston, South Carolina, U.S.A.; ‖Department of Biomedical Informatics, College of Medicine; ¶Center for Biostatistics, Wexner Medical Center, Biostatistics Resource at NCH (BRANCH), Nationwide Children's Hospital and The Ohio State University, Columbus, OH, U.S.A.; #Stead Family Department of Pediatrics, University of Iowa, Iowa City, Iowa, U.S.A.; **Department of Pediatrics, Division of Pediatric Neurology, University of Texas Southwestern, Dallas, Texas, U.S.A.; ††Department of Neurology, Connecticut Children's Hospital, University of Connecticut, Hartford, Connecticut, U.S.A.; ‡‡Department of Pediatrics, Division of Pediatric Neurology, Nationwide Children's Hospital, Columbus, Ohio, U.S.A.; §§Department of Neurology, Division of Epilepsy, University of California, San Francisco, San Francisco, California, U.S.A.; ‖‖Department of Neurology, Division of Epilepsy, Harvard University, Boston, Massachusetts, U.S.A.; ¶¶Department of Neurology, Division of Child Neurology, University of North Carolina Hospitals, Chapel Hill, North Carolina, U.S.A.; ##Division of Neurology, Department of Pediatrics, University of Utah, Salt Lake City, Utah, U.S.A.; ***Department of Pediatric Neurology and Developmental Neuroscience, Texas Children's Hospital, Baylor College of Medicine, Houston, Texas, U.S.A.; †††Center for Biostatistics, The Ohio State University Wexner Medical Center, Columbus, Ohio, U.S.A.; ‡‡‡Biostatistics Resource at Nationwide Children's Hospital, Abigail Wexner Research Institute at Nationwide Children's Hospital, Columbus, Ohio, U.S.A.; §§§Division of Neurology, Orange California & Department of Pediatrics, Children's Hospital of Orange County, UC Irvine, Orange, California, U.S.A.; ‖‖‖Department of Pediatrics, Division of Pediatric Neurology, University of Michigan, Ann Arbor, Michigan, U.S.A.; ¶¶¶Department of Pediatrics, Division of Pediatric Neurology, Duke University Medical Center, Durham, North Carolina, U.S.A.; ###Children's Hospital and Medical Center, Omaha, Nebraska, U.S.A.; ****Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, U.S.A; ††††Department of Pediatrics, Division of Pediatric Neurology, University of Texas Southwestern Medical Center, Dallas, Texas, U.S.A; ‡‡‡‡Department of Neurology, Division of Child Neurology, Stanford Children's Hospital, Stanford University, Palo Alto, California, U.S.A; §§§§Division of Pediatric Neurology, Departments of Neurology and Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, U.S.A; ‖‖‖‖Comprehensive Epilepsy Program, Jane and John Justin Neuroscience Center, Cook Children's Medical Center, Fort Worth, Texas, U.S.A.

Purpose: The best possible outcomes in infantile epileptic spasms syndrome require electroclinical remission; however, determining electrographic remission is not straightforward. Although the determination of hypsarrhythmia has inadequate interrater reliability (IRR), the Burden of AmplitudeS and Epileptiform Discharges (BASED) score has shown promise for the reliable interictal assessment of infantile epileptic spasms syndrome. Our aim was to develop a BASED training program and assess the IRR among learners. We hypothesized moderate or better IRR for the final BASED score and the presence or absence of epileptic encephalopathy (+/−EE).

Methods: Using a web-based application, 31 learners assessed 12 unmarked EEGs (length 1–6 hours) from children with infantile epileptic spasms syndrome.

Results: For all readers, the IRR was *good* for the final BASED score (intraclass correlation coefficient 0.86) and +/−EE (Marginal Multirater Kappa 0.63). For all readers, the IRR was *fair* to *good* for all individual BASED score elements.

Conclusions: These findings support the use of our training program to quickly learn the BASED scoring method. The BASED score may be a valuable clinical and research tool. Given that the IRR for the determination of epileptic encephalopathy is not perfect, clinical acumen remains paramount. Additional experience with the BASED scoring technique among learners and advances in collaborative EEG evaluation platforms may improve IRR.

Key Words: BASED score, EEG, Infantile epileptic spasms syndrome, Pediatric epilepsy.

(J Clin Neurophysiol 2025;42: 230–234)

Infantile epileptic spasms syndrome is a severe form of early childhood epilepsy with significant consequences regarding development and later life epilepsy.[1] Early and effective treatment can improve developmental and epilepsy outcomes.[2] The best possible child-specific outcomes require rapid electroclinical remission defined as complete cessation of clinical infantile spasms and the resolution of any EEG features consistent with an epileptic encephalopathy (EE). Although the presence of hypsarrhythmia is an important EEG marker of EE, the EEGs of many children with infantile epileptic spasms syndrome do not show hypsarrhythmia and its determination has poor interrater reliability (IRR).[3–5] The Burden of AmplitudeS and Epileptiform Discharges (BASED) score (Table 1) has shown promise for the interictal assessment of children with infantile epileptic spasms syndrome with high levels of IRR between three reviewers

at a single center.[4,6] Using a multicenter network of EEG readers, the aim of this study was to develop a formal BASED training program and to assess the IRR among learners. In the reading of EEGs from children with infantile epileptic spasms syndrome, we hypothesized moderate or better IRR for the final BASED score and the presence or absence of epileptic encephalopathy (+/−EE).

METHODS

As part of a multicenter biomarker and education grant funded by the Pediatric Epilepsy Research Foundation, we developed a formal BASED training program. This program included the independent review of the 2021 BASED score article (including the brief instructional within Mytinger 2021, et al.)[6] and two less than one-hour virtual sessions of discussion with guided EEG tracing review (**Supplements: Video 1, Video 2** http://links.lww.com/JCNP/A279, http://links.lww.com/JCNP/A280). To assess IRR, learners completed a posttest that included independent EEG tracing review and scoring. Three experts (J.R.M., D.V.F.A., and J.V.), who helped to create the BASED score and had extensive experience in its clinical application, independently reviewed all EEGs to determined gold standard scores (using a best 2/3 method). The IRR values for the three expert reviewers were excluded from the "All Readers Combined" and all other subsets. To assess accuracy, learner scores were compared with gold standard scores, and the average number and average percent correct were calculated. All learners and experts were blinded to clinical information for EEG review. Readers assessed 12 deidentified studies (six pretreatment and six posttreatment) from seven different children using a web-based application (Persyst Mobile version 1.2.15, Persyst Development Corporation, Solana Beach, CA). EEGs were listed in random order so that readers were not aware of pretreatment/posttreatment status or any relationship between studies. The length of EEG epochs ranged from 1 h (for outpatient studies) to 6 h (for long-term monitoring studies). To limit bias, EEG files were unmarked and readers were asked to self-identify epochs for review. For example, to determine a final BASED score of 3 to 5, readers were asked to identify the most epileptic five-minute sleep epoch. If no score was reached using this 5-minute epoch, readers were instructed to review an "adequate" portion of the remaining EEG to determine the 0 to 2 score.

All readers were provided a synopsis of the BASED score with associated rules of application as summarized in Table 1. Compared with the summary table of the 2021 BASED score in our prior publication,[6] we modified the current Table 1 and its associated rules to improve clarity. Importantly, for the description of BASED scores 4 and 5, we chose language more consistent with the concept of EE as an electroclinical diagnosis. We replaced "probable EE" and "definite EE" with *suggests EE* and *strongly suggests EE* for a BASED 4 and 5, respectively. Throughout this article, although we use the phrase *presence or absence of EE* (+/−EE), we acknowledge that the BASED score cannot be used to diagnose EE without clinical correlation. EEG findings to suggest the absence of EE were designated by a BASED score of 3 or less. After the posttest, readers were provided a pictorial answer key based on gold standard scores (see **Figure, Supplemental Digital Content 1,** http://links.lww.com/JCNP/A266, Pictorial Posttest Answer Key).

All EEGs were performed at Nationwide Children's Hospital and were pre-reviewed by JRM to assure technical adequacy. EEGs were deidentified, converted to European Data Format, and securely transferred to a data repository for the previously noted multimodal biomarker study (primary site: Children's Hospital of Orange County). We used a consecutive sample of five enrolled children (nine EEGs). Three EEGs from two additional nonconsecutively enrolled children were also included—one posttreatment EEG that replaced a technically inadequate study and two EEGs from another child who had large amplitude background slow waves noted on the posttreatment EEG during pre-review. The latter was performed to assure sufficient opportunity for learners to evaluate background amplitude. Levels of IRR for the final BASED score, ±EE, and all BASED score elements were assessed by reader demographics.

Statistical Analysis

For the BASED score, IRR values for various subsets were calculated using two-way random, single measures, absolute agreement using intraclass correlation coefficient (intraclass correlation coefficient command of psych package in R). The intraclass correlation coefficient agreement was designated as follows: < 0.5 = poor, 0.5 - < 0.75 = moderate, 0.75 to 0.9 = good, and > 0.9 = excellent.[7] For the component scores and ±EE, interrater agreement values were calculated using Free-Marginal Multirater Kappa (κ_{Free}) to adjust for binary outcomes and low prevalence.[8] Agreement for κ_{Free} was designated as follows: <0.4 = poor, 0.4 to <0.6 = fair, 0.6 to 0.75 = good, and >0.75 = excellent. All analyses were performed in R version 4.3.0.[9]

The authors have no funding or conflicts of interest to disclose.
This study was approved by the Children's Hospital of Orange County institutional review board.
Supported by the Pediatric Epilepsy Research Foundation (88-06042021, 2021); the Children's Hospital of Orange County Chief Scientific Officer Award (2022); the Children's Hospital of Orange County Neuroscience Institute (2021). This study was funded by the Pediatric Epilepsy Research Foundation. Nonfinacial support comes from the Pediatric Epilepsy Research Foundation.
This work will be presented in part at the annual meeting of the American Epilepsy Society, December 3, 2023.
These funding sources had no role in the study design, the collection, analysis or interpretation of data, the writing of the manuscript, or the decision to submit the article for publication.
Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.clinicalneurophys.com).
Address correspondence and reprint requests to John R. Mytinger, MD, Infantile Spasms Program, The Ohio State University, Nationwide Children's Hospital, 700 Children's Dr., Columbus, OH 43205; e-mail: John.Mytinger@nationwidechildrens.org.

ISSN: 1537-1603/24/4203-0230
DOI 10.1097/WNP.0000000000001101

TABLE 1. The 2021 BASED Score and Synopsis

	2021 BASED Score	Description
	0	Normal
	1	Any definite nonepileptiform abnormality
Apply scores 0–2 to remainder of study	**2**	1 or 2 spike foci **AND** no channel with abnormal large amplitude
	3	*There are either of two ways to arrive at a score of 3:* 1. ≥3 spike foci <50% of one second bins **AND** no channel with abnormal large amplitude 2. 0 – 2 spike foci but ≥1 channel with abnormal large amplitude
Apply scores 3–5 to a 5-min sleep epoch	**4 (Suggests EE*)**	*There are either of two ways to arrive at a score of 4:* 1. ≥3 spike foci <50% of one second bins **AND** ≥1 channel with abnormal large amplitude 2. Not meeting criteria for 5 but includes GMFS or paroxysmal voltage attenuations
	5 (Strongly suggests EE*)	≥3 spike foci that are ≥50% of one second bins

BASED: Burden of AmplitudeS and Epileptiform Discharges, GMFS: grouped multifocal spikes, EE: epileptic encephalopathy; Table and Text adapted from Mytinger et al.[6]

*The diagnosis of EE requires clinical correlation with score of 4 or 5 only suggesting EE.

BASED Score Rules:

1. Apply scores 3–5 to the most epileptic 5-minute sleep epoch; if no score reached, apply scores 0–2 to an adequate sample of the remaining tracing.
2. Do not apply the BASED score to EEG studies without sleep.^
a. An epoch with stuttering sleep with intermittent arousals is acceptable.
3. The BASED score is applicable for children 1–24 months corrected age.

^Utilization of a wakefulness-only EEG may result in a lower BASED score compared with sleep and thus may not adequately represent outcome. BASED was developed using EEGs inclusive of sleep; future study may clarify the utility of applying BASED to an EEG without sleep.

≥3 Spike Foci Rules:

1. May be at least one from each hemisphere OR all from one hemisphere (may include midline).
2. If ≥50%* ≥3 spike foci in entire study but no 3 spike foci within 5 minutes, and no channel with abnormal large amplitude, BASED score is 2 + uncommon ≥3 spike foci.
3. Only a single spike is necessary to designate an independent spike foci.

Spike Burden Rules:

1. Calculate ≥50%* ≥3 spike foci by determining whether 10 or more 15 seconds pages in a 5-minute sleep epoch include ≥8/15 1 s bins with at least 1 spike.

*The BASED score does not use a "spike wave index" and thus the actual number of bins with at least 1 spike may be <50%, but if there are at least 10 pages with ≥8/15 1 s bins with a spike, and there are 3 spike foci, the BASED score is 5. This method was preferred to improve efficiency—often less than the total 5 minutes will need to be assessed.

Amplitude Rules:

1. Peak-to-peak (most negative to most positive [or vice versa] along a continuous ascending or descending wave) amplitude on a longitudinal bipolar montage, refers to background waves and excludes the slow wave with a preceding spike and the field of these waves in other channels
2. Waves must be common: when assessing a single channel, present at least once in 10 or more 15 seconds pages in a 5-minute sleep epoch

3. Abnormal Large Amplitude:

a. ≥200 μv: Fp1-F7, F7-T3, Fp1-F3, F3-C3, C3-P3, Fp2-F4, F4-C4, C4-P4, Fp2-F8, F8-T4
b. ≥300 μv: T3-T5, T4-T6
c. Excluded: Fz-Cz, Cz-Pz, T5-O1, P3-O1, P4-O2, T6-O2‡

‡As previously described [6], because of the common occurrence of large amplitude background waves within midline and occipital channels among normal children, these channels were excluded for the amplitude assessment in the 2021 BASED score. Similarly, the common occurrence of ≥200 μv slow waves in posterior temporal channels in normal children required a higher cutoff (≥300 μv) for these channels.

Grouped Multifocal Spikes Definition:

1. At least 2 independent spike foci in each hemisphere within a fairly well-delineated group (may include midline).
2. For hemispheric grouping, at least 3 independent spike foci in one hemisphere (may include midline) within a fairly well-delineated group.

Paroxysmal Voltage Attenuation Definition:

1. Definite sudden change from ongoing background activities appearing as a relative attenuation most often lasting 1 s (but may last several seconds), often occurring after an epileptic discharge, and usually diffuse (but may be present in only one hemisphere).

Remission Rules:

1. Pretreatment score of 4 or 5, must improve to ≤3.
2. Pretreatment score of 3, must improve to ≤2.

RESULTS

Learner Demographics

The program was completed by 31 learners/EEG readers (2 child neurology residents, 3 clinical neurophysiology fellows, and 26 faculty physicians) who represented 18 different medical centers. Most learners were participants in the above noted multicenter biomarker study and or were members of the Pediatric Epilepsy Research Consortium Infantile Spasms Special Interest Group. Some of these participants recruited additional learners from their respective medical centers. Additional learner demographics are summarized in Table 2.

Interrater Reliability

Interrater reliability values are summarized in Table 2. For all readers combined, and within all reader subsets, the IRR was *good* for the determination of the final BASED score. For all readers combined, the IRR was *good* for >50% > or = 3 spike foci, 1–2 spike foci, grouped multifocal spikes, and ±EE. For all readers combined, the IRR was *fair* for <50% > or = 3 spike foci, paroxysmal voltage attenuations, and the common presence of abnormal large amplitude background waves in at least one channel. When considering all subsets, we noted a gradient of agreement for the background amplitude element: *poor* IRR among trainees, *fair* to *good* IRR among attending faculty physicians, and *excellent* IRR for experts.

Learners' Average Number Correct and Average Percent Correct Compared With Gold Standard Scores

When compared with gold standard scores, the average number of correct scores for the final BASED score and all elements were similar among learners in all subsets (see **Table, Supplemental Digital Content 1,** http://links.lww.com/JCNP/A267). For all readers combined, the average number of correct determinations of the BASED score of 12 was 8.5, with an average percent correct of 70%. For all readers combined, the average number of correct determinations for ±EE of 12 was 10, with an average percent correct of 83%. For all readers combined, 27 learners (87%) correctly determined ±EE for at least 9 of 12 studies.

DISCUSSION

After a formal training program, the IRR of all learners was *good* for our primary outcomes—the determination of the final BASED score and the presence or absence of EE. For all readers combined, the IRR of BASED score elements was *fair* to *good*. These findings support the use of our training program to learn the BASED scoring method.

Previous IRR studies of the BASED score used preselected 5-minute EEG clips reviewed by three reviewers at a single center.[4,6] By contrast, we used a large number of EEG reviewers (31 learners) from 18 different medical centers who reviewed EEG tracings ranging in duration from 1 to 6 h. Our findings support the use of the BASED score for multicenter research collaboration.

Before this training program, only one learner had used the BASED score clinically for more than one year and three learners had used the BASED score for less than 1 year—all other learners were not using the BASED score clinically. Despite this limited use, we showed that learners who completed just a few hours of training, including independent and guided learning, were able to achieve *good* IRR for our primary outcomes. Future learners can complete our training program by studying the previous[6] and current articles (including the Table 1, and **Supplements: Video 1, Video 2**, http://links.lww.com/JCNP/A279, http://links.lww.com/JCNP/A280).

It is not surprising that the determination of one or two independent spike foci had only *fair* IRR for all readers combined. It is known that the determination of epileptiform discharges among individual sharply contoured waveforms is only *fair* among EEG readers.[10] The evaluation of epileptiform discharges is particularly difficult in children with infantile epileptic encephalopathies given that many actual spikes are poorly formed. Within our training program, we emphasized a high degree of sensitivity in the determination of spikes. We likewise recommended a high degree of sensitivity when evaluating for grouped multifocal spikes and paroxysmal voltage attenuations. Although we noted *good* IRR among all learners combined for grouped multifocal spikes, there was only a *fair* degree of IRR in the determination of paroxysmal voltage attenuations and the assessment of at least one channel with common abnormal large amplitude background waves. Paroxysmal voltage attenuations can be difficult to discern, and our definition may be too nonspecific for high levels of IRR. Yet, among the three expert reviewers, we noted that the IRR for paroxysmal voltage attenuation was *good*, and for the background amplitude assessment, the IRR was *excellent*. This suggests that the IRR of these elements may improve with experience. In fact, we noted an IRR gradient for the background amplitude assessment—*poor* for trainees, *fair* to *good* for faculty physicians, and e*xcellent* for expert reviewers.

Given that the ideal short-term outcome for children with infantile epileptic spasms syndrome is electroclinical remission, it is important that we can reliably determine whether an EEG

TABLE 2. Interrater Reliability by EEG reader Subsets

Measurement	All Readers Combined ($N = 31$)	3 Expert BASED Readers (Gold Standard)*	Boarded Either CNP or Epilepsy ($N = 24$)	Boarded CNP ($N = 14$)	Boarded Epilepsy ($N = 21$)	No Board, Still in Training ($N = 7$)	Years Posttraining <5 ($N = 10$)	Years Posttraining >5 ($N = 14$)
BASED†	0.86‖	0.89‖	0.88‖	0.87‖	0.90‖	0.79‖	0.88‖	0.88‖
≥50% ≥ 3 SF‡	0.75‖	0.89¶	0.77¶	0.73‖	0.78¶	0.70	0.90¶	0.68‖
< 50% ≥ 3 SF‡	0.52§	0.67‖	0.58§	0.54§	0.59§	0.33#	0.67‖	0.51§
1–2 SF‡	0.66‖	0.67‖	0.67‖	0.67‖	0.67‖	0.64	0.68‖	0.65‖
GMFS‡	0.60‖	0.56§	0.60‖	0.57§	0.58§	0.56§	0.59§	0.59§
PVA‡	0.57§	0.67‖	0.61‖	0.59§	0.61‖	0.44§	0.62‖	0.59§
≥1 Amp‡	0.56§	1.00¶	0.64‖	0.61‖	0.63‖	0.32#	0.72‖	0.57§
+/−EE‡	0.63‖	0.67‖	0.65‖	0.68‖	0.66‖	0.54§	0.63‖	0.66‖

≥1 Amp, abnormal large amplitude waves in at least 1 channel; ±EE, suggests EE (BASED 4 or 5) or not (BASED 0–3); BASED, final Burden of AmplitudeS and Epileptiform Discharges score; CNP, American Board of Psychiatry and Neurology, Clinical Neurophysiology; GMFS, grouped multifocal spikes; *N*, number of EEG readers; PVA, paroxysmal voltage attenuations; SF, independent spike foci.

*Expert data were not included in any other interrater reliability assessments (i.e., excluded from all other columns).

†Interrater agreement using Intraclass Correlation Coefficient: 0.75 to 0.9 = good (‖=orange).

‡Interrater agreement using Free-Marginal Multirater Kappa (κ_{Free}): <0.4 = poor (#=yellow), 0.4 to <0.6 = fair (§=light orange), 0.6 to 0.75 = good (‖=orange), >0.75 = excellent (¶=red).

has sufficiently improved to designate electrographic remission. Yet, it is critical to point out that the IRR for determining the presence or absence of EE using the BASED score is not perfect. This was true even among our three expert reviewers. Thus, although the BASED score can be used to determine the presence or absence of EE with *good* IRR in the research setting, the BASED score must not be the sole determinant of treatment for a child in the clinical setting. Instead, we emphasized in our training program that clinical acumen is more important than the BASED score when making clinical decisions.

Several challenges arose from our study design. The IRR may be better if EEG reviewers were permitted to use their native EEG review systems. In addition, the version of the Persyst Mobile on-line platform that we used did not include a dynamic wave amplitude measurement tool. Instead, we relied on an amplitude calibration mark. At the time of the study, a software limitation prevented readers from moving the calibration mark over the four upper most channels of the tracing. To overcome this, we used a hand-held calibration mark for these channels. This limitation may have affected the IRR of the background amplitude assessment, with a potentially significant downstream effect on determining BASED scores and the presence or absence of EE. However, the *excellent* agreement among experts for the background amplitude assessment suggests that the IRR of this element may improve with experience. An additional limitation of our study is the relatively small number of EEGs reviewed; this number was chosen to improve study feasibility, particularly given the relatively long duration of EEGs. Although prior IRR studies of the BASED score used preselected five-minute EEG clips, this study used longer duration EEGs ranging from 1 to 6 h. Although longer duration EEGs more accurately reflect the clinical application of the BASED score is some practices, this may have contributed to lower degrees of IRR. Greater guidance focused on epoch selection may further improve IRR. Finally, regarding our statistical analyses, the Free-Marginal Multirater Kappa was the ideal statistic for assessment of binary outcomes. Yet, because methods of calculating κ_{Free} are not available in standard software packages, we manually calculated this in R. Variance calculations for κ_{Free} are not standardized, and thus, confidence intervals were not created for this study.

CONCLUSIONS

Our study shows that the BASED scoring method can be learned quickly using a formal training program. Learners had *good* IRR for both the determination of the BASED score and the presence or absence of EE, suggesting that the BASED score may be a valuable clinical and research tool. Given that the IRR for the determination of EE is not perfect, clinical acumen remains paramount. Regarding the BASED scoring method, additional experience among learners, broader utilization by pediatric epileptologists, exposure during fellowship training, and advances in collaborative EEG evaluation platforms may improve IRR.

ACKNOWLEDGMENTS

The Pediatric Epilepsy Research Consortium Infantile Spasms Special Interest Group provided nonfinancial support for this work.

REFERENCES

1. Shields WD. Infantile spasms: little seizures, BIG consequences. Epilepsy Curr 2006;6:63–69.
2. O'Callaghan FJK, Edwards SW, Alber FD, et al. Vigabatrin with hormonal treatment versus hormonal treatment alone (ICISS) for infantile spasms: 18-month outcomes of an open-label, randomised controlled trial. Lancet Child Adolesc Health 2018;2:715–725.
3. Hussain SA, Kwong G, Millichap JJ, et al. Hypsarrhythmia assessment exhibits poor interrater reliability: a threat to clinical trial validity. Epilepsia 2015;56:77–81.
4. Mytinger JR, Hussain SA, Islam MP, et al. Improving the inter-rater agreement of hypsarrhythmia using a simplified EEG grading scale for children with infantile spasms. Epilepsy Res 2015;116093–98.
5. Mytinger JR. Definitions and diagnostic criteria for infantile spasms and west syndrome - historical perspectives and practical considerations. Semin Pediatr Neurol 2021;38:100893.
6. Mytinger JR, Vidaurre J, Moore-Clingenpeel M, Stanek JR, Albert DVF. A reliable interictal EEG grading scale for children with infantile spasms – the 2021 BASED score. Epilepsy Res 2021;173:106631.
7. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–163.
8. Randolph J. Free-marginal multirater kappa: an alternative to Fleiss' fixed-marginal multirater kappa. Joensuu, Finland, 2005.
9. Fleiss J. Statistical methods for rates and proportions. Hoboken, NJ: Wiley, 1981.
10. Jing J, Herlopian A, Karakis I, et al. Interrater reliability of experts in identifying interictal epileptiform discharges in electroencephalograms. JAMA Neurol 2020;77:49–57.