Quantitative analogues to visual spike interpretation in EEG

Eivind Aanestad

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2024

UNIVERSITY OF BERGEN

Quantitative analogues to visual spike interpretation in EEG

Eivind Aanestad

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 03.05.2024

Year: 2024

Title: Quantitative analogues to visual spike interpretation in EEG

Name: Eivind Aanestad

Print: Skipnes Kommunikasjon / University of Bergen

Acknowledgements

I would like to thank my supervisors, Jan Christian Brøgger and Nils Erik Gilhus, for their excellent scientific guidance, bureaucratic support, patience, writing expertise, and for providing the structure this project needed. Thanks to Jan Brøgger and Henning Olberg for devoting time and effort to collect data (marking spikes in hundreds of EEGs). We had a pleasant collaboration with a Danish research group, Sándor Beniczky and Mustafa Aykut Kural, enriching our second and third paper considerably, for which I am thankful. I would also like to thank Tom Eichele for helpful advice and answers to countless questions (mainly about signal analysis) and Harald Aurlien for healthy rational advice. Thanks to the Department of Clinical Neurophysiology at Haukeland University Hospital for the facilities and access to the data, and to my colleagues for constructive criticism, advice and cheers throughout this project. Thanks to Bergen Epilepsy Research Group and professor Marte Bjørk for several years of enriching research related lectures and discussions. Thanks to Noeska Smit for the best figure in Paper II. Last but not the least, I would like to thank my parents for being supportive and curious.

Scientific environments

- Bergen Epilepsy Research Group (BERG) at the Department of Clinical Medicine, University of Bergen. BERG's official web page can be found at https://www.uib.no/en/rg/epilepsy (last accessed at 04.06.2022).
- Department of Clinical Neurophysiology, Haukeland University Hospital, Bergen.

Table of contents

List of papers 5

Key acronyms 5

Ethical considerations 7

Introduction 10

The start of our research project 10

Epilepsy 11

EEG 12

Interictal epileptiform activity 13

Old and new definitions of interictal epileptiform activity 14

Terminology 15

Ictal or interictal epileptiform discharges 16

Quantitative measures of IEDs 16

Diagnostic accuracy in visual IED analysis 17

Interrater reliability in visual IED analysis 18

Other neurophysiological markers for epilepsy 18

Materials 19

Patients 19

Paper I 19

Paper II - III 20

Sharp transients in EEG 21

Methods 23

Study design 23

Qualitative or quantitative analysis (or both) 24
Using quantitative measures as decision support 26
How to measure a spike 26
Creating the Bergen Epileptiform Morphology Score (BEMS) 28
Aggregated BEMS variables 30
IED candidate morphology and IED candidate count 30
Interrater reliability (IRR) 31
Statistics 33
Results 34
Paper I 34
Paper II 36
Paper III 38
Discussion 42
Conclusions 56
Future perspectives 56
References: 58
Papers I-III: 66

List of papers

Paper I: Aanestad, E., Gilhus, N. E., Brogger, J (2020). **"Interictal epileptiform discharges vary across age groups."** Clin Neurophysiol 131(1): 25-33.

Paper II: Aanestad, E., Gilhus, N. E., Brogger, J (2021). **"A New Score for Sharp Discharges in the EEG Predicts Epilepsy."** J Clin Neurophysiol 40(1):p 9-16, January 2023.

Paper III: Aanestad, E., Gilhus, N. E., Olberg, H. K., Kural, M.A., Beniczky, S., Brogger, J (2023). **"Spike count and morphology in the classification of epileptiform discharges"** Front. Neurol. Volume 14:1165592, May 2023.

Key acronyms

AUC:	Area under the receiver operating characteristics curve
BEMS:	Bergen Epileptiform Morphology Score
$BEMS_{first}$:	BEMS of the first marked IED candidate in an EEG
$BEMS_{max}$:	BEMS of the IED candidate with the highest BEMS in an EEG
$BEMS_{sum}$:	BEMS of all IED candidates in an EEG summed together.
CC:	Correlation coefficient
EEG:	Electroencephalogram
ICC:	Intraclass correlation coefficient
IEA:	Interictal epileptiform activity (synonym: IED)
IED:	Interictal epileptiform discharge (synonym: IEA)
IEDC:	Interictal epileptiform discharge candidate
IRA:	Interrater agreement (synonym: IRR)
IRR:	Interrater reliability (synonym: IRA)
SCORE:	Standardized computer-based organized reporting of EEG
SQL:	Structured query language

Aims of this project

The common objective for all three papers was to find reproducible and visually relatable quantitative measures of interictal epileptiform discharge (IED) morphology that could inform, assist and improve the diagnostic work-up of patients with suspected epilepsy. Each paper had specific aims as outlined below.

Paper I

The primary aim in Paper I was to investigate whether morphological features of focal IEDs depended on age. Our clinical experience suggested that the spike amplitudes diminish with increasing age and we wanted to examine this using quantitative measures of amplitude as well as other morphological IED features. The secondary aim was to assess the occurrence of epilepsy type by age.

Paper II

Our primary aim in Paper II was to create a composite score from reliable quantitative morphological measures of the first sharp transient in the EEG that could distinguish between epileptiform and non-epileptiform sharp transients.

Paper III

The primary aim in Paper III was to examine whether the diagnostic accuracy of our classification model from Paper II could be improved by adding the marker *interictal epileptiform discharge candidate count* (IEDC count) and by applying the morphological score to several sharp transients in the EEG. IEDC count represents the total number of sharp transients suspected as IEDs in an EEG.

Ethical considerations

This project was approved by the Regional Committees for Medical and Health Research Ethics (reference code 2017/1512/REK Vest). We have not identified any ethical challenges regarding the project.

Abstract

Background: Some sharp transients in EEG represent interictal epileptiform discharges (IEDs) and are biomarkers for epilepsy. The current gold standard for detecting IEDs is by visual analysis. Detection relies mostly on experience, guided by qualitative descriptions and criteria from available guidelines.

Objectives: The main aim for our first paper was to investigate whether focal IED morphology depended on age. In our second paper, we hypothesized that selected quantitative morphological features of sharp transients can be combined into a score to classify EEGs as epileptiform or non-epileptiform. In our third paper, we hypothesized that the number of suspected IEDs, or IED candidates, represents an important variable in the classification of EEGs as epileptiform or non-epileptiform.

Materials and Methods: We included patients who had a routine EEG recorded in our EEG laboratory at Haukeland University Hospital during 2013– 2017, and who had an EEG report in our SCORE EEG database. In paper I, we examined whether focal IED morphology depended on age. In paper II, we examined whether quantified focal IED candidate morphology can be applied to classify EEGs as epileptiform or non-epileptiform. In paper III, we assessed whether the number of focal IED candidates and their morphology are relevant in the classification of EEGs as epileptiform or non-epileptiform. The clinical EEG conclusion was used as the primary outcome.

Results and Conclusions: IED morphology depended on patient age. IEDs became blunter, smaller, wider, and had a less prominent slow after-wave with increasing age. The quantitative morphological score for IED candidates classified EEGs as epileptiform or non-epileptiform with a sensitivity of 55% and a specificity of 91%. By applying the combination of quantified IED candidate morphology and IED candidate count, we could classify EEGs with a high IRR, sensitivity of 60% and specificity of 99%.

Sammendrag (abstract in Norwegian)

Bakgrunn: Noen skarpe transienter i EEG er interiktale epileptiforme utladninger (IEU) og representerer en biomarkør for epilepsi. Visuell analyse er den gjeldende gullstandarden for å detektere IEUer. En slik analyse avhenger av erfaringsgrunnlag, og er veiledet av kvalitative kriterier og beskrivelser fra tilgjengelige retningslinjer.

Mål: I første artikkel ville vi undersøke hvorvidt fokal IEU-morfologi er avhengig av alder. I andre artikkel var målet å utrede om den første fokale skarpe transienten i EEG kan kvantifiseres til en skåre som klassifiserer EEG som epileptiforme. I tredje artikkel fremsatte vi hypotesen at antallet mistenkte fokale IEUer, eller IEU-kandidater, er en viktig variabel i klassifiseringen av EEG som epileptiforme eller ikke-epileptiforme.

Materiale og metoder: Vi inkluderte pasienter som hadde registrert et rutine-EEG i vårt laboratorium ved Haukeland Universitetssykehus i tidsrommet 2013-2017, og som hadde en EEG-rapport lagret i vår SCORE EEG-database. I første artikkel undersøkte vi om IEU-morfologi var avhengig av alder. I andre artikkel klassifiserte vi EEG som ikke-epileptiforme eller epileptiforme basert på IEU-kandidat-skåren. I tredje artikkel kvantifiserte vi en eller flere IEU-kandidater i hvert EEG for å vurdere hvorvidt antallet IEU-kandidater er relevant i klassifisering av EEG som ikke-epileptiforme eller epileptiforme. Den kliniske EEG-konklusjonen ble brukt som gullstandard.

Resultater og konklusjoner: Fokal IEU-morfologi var avhengig av alder. IEU ble buttere, lavere i amplitude, bredere og fikk en mindre fremtredende etterfølgende langsom bølge ved høyere alder. Den kvantitative skåren for IEU-kandidater klassifiserte EEG som epileptiforme eller ikke-epileptiforme med sensitivitet=55% og spesifisitet=91%. Antall IEU-kandidater kombinert med kvantifisert IEU-morfologi klassifiserte EEG som epileptiformt med en sensitivitet på 60% og spesifisitet på 99%.

Introduction

The start of our research project

This work began in 2017. We had considered examining various topics within the domain of quantitative EEG, and decided to focus on IEDs or "spikes". The IED is a monumental EEG signal, a biomarker for epilepsy, an everyday challenge for EEG readers everywhere, known and described for nearly a century. Initially, there was confusion between what we now know as interictal and ictal epileptiform activity [1], with beginning understanding of the separation of interictal and ictal activity around the 1940s [2]. A Norwegian paper from 1959 reflects this incomplete understanding [3]. Despite a long history of research on IEA, there are still no applicable quantitative reference values to aid in EEG interpretation, which causes clinical problems [4-6]. We accepted the challenge and started putting numbers to spikes. The definition of epileptiform activity got an update in 2017 [7], while we were still in the early planning stages of our project. The reworked definition signaled that the science of spike interpretation had not peaked, but that there were still opportunities to improve interrater agreement and diagnostic accuracy, core elements of a useful diagnostic marker. Since 1974, a spike had been defined by the International Federation of Clinical Neurophysiology as a sharp wave typically seen in the EEG of patients with epilepsy[8]. This was a useful definition if you were already an experienced EEG reader, but not very helpful for EEG apprentices. However, the 2017 update introduced morphological criteria for typical spike features, essentially dissecting the waveform and shifting the focus over to the various elements that make up spikes. The new criteria are still qualitative and experience-based like the old definition, and therefore they continue to present similar challenges to the EEG reader. For example, the first criterion states that an IED should have a pointy peak. Normal EEG background activity contains many pointy peaks, as several physiological transients have pointy peaks as a typical feature [9]. When is a peak pointed to the degree that it should be considered a possible IED, or even a definite IED? Our hypothesis was

that quantifying such morphological features would improve our knowledge of which parameters are the most impactful and reliable in spike detection and classification.

We had access to a large SCORE EEG [10, 11] database with thousands of structured reports spanning several years of clinical EEG interpretation at the Department of Neurophysiology, Haukeland University Hospital, Bergen. The EEG findings in the SCORE reports, such as the spikes that we focused on, had a timestamp that pointed to the corresponding raw EEG-signal, providing an unambiguous link between the clinical classification of findings and the EEG itself. This gave us a very robust starting point by providing easy access to the visual gold standard for spike classification and saving us a lot of time harvesting and managing data. We used SQL to extract data from the SCORE database, MATLAB for signal analysis, STATA for statistics and dataset management, and GitHub for script management and to enforce reproducible research [11-14].

Epilepsy

Epilepsy is a brain disease that generates unprovoked epileptic seizures. It affects more than 50 million people worldwide with a prevalence around 4-12 per 1000 people and an incidence around 50 per 100 000 people per year [15, 16]. The annual worldwide costs have been estimated to about 119 billion USD, where 83% of the costs benefits only 15% of the epilepsy population in high-income countries [15]. Epilepsy has a wide range of causes (structural, genetic, infectious, metabolic, immune or unknown) and heterogenous seizure semiology (focal seizures, generalized seizures, with or without loss of consciousness etc.). The prognosis depends on epilepsy type or syndrome [17, 18]. Epileptic seizures can have deadly consequences such as in sudden unexpected death from epilepsy (SUDEP) and severe status epilepticus [19, 20]. Epilepsy is associated with many comorbid conditions like cognitive impairment, learning disabilities, memory difficulties, physical disabilities, psychiatric and behavioral problems, and stigma [21].

A task force for The International League Against Epilepsy published a practical definition of epilepsy in 2014 as either at least two unprovoked seizures that occur with greater than 24 hours apart, only one unprovoked seizure with a recurrence risk estimated to be above 60% during the next 10 years, or the diagnosis of an epilepsy syndrome [22]. Interictal epileptiform discharges (IEDs) in scalp EEG represent an important biomarker for epilepsy and are particularly relevant in the diagnostic work-up of patients that had only one unprovoked seizure, since IEDs will heavily impact the assessment of recurrence risk.

EEG

Scalp electroencephalogram, hereafter referred to as just EEG, represents a recording of electric potential differences between electrodes that are placed on the scalp. An advantage of EEG is the high temporal resolution, hundreds or thousands of samples are recorded per second, making it possible to record sudden and brief changes in brain activity, like IEDs. The spatial resolution and localizing value are generally not strengths of EEG, although there are methods to localize sources of synchronized neuronal activity, like in electrical dipole analysis and averaging [23]. Spatial resolution can be improved by high density electrode arrays, but for general purpose routine EEG recordings the 10-20 or 10-10 electrode montage is used (21-74 electrodes) [24]. The 10 and 20 refers to the distance between electrodes in percentage relative to the scalp size. The EEG signal is measured in microvolts and is relatively weak compared to extracerebral artifacts like body movements. In the context of spike detection, the normal or physiological background activity is considered as background noise [23]. A large cortical area that generates synchronous activity, usually >10 square centimeters, is required for an electric potential difference to be visible in free running EEGs [25].

EEG recordings are non-invasive and relatively cheap to perform. The duration of a routine EEG recording is usually 20-30 minutes, and the time required for analysis is

roughly 12 minutes, depending on whether there are abnormal findings or not [26, 27]. There are many clinical indications for an EEG. Referral reasons can be summarized as a question whether there is cerebral dysfunction and if so, what type of cerebral dysfunction that occurs. Suspicion of epilepsy is the most common diagnostic question in the EEG referrals at our department. Interictal epileptiform activity (IEA) is a biomarker for epilepsy, and IEA represents the core subject in our three papers I-II-III.

Interictal epileptiform activity

IEA in EEG is the most relevant interictal biomarker for epilepsy [28] . When present, the correct diagnosis of epilepsy can be given earlier and with higher precision, and appropriate treatment can be initiated without depending on further seizures to confirm the diagnosis [29]. Constellations of IED morphology, topography and activating patterns can point to specific epilepsy types and syndromes [28]. Figure 1 shows the typical signal morphology of an IED signal after eliminating background noise by averaging independent IEDs. The mechanism of how IEDs are generated has been studied in magnetoencephalography and intracortical EEG recordings, and appears to be a widespread slow oscillations before the hypersynchronization of the IEDs themselves {Sheybani, 2021 #797;Westin, 2022 #910.

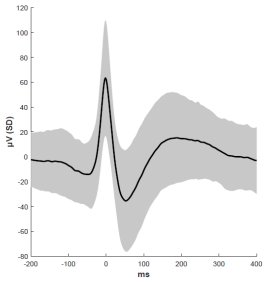

Figure 1. Averaged independent IEDs (N=349) [30].

Old and new definitions of interictal epileptiform activity

Interictal epileptiform activity represents a main topic of this project. IEDs are usually brief, visible in the EEG for less than a second, they always include at least one sharp or spiky wave, they are often followed by a slow after-wave, but there are otherwise many morphological variations. Several discharges may appear in trains, rhythmical or arrhythmical [31]. The old and new definitions of IED are shown below [7, 8, 32].

Oldest definition (1974) [8]:
"Epileptiform pattern. Interpretive term. Applies to distinctive waves or complexes, distinguished from background activity, and resembling those recorded in a proportion of human subjects suffering from epileptic disorders and in animals rendered epileptic experimentally. Epileptiform patterns include spikes and sharp waves, alone or accompanied by slow waves, occurring singly or in bursts lasting at most a few seconds."

Old definition (1999) [32]:
"Describes transients distinguishable from background activity, with a characteristic spiky morphology, typically, but neither exclusively nor invariably, found in interictal EEGs of people with epilepsy."

New definition (2017) [7]:
"Epileptiform discharges: Describes transients distinguishable from background activity with a characteristic morphology typically, but neither exclusively nor invariably, found in interictal EEGs of people with epilepsy. Epileptiform patterns have to fulfill at least 4 of the following 6 criteria:
(1) Di- or tri-phasic waves with sharp or spiky morphology (i.e. pointed peak).
(2) Different wave-duration than the ongoing background activity, either shorter or longer.

(3) Asymmetry of the waveform: a sharply rising ascending phase and a more slowly decaying descending phase, or vice versa.
(4) The transient is followed by an associated slow after-wave.
(5) The background activity surrounding 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 (see dipole). This is best assessed by inspecting voltage maps constructed using common-average reference."

Terminology

When referring to sharp EEG waveforms in general, whether epileptiform or non-epileptiform, sharp transients is an appropriate term, although they are sometimes also referred to as spikes, sharps, or just transients [7, 11]. When the distinction between epileptiform and non-epileptiform is important, then explicit terms like IED or non-epileptiform transient prevent misunderstandings. The prefix "interictal" means "between seizures", and is often left out for readability, but is sometimes needed to make a distinction from ictal epileptiform activity that is seen during seizures. The choice of terminology is context-dependent, and the meaning of a term should be obvious from the context. When referring to epileptiform discharges in this thesis, they will be specified explicitly as interictal epileptiform discharges (IEDs) or interictal epileptiform activity (IEA). Considering the vast plethora of morphological IED variants it is not surprising that creative terms may arise in EEG reports described in free text. Harald Aurlien's PhD research project addressed this problem and demonstrated the use of structured EEG reports stored in a computer database for research purposes [33-35]. His work led the way to Standardized computer-based organized reporting of EEG (SCORE), a standardized method for reporting EEG findings, that was published in 2013, to improve communication of EEG findings from EEG reader to the treating clinician [31]. As mentioned previously, the data for this project relied on a structured database of EEG

findings, which is much preferred to searching through unstructured free text documents in electronic patient journals.

Ictal or interictal epileptiform discharges

The distinction between interictal and ictal epileptiform activity is not always straightforward, probably because both EEG patterns stem from synchronized neuronal activity [22, 36]. Some studies indicate that IEDs can have a measurable impact on cognitive functioning independently of epileptic seizures [37, 38]. A short burst of rhythmic 3 Hz polyspike and slow wave in the EEG can be both interictal and ictal in juvenile myoclonic epilepsy, depending on whether they present together with any clinical symptoms or not [39]. Some periodic and rhythmic patterns with epileptiform morphology in critically ill patients do not quite fit the criteria as either ictal or interictal activity but is rather classified somewhere along an ictal-interictal continuum [40]. Despite these entanglements, sharp transients can for most purposes be safely categorized as interictal or ictal.

Quantitative measures of IEDs

EEG and IEDs are recordings of voltage fluctuations, so it is rather obvious to attempt precise quantifications. Descriptions of frequencies and amplitudes in EEG have a long history, dating back to 1929 in Hans Berger's paper "Über das Elektrenkephalogramm des Menschen" [41]. It is impossible to decipher exactly when IEDs became a marker for epilepsy, but EEG investigators were aware of EEG abnormalities occurring between seizures already in the 1930s [42]. Interictal waves and spikes were described in the 1940s and 50s [2, 43, 44]. A definition of IEDs that is familiar to modern EEG interpreters was published as part of a glossary in 1962 [45], updated in 1974 [8], 1999 [32] and 2017 [7]. The definitions contain only qualitative descriptions that can aid in visual spike detection. The ILAE guidelines for classification of epilepsy syndromes

contain some information regarding voltages and frequencies for specific epilepsy syndromes [28].

Despite the lack of quantitative information in definitions and guidelines, there are several studies that have quantified morphological IED features either as a step towards the development of automatic spike detectors [46-64], to discover associations between IED features and seizure risk [65], to assess reproducibility of visual analysis [66, 67], or to define their potential role in epilepsy syndrome classification [33, 68, 69]. These studies applied mimetic analysis, which is the reduction of more or less complex morphological features of the transient to simplified quantities (e.g. amplitude, sharpness, duration and area) [53]. Other non-mimetic methods are often included in automatic spike detectors, such as artificial neural networks [70], fourier and wavelet transform analysis [71, 72], template matching [73] and independent component analysis [48]. Older studies are often small in terms of patient numbers, often due to computational burden at the time. The computer code behind the previous automated spike detector work was not available, as code sharing was not usual then.

Diagnostic accuracy in visual IED analysis

The ability of a diagnostic test to accurately identify patients with and without a disease is referred to as diagnostic accuracy [74]. Diagnostic accuracy is often reported as sensitivity and specificity. Sensitivity denotes the proportion of subjects with the disease that is correctly identified by a positive test. Specificity denotes the proportion of subjects without the disease that is correctly identified as negative. False positive EEGs is a common problem due to over-reading and the faulty classification of physiological transients and various artifacts as IEDs [75]. A high specificity is especially important in patient populations with a low pretest probability of epilepsy. According to a systematic review by Bouma et al. [76], the pooled sensitivity of routine EEGs regarding seizure recurrence after an unprovoked first seizure was 17% for adults and 58% for children. The pooled specificity was 95% for adults and 70% for children. Time will tell if the

introduction of additional morphological criteria to the 2017 definition of epileptiform discharges [7] will improve the diagnostic accuracy of visual IED interpretation. Some recent studies have shown promising results when applying the new criteria in visual IED detection [77, 78].

Interrater reliability in visual IED analysis

A high diagnostic accuracy can not be expected unless EEG readers are in reasonable agreement regarding the presence of epileptiform discharges in EEGs. Variability in diagnostic interpretation is also important in other medical fields such as radiology. There is a longer standing tradition in radiology of more advanced research designs, larger datasets than has been in common clinical neurophysiology[79]. IRR can be reported numerically as percentage agreement or by chance-adjusted agreement statistics, e.g. Cohen's kappa or Gwet's AC1 [80, 81]. It is also common to describe IRR by ordinal categories as follows: A Kappa value of ≤ 0 as "no agreement", 0.01-0.40 as "fair agreement", 0.41-0.60 as "moderate agreement", 0.61-0.80 as "substantial agreement" and 0.81-1 as "almost perfect agreement" [82]. The IRR among expert EEG readers was found to be moderate for identifying individual IEDs and substantial regarding whether an EEG contained any IEDs, in a recent study by Jing et al [67]. Precise terminology, definitions and criteria represents the foundation for consistent and reliable interpretation of IEDs [83].

Other neurophysiological markers for epilepsy

While IEA in EEG is the most commonly used modality in the diagnostic work-up of patients with suspected epilepsy, other electrophysiological methods are also available. Ictal EEG, that is an EEG recording that contains an epileptic seizure, can verify and localize epilepsy, especially with added video for seizure semiology [84]. Magnetoencephalography can be used to detect interictal epileptiform activity[85, 86]. High-frequency oscillations are a new aspect of interictal epileptiform activity

characterization[87]. Measures of functional brain network connectivity is an interictal non-epileptiform EEG marker for epilepsy [88].

Materials

Patients

The source for our patient material in Paper I-II-III was a SCORE EEG database located at the Department of Clinical Neurophysiology, Haukeland University Hospital. Patients, EEGs and EEG reports were gathered in the period of March 4th, 2013 - October 29th, 2017. A follow-up diagnosis of epilepsy was gathered from the Haukeland University Hospital records covering all clinical departments from January 1st, 1999 until November 27th, 2019. This means that we had a 2 year minimum follow-up after the EEG recording, and a good coverage of previous EEG recordings. The upstream preselected EEG database material consisted of 10,547 consecutive patients who had one or more EEG reports in the SCORE database. Patients who had their EEG recorded at the intensive care unit (ICU) were excluded since this population tends to differ with regards to patient state, medication (sedatives), prevalence of epilepsy and comorbidities. Further selection of patients differed in each paper. Specific selection criteria were defined so that we should be able to ensure optimal answers to our research questions, minimize the risk of selection bias, obtain a sufficient sample size, but still with an acceptable work load..

Paper I

We selected one EEG from each patient as the material for Paper I, giving us a total of 10,547 EEGs. The patients were categorized into ten groups according to age in years as follows: <1 year, 1–9 years, 10–19 years, 20–29 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years, 70–79 years, and 80–101 years.The EEGs were grouped according to the diagnostic conclusion in the EEG report. 9,238 EEGs had a conclusion other than epilepsy while 1309 EEGs had a conclusion of epilepsy. The latter EEGs were further

grouped into generalized (207 EEGs), focal (875), or undefined epilepsy (227). Demographic characteristics were analyzed in all 10547 patients, and quantitative IED features were described in the focal epilepsy group.

Paper II - III

The main aims for Paper II and III were to establish sensitive and specific quantitative diagnostic markers for epilepsy that were reproducible. We applied a retrospective cohort study design using routine EEGs. Only spike- and epilepsy-naive patients were included in order to avoid unwanted bias in the form of reduced spike detection thresholds due to information from previous EEGs or a known hospital diagnosis of epilepsy. This reduced our EEG database material from 10,547 to 10,138 patients.

In Paper II, we included one EEG from each of 2063 patients that had either sharp transients or focal IEDs in their EEG. 37 patients were excluded due to missing EEG data. 350 EEGs contained focal IEDs, while the remaining 1713 contained non-epileptiform sharp transients. The EEGs were randomized half and half into a training dataset (N=1013) and a validation dataset (N=1013). Randomization was obtained using the STATA command “runiform()” to generate a column with random numbers, sorting the dataset according to those random values, and then selecting N ordered rows [13]. A random subset of 244 EEGs from the training dataset was extracted to assess IRA by two independent raters. A subset of 345 EEGs that overlapped between Paper I and Paper II was used to examine inter-method agreement between the automatic one-click algorithm and the manual 4-click annotation method for the quantification of sharp transients. We were able to include an additional external validation dataset (N=100) described in a paper by Kural [89].

The material for Paper III was a randomly extracted subset including 400 patients from the Paper II dataset of 2063 patients. These EEGs were randomized and split into a training dataset (N=200) and a validation dataset (N=200). An external dataset containing

60 routine EEGs [90] was used for further validation. The patients in the external datasets for Paper II-III had all undergone long-term video-EEG monitoring that contained a habitual paroxysmal event, either a non-epileptic event or an epileptic seizure, that served as the reference standard.

Sharp transients in EEG

The subject of analysis in Paper I-II-III is sharp transients. Figure 2 illustrates the general concepts for EEG and spike selection in our three papers. Our main objective in Paper I was to examine whether IED morphology depended on age. Accordingly, we analyzed only those transients that were classified as IEDs during routine clinical EEG evaluation at our department. We selected the first IED that had been marked previously in each EEG. In Paper II, our goal was to identify morphological measures that differed between non-epileptiform and epileptiform discharges. Accordingly, the first transient was selected and measured in all EEGs using the previous clinical classification as epileptiform or non-epileptiform as the primary outcome variable. The main objective in Paper III was to investigate whether the number of sharp transients in the recorded EEG was relevant for the visually based classification of IEDs. It is uncommon to mark every transient suspected to be an IED in clinical EEG evaluation practice, so we had to gather this information by re-marking transients in the EEG as part of the research project. We decided to limit the number of markings to 40 IED candidates per EEG, as we hypothesized that this number was well above the threshold for when there is likely to be enough evidence that an EEG would be classified as epileptiform. This number at the same time gave us sufficient data regarding the variance of morphological measures in EEG.

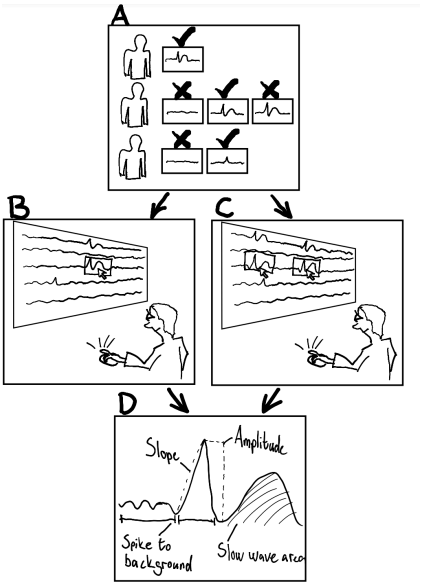

Figure 2. We selected the first EEG for each patient that contained a sharp transient (A), only IEDs for Paper I, either IEDs or non-epileptiform transients for Paper II-III. Pre-classified transients were re-marked for Paper I-II (B). One or several IED candidates were marked blindly in each EEG for Paper III (C). Quantitative measures were calculated from all marked transients (D).

The marking of sharp transients was carried out by one rater in Paper I, two raters in Paper II, and three raters in Paper III. The number of raters varied for various datasets within Paper II and III. Figure 3 illustrates the relationship between datasets and raters for each paper.

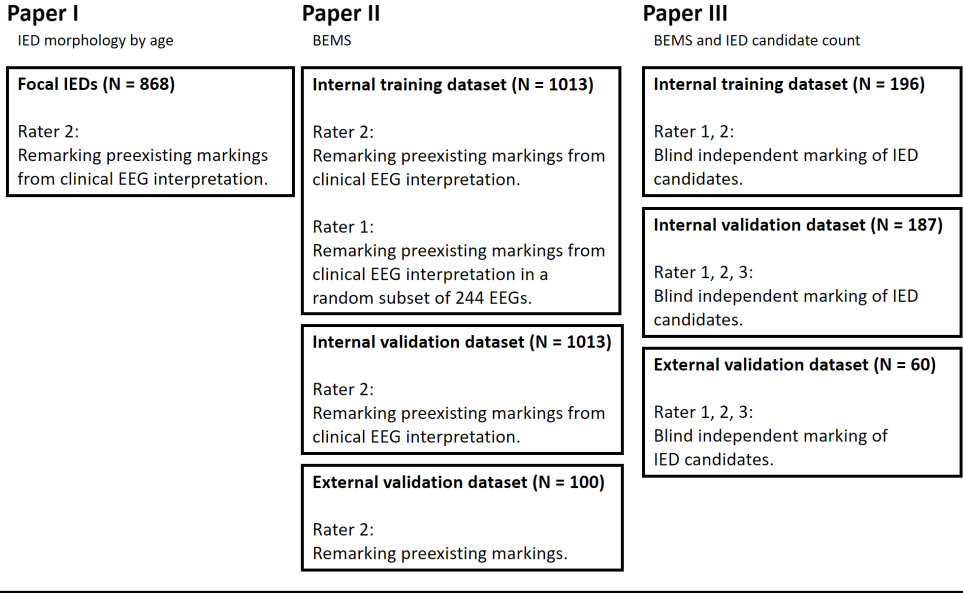

Figure 3. Illustration of the relationship between datasets and raters for Paper I-II-III.

Methods

Several aspects of study design and key methodological concepts were common for all three papers within this project. The quantitative measurements of sharp transients were nearly identical across the three papers. The only difference was that the algorithm that required four mouse clicks to quantify a transient in Paper I was replaced by a one-click algorithm for Paper II-III.

Study design

Paper I was a retrospective cross-sectional study where we studied the morphology of IEDs by age groups and occurrence of epilepsy type as per the clinical diagnostic EEG conclusion. We aimed for a large sample size in Paper I so that we would be able to compare many age groups while minimizing the statistical uncertainty in each group.

Paper II-III were retrospective case-control studies using two separate outcomes. The primary outcome was the conclusion based on EEG evaluation only of either focal IEDs or non-epileptiform transients (EEG-outcome). The secondary outcome was the follow-up diagnosis of epilepsy in the hospital records based on all clinical and non-clinical information available (clinical outcome). The follow-up time period ranged from 2-6 years (more precisely 769-2,447 days). Both Paper II and III used a conservative validation design where one half of the EEGs were used for building the regression model (Paper II) or setting cut-points for BEMS derived variables and IED candidate count (Paper III), while the other half of the EEGs were reserved for validation. The models were also validated applying independent, external datasets in Paper II-III.

Qualitative or quantitative analysis (or both)

Qualitative visual analysis, descriptions, definitions and criteria have been the traditional approach for EEG interpretation and spike detection. A standardized qualitative nomenclature covers the complexity and variety of IED patterns quite well and has been regarded as an effective tool for communicating and reporting EEG findings. However, there are good reasons to apply quantitative analysis in spike detection and description.

"If you can not measure it, you can not improve it." (Lord Kelvin)

Inter-rater agreement in qualitative visual EEG analysis is far from perfect, calling for a more objective classification method [67, 91, 92]. Morphological IED features are not dichotomous or categorical by nature, but rather exist on a continuous spectrum where quantitative measures should be appropriate. Numbers can be a resource in visual analysis by placing quantified features along a numerical spectrum, thereby providing data as evidence and explanation for the chosen classification or evaluation.

According to standard nomenclature a slow after-wave is either present or it is not. However, in real life after-waves come in various shapes and sizes. One might argue that every transient shown in the example in Figure 4 is followed by a slow after-wave. So, how can a simple quantitative value, like the area of the slow after-wave, add value to the EEG report? We will describe how three different measurement outcomes can improve the classification of slow-waves and give more precise information.

1: A large slow-wave area indicates that there is no ambiguity, thus supporting the classification.

2: A slow-wave area in the mid-range indicates that the classification was not clear cut.

3: A small slow-wave area would not support the classification, or even indicate a misclassification.

Similar to all other morphological features, the classification of a slow after-wave is not always as definite as we would like it to be.

In our three papers, we have quantified several morphological IED features, and combined them into a total score, the novel Bergen Epileptiform Morphology Score (BEMS). A higher BEMS value means that the transient has more typical IED features. Figure 4 illustrates how sharp transients look like when they are grouped according to BEMS percentiles. BEMS will be discussed in more detail later.

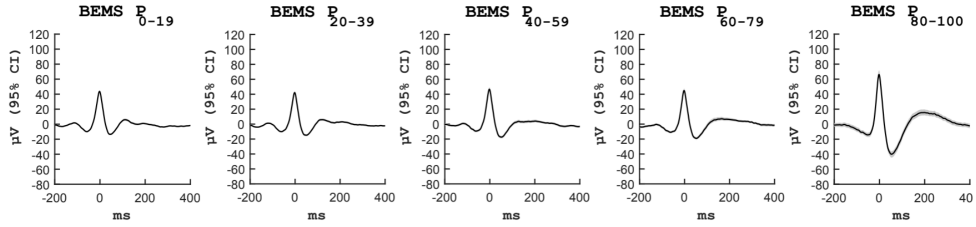

Figure 4. Average signals of all sharp transients from Paper II (N=2026) categorized into five BEMS-score percentile quantiles. 95% confidence intervals are shown as shaded gray areas.

Using quantitative measures as decision support

This work was firmly grounded on the standard visual analysis by clinical neurophysiologists. All sharp transients analyzed in this thesis were identified and gathered through standard visual EEG inspection and not by automatic spike detection. These were suspected IEDs with various degrees of certainty. Some exceptions will be further discussed in the limitations section. Selection bias for our included sharp transients can be regarded as a human EEG filter. The quantitative analysis and morphological measures were applied by us after the EEG had been filtered through this visual inspection. Therefore our measures can only describe the subset of EEGs that contains visually suspected IEDs, excluding the remaining EEG background waves. An advantage of analyzing sharp transients detected by human EEG readers is that the resulting measures give information about sharp transients selected by visual analysis, the current gold standard for EEG assessment. Furthermore, this selection should be well suited to establish and define markers for the classification of sharp transients as either epileptiform or non-epileptiform activity. We did this in Paper I. A detailed and computerized analysis of sharp transients can identify the essential features that separate non-epileptiform from epileptiform activity, as demonstrated in our Paper I-II. We gathered and classified our transients as follows: First, the human EEG reader marked sharp transients suspected of being IEDs. Then, the computer evaluated whether the particular transient represented an IED by calculating a score from the morphological measurements.

How to measure a spike

There are many ways to measure any property of a spike. There are no generally agreed-upon guidelines or recommendations. In our project, we put emphasis on deriving measures that had obvious associations with visual interpretation. The new definition of epileptiform discharges with morphological criteria in 2017 influenced our decision regarding which features to focus on. Before doing any actual spike measurements, the

transient signal needs to be separated from surrounding EEG activity by deciding where the transient starts and ends.

We developed a program in MATLAB for Paper I to calculate morphological measures from IEDs that had already been marked in the EEG during clinical routine EEG interpretation [93]. During clinical EEG interpretation montages could be changed freely, while the program was limited to a common average reference montage. The program required four clicks to measure each spike; one at the start of the ascending spike, one at the peak of the spike, one at the end of the descending spike, and one at the end of the slow after-wave if present. Four clicks per IED was laborious, time intensive, and prone to human error. Therefore, we improved the algorithm for Paper II, so that only one mouse click near the spike peak was sufficient for all spike measurements. This openly available algorithm, titled "EpiOneClick" [93], was able to locate the start and end points of the spike as well as the slow-wave component using idiosyncratic rules (Figure 5). EpiOneClick detected the proper peak within a 25 ms time window around the clicked sample. Then it crawled along the signal forwards and backwards to define three more time points; the spike start, the spike end which is also the start of the slow after-wave, and the end of the slow after-wave. The start or initial trough of the spike was identified by iterating backwards through all local voltage minima between the sample point of the peak and up to 200 ms before the peak. The marking for the spike start was updated to the current iteration of the local minimum as long as the slope was greater than 0.3 μV/ms between itself and the peak. The iteration process was terminated if the next iterative local minimum had a higher voltage than the previous one. The spike end or trailing trough of the spike was detected by a similar procedure, only that it was iterating forwards through local minima instead of backwards. The spike end at the same time marked the start of the slow after-wave. The end of the slow after-wave was estimated by first applying a smoothing average filter for the duration of 800 ms following the slow-wave start, and then marking the sample with the minimal voltage outside the time window and at least 166 ms from the slow-wave start, to exclude waves with a frequency

>6 Hz. Then, a figure was presented in a popup-window that showed the acquired signal with the automatically generated markings so that the EEG reader could verify that the intended IED or IED candidate had been properly marked.

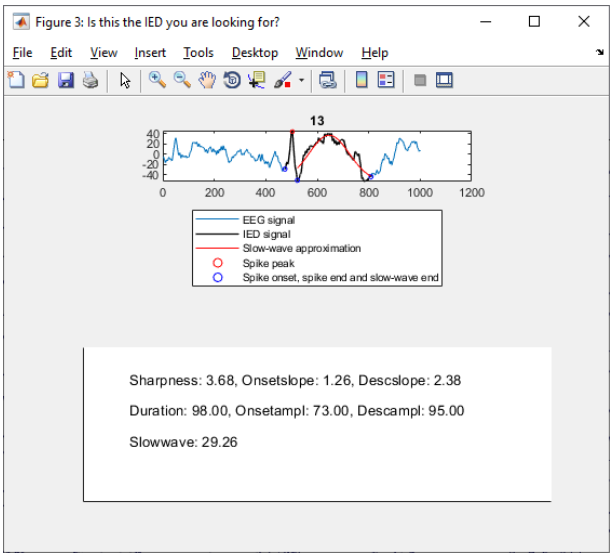

Figure 5. Illustration of how IED candidates were presented to the rater during the marking process. The purpose was for the rater to verify that the intended IED was properly quantified.

Creating the Bergen Epileptiform Morphology Score (BEMS)

In Paper II, we developed the Bergen Epileptiform Morphology Score (BEMS) for sharp transients in order to classify them as epileptiform or non-epileptiform discharges. A higher score meant that a transient had a more typical epileptiform morphology and was more likely to be an IED. The score combined several coefficients in a logistic regression model using the four most significant morphological features plus patient age in years as independent variables. To build the model we first selected 11 morphological measures for univariate analysis of dose response and distribution. The measures were: Spike

ascending amplitude (μV), spike descending amplitude (μV), spike to background power (%), sharpness (2nd derivative around peak), spike duration (ms), background amplitude (root-mean-square μV), slow after-wave area (weber), spike onset slope (μV/msec), spike descending slope (μV/msec), Henze asymmetry (ms/ms), and the number of channels where the sharp transient had been marked. All these morphological measures as well as age in years were then included in a preliminary logistic model using the clinical EEG classification regarding the presence of IEDs as the outcome. We conducted a stepwise elimination of independent variables one after the other, until the five most important measures remained. Each variable was divided into four categories based on its univariate dose-response curve. The maximal number of predictor variables (or degrees of freedom when categorized) in a multivariate regression model for a given dataset size is not known [94]. More than 10 events per predictor has been considered to be an acceptable methodological practice [94]. Fewer variables and larger datasets increase model stability and performance when applied to external datasets. Based on this, we decided that 10 or more observations per variable category was appropriate to ensure a statistically reliable model.

Table 1. The categorized predictor variables and their individual point contribution for the total BEMS-score.

Predictor variable	Category	Points
Spike descending amplitude (μV)	0-69	1
	70-89	0
	90-119	7
	>119	17
Spike onset slope (μV/ms)	0-0.9	0
	1.0-1.4	4
	1.5-1.9	5
	>1.9	11
Spike to background power (%)	>8.5	0
	4.7-8.5	9
	2.6-4.6	6
	0-2.5	14
Slow after-wave area (weber)	0-4	0

	5-9	6
	10-19	11
	>19	19
Age (years)	0-9	16
	10-19	0
	20-59	12
	>59	25

Aggregated BEMS variables

In Paper III, BEMS was calculated for one or several IED candidates for each EEG recording. Aggregated values from the distribution of BEMS-scores were analyzed and used to classify EEGs as epileptiform or not instead of using only the first IED candidate as we had done previously in Paper I-II. The variables used in Paper III were the following:

$BEMS_{max}$: The maximum BEMS of any IED candidate in an EEG

$BEMS_{sum}$: The sum of BEMS for all IED candidates in an EEG

$BEMS_{mean}$: The mean BEMS for all IED candidates, which was calculated for the purpose of investigating the correlation between morphology and IED candidate count.

IED candidate count: The number of IED candidates in an EEG.

Diagnostic classifier: The combination of BEMS and IED candidate count was applied into three criteria sets where at least one criteria-set had to be fulfilled in order to classify an EEG as epileptiform. The three criteria sets were defined as follows: In order for an EEG to be classified as containing IEDs, it had to contain either one IED candidate with BEMS $\geq$ 58, two IED candidates with BEMS $\geq$ 47 or seven IED candidates with BEMS $\geq$ 36.

IED candidate morphology and IED candidate count

In Paper III, we investigated whether the number of IED candidates in an EEG was relevant for classification of IEDs. According to our clinical experience, the occurrence of sharp transients that are recurring frequently during the EEG recording influences the

evaluation and conclusion, especially in those EEGs that have uncharacteristic IED candidates. The concept and relevance of IED candidate count is presumably well known among EEG interpreters, but surprisingly there was only one paper by Kural et al. [90] that had actually examined the influence of IED candidate count on IED classification. In Paper III, we combined IED candidate morphology and IED candidate count into one classification model. Our model was built using the training dataset (N=196) where two raters had marked IED candidates independently. The model consisted of three criteria-sets, where at least one criteria-set had to be fulfilled in order to classify an EEG as epileptiform, i.e. containing IEDs. Each of the three criteria-sets were in the format of "The EEG must contain at least *x* IED candidates with a BEMS above *y*". To find the optimal values for *x* and *y*, we had to test various combinations of three criteria-sets, and then choose the one with the best diagnostic performance. An exhaustive search through all possible combinations would be demanding since it would require classification of the whole dataset for each combination in order to obtain diagnostic performance metrics. The range of *x* would be 1-40 and of *y* 1-86, giving possible combinations in the order of billions. We decided to apply some limitations on the ranges of *x* and *y* such that we only had to evaluate a total of 5456 combinations of BEMS and IED candidate counts. We chose the combinations according to the sum of mean accuracy and kappa, requiring that the specificity was greater than 90%.

Interrater reliability (IRR)

IRR represents a crucial element for diagnostic tests, such as for our classification models applied in Paper II-III. Paper I contains no IRR-data.

In Paper II, we provided IRR data for two independent raters applying the BEMS algorithm on sharp transients that had already been marked during clinical routine EEG analysis. The transients were presented to the raters by automatically showing the EEG page containing the current transient with its pre-existing marking. This means that the

IRR results in Paper II depend only minimally on spike selection, except for channel selection by the rater. In some cases, the raters also had to choose one transient amongst several within the same EEG page. When we upgraded the BEMS-algorithm to only require one click on the spike peak instead of four (spike start, peak, end, and end of the slow after-wave), we also assessed inter-method reliability for a subset of transients that were common for Paper I and Paper II. The measures from Paper I that had required four clicks were compared to the measures from Paper II that only required one click.

In Paper III, two independent raters (rater 1 and 2) marked all transients in both the internal training dataset and the validation dataset, as well as in the external dataset. A third independent rater (rater 3) was recruited later in the project and only marked transients in the internal validation and external datasets. The classification model was built using the markings of rater 1 and 2 in the training dataset. The chosen cut-points were therefore independent of rater 3's markings. The optimal cut-points were determined by accuracy, while requiring the specificity to be > 90%. First, the optimal cut-point was determined for each individual rater. To determine a common cut-point that would be applied for all raters in the validation dataset, the average of the two cut-points given by the two raters individually was calculated for each diagnostic marker. The IRR calculated in Paper III depended on the selection of IED candidates, this in contrast to Paper II, since the marking process was blinded to previous markings from clinical routine analysis. The IRR resulted from accumulated data on a per EEG basis, not a per transient basis. The variables $BEMS_{max}$, $BEMS_{sum}$, and IED candidate count represent EEG-level variables obtained by finding the marked IED candidates with the highest BEMS ($BEMS_{max}$), the summed BEMS of all marked IED candidates ($BEMS_{sum}$), and the number of marked IED candidates (IED candidate count).

The algorithm that calculated quantitative morphological measures was improved for Paper II, so that it would only require one click on the spike peak instead of four clicks (see "How to measure a spike" previously). There is a risk of discrepancy when different

methods are used to measure the same object. Inter-method reliability was assessed for sharp transients that had been marked both for Paper I using the 4-click method, and for Paper II using the 1-click method.

Statistics

The subjects of analysis in Paper I-II were independent IEDs or IED candidates obtained by including the first IED (Paper I) or IED candidate (Paper II) from one EEG per patient. In Paper III, the subject of analysis was aggregated variables for many IED candidates, that is $BEMS_{max}$, $BEMS_{sum}$ and IED candidate count. Each of these measures was obtained from one EEG per patient. External datasets were used for Paper II (N=100) and Paper III (N=60), and they contained one EEG per patient. The transients were included in a similar way as for the datasets collected by us.

In Paper I, we examined whether IED morphology depended on age. Each of the quantitative measures (sharpness, ascending slope, ascending amplitude, duration, spike asymmetry, and area of slow-wave) was grouped according to age categories and visualized in box plots. Kruskal-Wallis-tests were used to test the dependency of the morphological measures on age category [95]. To examine whether age category had any effect on each morphological measure we performed multiple linear regressions with each measure as an outcome variable, and age category, etiology, brain region, and laterality as independent variables.

In Paper II, we combined morphological measures and patient age into a multivariate logistic regression model that classified sharp transients as epileptiform or non-epileptiform. The transients were grouped according to two outcomes; the EEG conclusion regarding the presence of IEDs, and a future clinical diagnosis of epilepsy during long-term follow-up. The stepwise construction of the logistic regression model has been described previously (Creating the Bergen Epileptiform Morphology Score,

page 28). Area under the receiving operating characteristics curve (AUC) was calculated for each morphological variable univariately, as well as for the BEMS score. Diagnostic performance, as evaluated by sensitivity, specificity and AUC, was examined for the BEMS score. The intraclass correlation coefficient was used to assess IRR and inter-method agreement for individual morphological measures and for the composite BEMS score.

In Paper III, we included one or several IED candidates per EEG and examined aggregated BEMS variables from the IED candidates as well as the number of IED candidates, limited to a range of 1-40. The mean IED candidate count was estimated for each of the raters in all included EEGs by a censored poisson model, where the lower censoring limit was 1 and the higher censoring limit was 40. Diagnostic performance was calculated as sensitivity, specificity and accuracy for $BEMS_{max}$, $BEMS_{sum}$, and IED candidate count. Our model combined $BEMS_{max}$ with IED candidate count, and this represented the diagnostic classifier. Intraclass correlation coefficients were used to assess IRA of the "raw" measures between raters. Gwet's AC1 was used to assess IRA after the cut points had been acquired for binary classification.

Results

Paper I

10,547 patients were included for analysis of epilepsy type by age. Occurrence of focal, generalized, or unspecified epilepsy type depended on age ($p<0.001$). An EEG conclusion of focal epilepsy was most common in all age groups, with the highest occurrence in children and the elderly. Adolescents had the highest occurrence of generalized epilepsy.

868 IEDs were available for quantitative measures. Except for in the youngest children (age groups <1 year and 1-9 years) the IEDs became shorter, blunter, wider, and the slow after-waves had a smaller area with increasing age, visualized in Figure 6. Spike asymmetry did not vary with age group. Spike sharpness (measured as the second derivative d^2V/dt^2 around the peak), was the highest in infancy with a median sharpness of 4.5, and lowest above age 80 years with a median sharpness of 2.0. Ascending and descending spike slopes had a maximum steepness in the age group 1-9 years with a median steepness of 2.5 μV/ms. Spike slopes became gradually less steep with increasing age down to a median of 1 μV/ms. Spike amplitude increased from infancy to childhood, and the median amplitude was 100μV in the age group 1-9 years. Spike amplitude gradually declined to a median of 60 μV with increasing age. Spike duration decreased to a median of 90 ms from infancy to childhood, then increased to a median of 130 ms in the elderly. The slow after-wave area was maximal in infancy with a median of 20 weber, then gradually decreased before stabilizing at around 10 weber for the age groups 10-101 years. Spike asymmetry had a stable median of around 0.8 ms/ms for all age groups. We applied multiple linear regression models for the quantitative IED measures as dependent variables and used brain region, etiology, laterality and age, as independent variables. This showed that age had the strongest effect on all the quantitative IED measures.

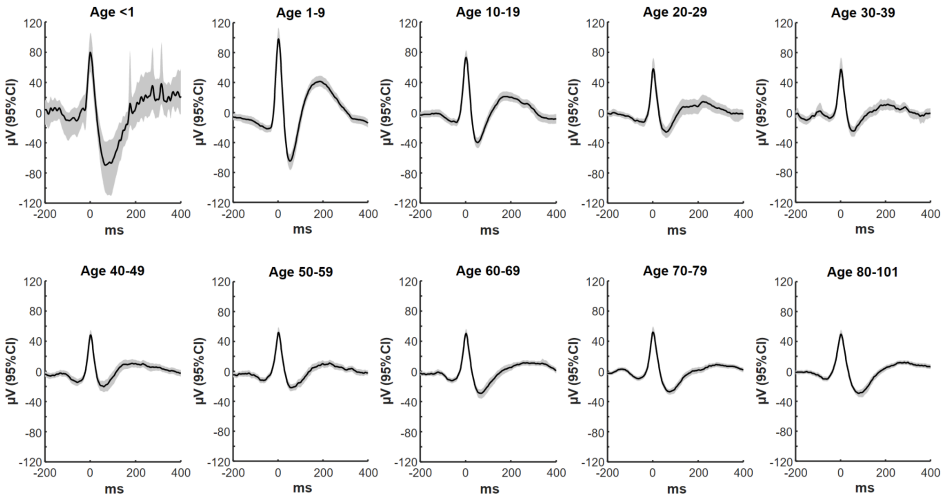

Figure 6 (Paper I). Averaged focal IEDs by age categories in years. 95% confidence levels are shown by a shaded gray area. (N = 868).

Paper II

The distributions for all quantitative morphological measures were examined in the complete dataset of 2026 sharp transients. All features except for spike duration differed significantly between the EEG outcomes of either focal IEDs or non-epileptiform transients ($p<0.001$). In univariate analysis, preceding background power, slow after-wave area, and descending spike amplitude had AUC > 0.7. Onset slope, descending slope, ascending amplitude, duration, spike sharpness and spike to background power had AUC ≤ 0.7. The BEMS score had an AUC = 0.84. The cumulative incidence of receiving a clinical diagnosis of epilepsy in the hospital records during long-term follow-up was 10% for patients with a BEMS score of 0-16, 14% with a BEMS score of 17-23, 23% with a BEMS score of 24-32, 34% with a BEMS score of 33-43 points, and 50% with a BEMS score of 44-79 points. A BEMS score ≥ 54

corresponded to a 60% cumulative incidence of epilepsy. The AUC was 0.70 for BEMS regarding the clinical diagnosis of epilepsy. The clinical EEG classification of focal IEDs based on the previous visual EEG interpretation, the reference standard, had a sensitivity of 52% and a specificity of 95% for a follow-up diagnosis of epilepsy in the hospital records.

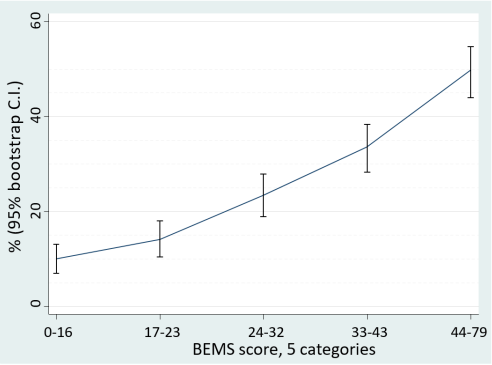

Figure 7. The cumulative incidence of a clinical diagnosis of epilepsy in the hospital records after 6 years of follow-up according to the BEMS score grouped into five quantiles.

The BEMS score had an AUC = 0.86 for the EEG outcome in the internal validation dataset (N=1013). A cut point at BEMS ≥ 46 had a sensitivity of 55% and a specificity of 91%. A cut point at BEMS ≥ 29 had a sensitivity of 90% and a specificity of 57%. BEMS had an AUC of 0.80 (N=100) in the external validation dataset. A cut point for BEMS at 53 gave a sensitivity of 41%, specificity of 91%, and an accuracy of 64%. A cut point at 29 gave a sensitivity of 98%, specificity of 50% and accuracy of 76%.

Inter-rater and inter-method reliability

Each of the quantified morphological IED features had an Intraclass correlation coefficient (ICC) > 0.69 between the two raters. BEMS had ICC = 0.91 (95%CI 0.88-0.93), while visual evaluation of transients on a Halford's 5-point likert scale gave a

Cohen's kappa of 0.43 (95%CI 0.15-0.32), which is regarded as moderate agreement [96]. The ICC for BEMS when comparing the automated one-click algorithm with the manual 4-click annotation method was 0.84 (95%CI 0.81-0.87).

Paper III

The mean IED candidate rate was estimated for each rater by a poisson model for all included EEGs. The mean IED candidate rate had a wide range between the three raters, ranging from 0.1 to 0.4 per minute (N=383 for rater 1 and 2, and N=187 for rater 3). This corresponds to 2-8 IED candidates in a 20-minute EEG. The estimated mean IED candidate rate was 0.35 per minute for rater 1, 0.10 per minute for rater 2, and 0.14 per minute for rater 3.

IED candidate count and IED morphology, as measured by $BEMS_{max}$ and $BEMS_{mean}$, had a positive correlation in all the three datasets (Figure 8). $BEMS_{max}$ and IED candidate count had a correlation coefficient (CC) in the range of 0.61-0.67 for the three raters in all three datasets. $BEMS_{mean}$ and IED candidate count had a CC in the range of 0.37-0.42 for the three raters in the internal datasets (N=383 for rater 1 and 2, and N=187 for rater 3).

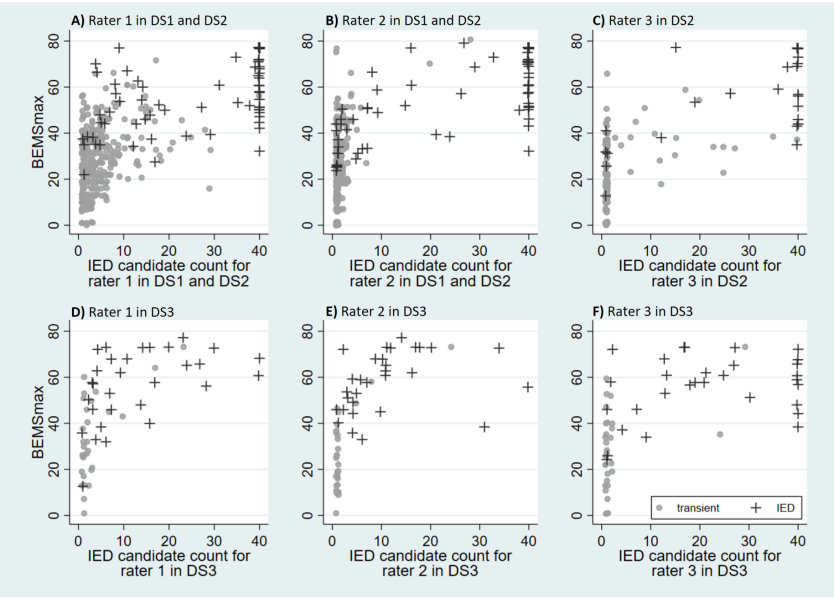

Figure 8 (Paper III). Scatter plots with IED candidate count and BEMSmax for the three raters.

Optimal cut-points in the training dataset

Optimal rater-specific cut-points were first established for each of the diagnostic markers for rater 1 and 2, before identifying the common cut-points that would later be used for classifying transients as epileptiform or non-epileptiform. The optimal cut-points for rater 1 in the training dataset with regards to the univariate classification of transients as IEDs or non-epileptiform transients were 52 for $BEMS_{max}$, 618 for $BEMS_{sum}$, and 28 for IED candidate count. The optimal cut-points for rater 2 were 48 for $BEMS_{max}$, 312 for $BEMS_{sum}$, and 7 for IED candidate count. The common diagnostic marker cut-points, or the "common ground" for the two raters, were 50 for $BEMS_{max}$, 465 for $BEMS_{sum}$, and 18 for IED candidate count. These cut-points were then applied for all three raters to assess diagnostic performance in the training dataset (rater 1 and 2) and in the validation dataset (rater 1, 2 and 3). The best combination of all three criteria-sets for the Diagnostic

classifier was as follows: Either one IED candidate with BEMS above 57, two IED candidates with BEMS above 46, or seven IED candidates with BEMS above 35.

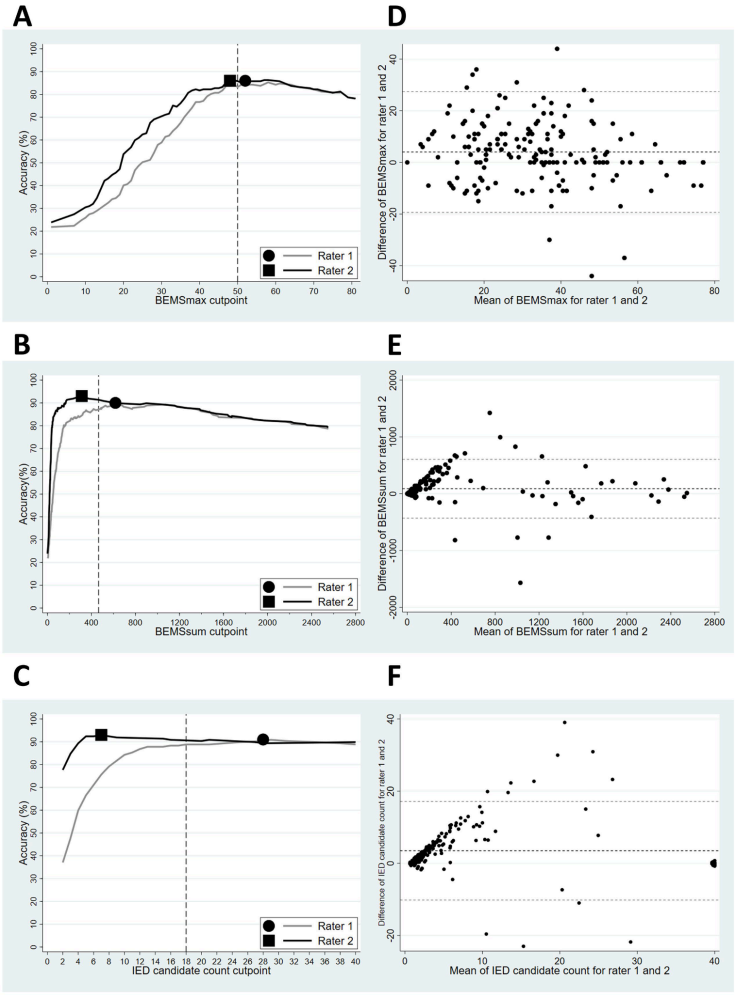

Figure 9 (Paper III). Diagnostic accuracy for the EEG conclusion as outcome for rater 1 and 2 in the training dataset is shown on the y-axes with the range of possible cut points of $BEMS_{max}$ **(A)**, $BEMS_{sum}$ **(B)** and IED candidate count **(C)** shown on the x-axes. Bland-Altman plots for $BEMS_{max}$ **(D)**, $BEMS_{sum}$ **(E)** and IED candidate count **(F)** for rater 1 and 2 in the training dataset.

Diagnostic performance in the internal validation dataset

Our reference standard, the previous clinical EEG conclusion, had a sensitivity of 41% and a specificity of 97% for the clinical outcome (a follow-up clinical diagnosis of epilepsy at the hospital). The mean diagnostic performance measures for all the three raters in the internal validation set for $BEMS_{max}$ gave a sensitivity of 60% and a specificity of 96% for the EEG-outcome, and a sensitivity of 29% and a specificity of 95% for the clinical outcome. For $BEMS_{sum}$, the sensitivity was 67% and the specificity was 98% for the EEG-outcome, whereas the sensitivity was 32% and the specificity was 98% for the clinical outcome. For IED candidate count, the sensitivity was 60% and the specificity was 98% for the EEG-outcome, and the sensitivity was 29% and the specificity was 97% for the clinical outcome. For the Diagnostic classifier the sensitivity was 60% and the specificity was 99% for the EEG outcome, and the sensitivity was 29% and the specificity was 97% for the clinical outcome.

Diagnostic performance in the external validation dataset

The mean diagnostic performance, calculated as sensitivity and specificity, for the three raters in the external validation set for the clinical outcome determined by the presence of habitual non-epileptic or epileptic seizures during long-term EEG monitoring was as follows: $BEMS_{max}$ had a sensitivity of 70% and specificity of 86%. $BEMS_{sum}$ had a sensitivity of 56% and a specificity of 94%. IED candidate count had a sensitivity of 33% and specificity of 96%. The Diagnostic classifier had a sensitivity of 63% and specificity of 91%.

Inter-rater reliability

Gwet's AC1 for three raters had a range of 0.90-0.96 for the diagnostic markers in the internal validation dataset, and had a range of 0.57-0.73 in the external validation dataset. The Diagnostic classifier had the highest IRR with Gwet's AC1= 0.96 and Gwet's AC1= 0.73 in the internal and external validation datasets, respectively.

Discussion

Paper I

We have shown in Paper I that focal IED morphology depends on age in an unselected dataset. The large size of our dataset provides solid evidence to support this conclusion. This represents a novel finding as IED morphology has not been systematically examined previously. To our knowledge, only the study by Aurlien et al. [33] has reported a comparable analysis, but for generalized and not focal IEDs. That study showed that the amplitude of generalized IEDs change with age. Some studies have examined the influence of age on IED localization in specific epilepsy types or epilepsy syndromes [97-101].

Focal IEDs were more common in children and in the elderly. With increasing age, focal IEDs became blunted, had lower slow after-wave area, lower spike amplitudes, and became more lateralized. Spike asymmetry was the only morphological IED measure that did not depend on age, suggesting that the corresponding criterion number three regarding asymmetry in the definition of epileptiform activity [7] should apply evenly for all ages. Spike detection, whether by human EEG readers or automatic spike detectors, should benefit from our findings because we clearly demonstrate that the interpretation of morphological IED features should depend on age. An IED with an ascending slope of 1 μV/ms would be a rarity in the age group 10-19 years, but a common event well within the interquartile range in the age group 80-101 years (Figure 10). The thresholds for quantitative IED features classification should differ according to age.

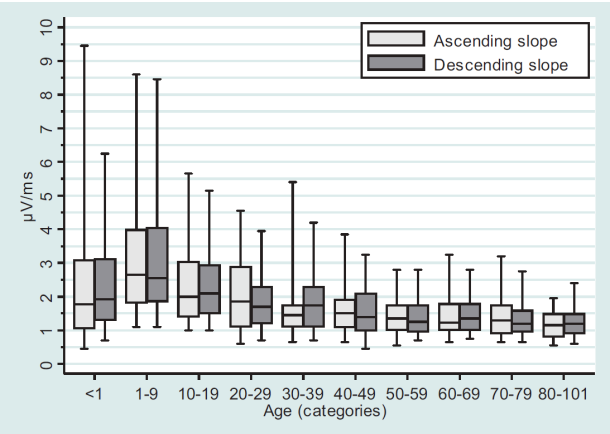

Figure 10 (Paper I)**.** The median, interquartile range, 5th percentile and 95th percentile for the ascending and descending slope of IEDs calculated as µVolt/ms (N = 868).

We demonstrated that the occurrence of focal and generalized epilepsy depended on age in a large dataset of routine EEGs. Focal epilepsy was the most common diagnostic conclusion regarding epilepsy type in the EEG report for all age groups. Focal epilepsy had a bimodal distribution by age, with the highest occurrence in children and in the elderly. Generalized epilepsy as the diagnostic conclusion in the EEG report occurred most frequently in the age group 10-19 years. These findings coincide with the incidence of epilepsy by age as described in etiological studies [16, 102-104], and shows that the patient population in our dataset, that were referred from a wide range of specialists and general practitioners, reflects what can be expected in the general patient population.

Paper II

We showed that carefully selected quantified morphological features of the first occurring IED candidate in an EEG can be combined into a BEMS score that classifies EEGs as epileptiform or non-epileptiform with high IRR. However the BEMS score had a lower sensitivity (38%) and specificity (86%) as compared to the clinical EEG conclusion from visual interpretation and information in the referral note (sensitivity 52% and specificity

95%) for the outcome of epilepsy in the hospital record. The context for IRR in Paper II was that two raters independently marked an IED candidate that was previously selected during clinical routine EEG analysis. This meant that the selection and marking of the IED candidate was limited to a short time segment, usually a few seconds, as opposed to a complete EEG recording. There were EEGs where the preexisting marking contained several IED candidates, i.e. in trains or in various EEG channels. This is most probably the main reason for a non-perfect IRR, since the algorithm has a deterministic nature and should compute exactly the same measures for a mouse click on a specific spike peak. It is important to examine IRR also on the IED candidate level, when the IRR is not influenced to a large degree by IED candidate selection. Knowing the IRR on such a detailed IED candidate level is necessary for correct assessment of IRR on the EEG level like we did in Paper III. In Paper III, IRR was calculated for aggregated morphological measures from the collection of many IED candidates in an EEG, like $BEMS_{max}$ and $BEMS_{sum}$. Findings in previous studies that examined IRR both on IED and EEG level indicate that IRR is lower for individual IEDs than for complete EEG recordings [67, 105-108]. We obtained a high IRR for single IED candidate measurements in Paper II. This implies that a low IRR for the EEGs in Paper III would very likely stem from IED candidate selection.

The BEMS score included four morphological IED features that corresponded well to three of the criteria in the definition of epileptiform activity [7]. Spike onset slope and descending spike amplitude both correspond to criterion 1 (pointed peak). The measure "spike to background power" illustrates how much relative power of the background activity is within the frequency band corresponding to the duration of the spike component of the IED capturing the essence of criterion 2 (different wave duration than the waves from the background activity). The area of the slow after-wave corresponds well to criterion 4 (which reads the transient is followed by a slow after-wave). We did not include the measure for spike asymmetry (criterion 3) in the BEMS because it did not differ significantly between epileptiform and non-epileptiform transients, and it did not

contribute to the multivariate classification model. This was unexpected, since asymmetry is included as a feature in the generally accepted classification criteria. Findings from our Paper 1 suggested that asymmetry was a relevant feature in all age groups. A paper by Jing et al. [67] found that asymmetric IED candidates were more likely to be classified as IEDs in visual spike analysis. We did not include appropriate measures for criterion 5 (disrupted background activity) or criterion 6 (a transient dipole suggesting a brain source). Quantitative measures for spike asymmetry, disrupted background activity and a transient dipole may be considered in future attempts to build an optimized classification model from quantitative IED features.

The annotation algorithm, EpiOneClick, was programmed in MATLAB to be used in conjunction with EEGLAB, and is publicly available for download [12, 93, 109]. This was an effective tool for us in marking and quantifying IED candidates. It required only one mouse click near the spike peak to precisely determine the relevant timestamps (spike start and end, slow after-wave start and end) and calculate the quantitative morphological features of an IED candidate in milliseconds. It then presented the IED candidate in a figure on screen so that the EEG reader could verify that the correct EEG signal had been quantified.

Paper III

In Paper III, we showed that IED candidate morphology is positively correlated to IED candidate count. EEGs that contained IED candidates with prominent epileptiform features, defined as a high BEMS score, had a high IED candidate count. IED candidates with less typical epileptiform features, measured as a low BEMS score, had a low IED candidate count. This represents a novel finding. Even though both the BEMS score and IED candidate count were shown to be relevant in classification of epileptiform EEGs, the consequences of their correlation for visual interpretation of IED candidates need to be further explored. An intuitive explanation might be that an EEG reader is more likely to detect numerous IED candidates if they are distinct and easily distinguishable from the

background activity. This would make $BEMS_{max}$ a confounder variable. The IED candidate count classified EEGs with an accuracy comparable to the other univariate diagnostic markers in our internal validation dataset, but had a considerable drop in sensitivity when applied to the external validation dataset. Our findings show that IED candidate count is relevant in IED classification and adds valuable information regarding the clinical diagnosis of epilepsy. The large threshold difference between individual raters for IED candidate count as opposed to the morphological measure $BEMS_{max}$ should be considered in future attempts of applying them in a classification model. $BEMS_{max}$ reflects the BEMS score of only one IED candidate, the one with the highest morphological BEMS score among all marked IED candidates. In contrast all IED candidates contribute to the IED candidate count. The count will therefore probably be more susceptible to low IRR [67]. A better defined threshold for when a transient should be considered an IED candidate is necessary to reduce the threshold difference between individual raters.

Selection bias

The focal spikes included for examination in our materials from Paper I-II-III were selected by clinical neurophysiologists. In Paper I-II, we analyzed one independent spike per EEG that had been marked previously during routine clinical EEG interpretation. This was not necessarily the first transient that occurred in the EEG, but it was the first transient that had been selected and marked by the clinical EEG interpreter based on visual examination. It was likely the first transient that raised a concern to the EEG reader about IEDs being present. A better methodical choice for Paper II might have been to include the transients with the most typical epileptiform characteristics, with the maximum amplitude, or with some other predefined relevant measurements. However, we did not re-mark several transients in each EEG for Paper I-II as we deemed it too time consuming for what we regarded as a relatively modest benefit. For Paper III, three blinded independent raters marked up to 40 consecutive spikes per EEG. We set the upper limit at 40 IED candidates to reduce the workload, assuming that 40 IED candidates were

well above the threshold for when an EEG would be classified as definite epileptiform. We found no relevant information in the literature that could assist in deciding this counting threshold. The primary goal for Paper III was to examine IED candidate count. That required an extensive marking of IED candidates, which is not commonly done during routine clinical EEG reviews. Having included several IED candidates per EEG allowed us to calculate BEMS for each individual IED candidate. We therefore decided to use $BEMS_{max}$, that is the BEMS with the highest score among many IED candidates from the same EEG, instead of $BEMS_{first}$, the first IED candidate that was marked, as the morphological component of the prediction model. $BEMS_{max}$ had a higher diagnostic performance than $BEMS_{first}$ in the training dataset of Paper III. We considered that the most epileptiform discharge ($BEMS_{max}$) was optimally suited for the epileptiform classification of EEGs. However, a limitation of the BEMS-score is that it was originally built on the first occurring IED candidate in the EEGs (Paper 2), while $BEMS_{max}$ might have been a better choice. Like many similar papers, we focused on focal suspicious sharp activity because we considered generalized sharp suspicious activity to be a different clinical diagnostic problem.

We chose to mark one or more IED candidates per EEG for Paper III. This proved to raise a challenge because included EEGs did not necessarily contain any IED candidates. In those cases where no convincing IED candidates were detected, we marked a background wave from the last EEG page. This represents a source of selection bias because we did not specify any formal rules for such a selection before the study started. For example, horizontal eye movements can produce high amplitude IED-like transients in the EEG because of the combined EMG-activity from the abductor muscle (spike) and eye movement artifacts (slow-waves). Marking such an artifact transient would result in a high BEMS score, while marking a negative peak from the posterior dominant rhythm would not. A better solution might have been to allow for no markings in EEGs with no IED candidates and then set the morphological score equal to zero in these EEGs that did not contain any IED candidates.

Our dependence on manually selected spikes means that any application of the classification models in Paper II-III would require an EEG reader to mark suspected IEDs as described in our methods. We did not compare features of IEDs to those of background waves which would have been a prerequisite for automatic spike detection. The optimal morphological features and their thresholds for classifying background waves into epileptiform or non-epileptiform activity would differ from classifying IED-candidates into epileptiform or non-epileptiform. Figure 11 illustrates how classification depends on whether the waveforms are selected by a neurophysiologist or if they are unselected EEG waveforms. This is an example of selection bias. Selection bias is in fact present at all steps, starting already at the patient's first visit to his or her treating physician, who decides to refer the patient to an EEG or not, this depending on the interpretation of the presenting symptoms. Next is the selection of which EEGs to include in our study. We sought relevance for the daily clinical routine situation by including all routine-EEGs recorded at our department. We excluded patients who had their EEG at the ICU since that patient population is highly different, as discussed in the methods section previously.

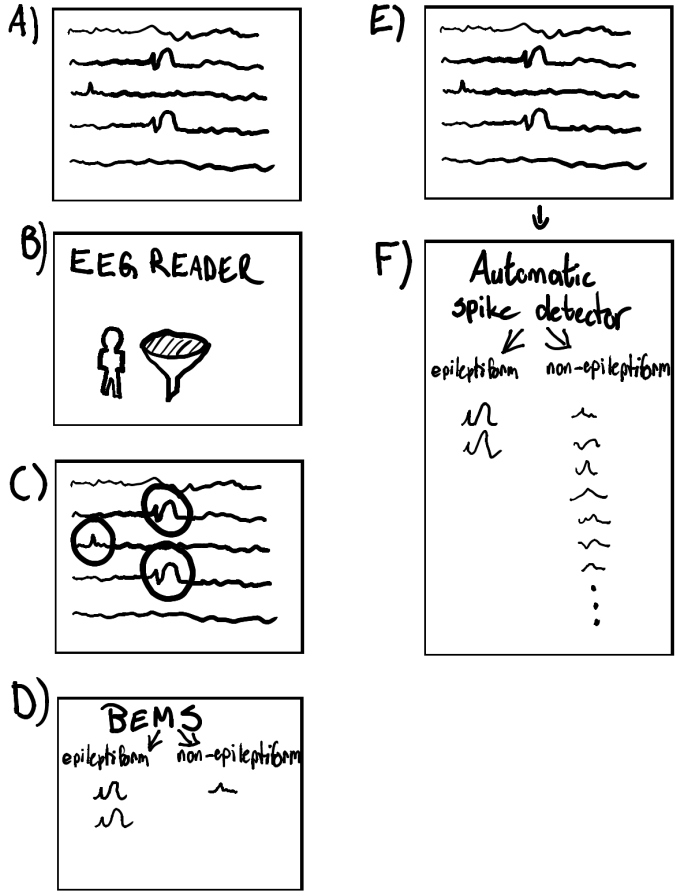

Figure 11. The quantification and classification of IED candidates depend on whether the waveforms are selected by a neurophysiologist (left column, A-D) or if they are unselected (right column, E-F). **A:** EEG background activity containing various epileptiform and non-epileptiform transients. **B**: A neurophysiologist reads the EEG and marks all IED candidates (waveforms that are suspected to represent IEDs). **C:** Only those IED candidates that were selected by the neurophysiologist serve as input to the classification model. **D:** BEMS is calculated for the selected IED candidates to classify them as either epileptiform or non-epileptiform. **E:** EEG background activity as in A. **F:** The automatic spike detector analyzes all EEG waveforms. Classification of all EEG waveforms into either epileptiform or non-epileptiform probably requires analysis of other morphological characteristics and have other thresholds than when classifying only selected IED candidates.

The threshold for when a transient should be marked as an IED candidate was decided before annotation of the EEGs for our Paper III. We decided to mark transients as IED candidates when they were "likely to represent an epileptiform charge", i.e. corresponding to 4 points on a 5-point Likert scale as described in a paper by Bagheri et al. [66]. Three points on this scale would correspond to "not sure if the transient is an IED, could go either way", which we found to be too inclusive with a loss of specificity, while 5 points on this scale would correspond to "definitely an epileptiform discharge", which we decided would lack sensitivity. Despite the differences in the individual rater thresholds for IED candidates, the IRR for the combined IED candidate count and BEMS into three criteria-sets was good. The high IRR might have resulted from the contribution of BEMS to the criteria-sets, functioning as a morphological filter where only those IED candidates that have a high enough BEMS score were counted. Since we suspected that a large proportion of both clearly pathological EEGs and normal EEGs could explain this high IRR, we performed a sensitivity analysis with regards to IRR for "difficult" EEGs that did not have the maximum IED candidate count of 40 or the minimum of 1 (not published). IRR was assessed for the Diagnostic classifier for two raters in the combined training and validation datasets. This selected group included 271 "difficult" EEGs out of the total of 394 EEGs. Gwet's AC was 0.91 as compared to 0.93 for the complete dataset. This illustrates that the IRR is robust also for EEGs that are more demanding to evaluate. Furthermore, this assessment shows that all types of EEGs contribute to the IRR.

IED features missing in our project

We covered only one domain of IED characteristics in Paper II, and that was spike morphology in one EEG-channel. We did not include any information from the temporal or spatial domains. The temporal domain was included in the variable IED candidate count in Paper III. We did not succeed in finding a relevant spatial measure for spikes that might have improved our prediction model. However, we made an attempt to quantify IED candidate topography to assess if topography might be a viable additional predictor

in our model (Figure 12). This study was discontinued after it did not perform well in a univariate analysis using the visually marked electrodes from clinical EEG interpretation as the gold standard in the Paper 2 training dataset (N=1013). Several quantitative methods were applied to identify EEG electrodes in which the IED candidate could be detected. Most of the methods were based on finding the subset of electrodes that were outliers regarding voltage around the spike peak (Grubb's outliers [110], median absolute deviation, standard deviation or the generalized extreme studentized deviate [111]). We also tried to identify IED candidate topography by assessing morphological similarity to the clicked spike peak using dynamic time warping [112] or our own quantitative IED features from Paper II. The best performance achieved by applying these methods was a sensitivity of 82% and a specificity of 91% when using visually marked electrodes during routine clinical EEG interpretation as the reference standard. The mean number of visually marked electrodes in the EEGs was three, and the total number of electrodes in the montage ranged from 21 to 26. Assuming that there were 25 electrodes in all EEG montages for the sake of simplification, the automatic detection of electrodes missed roughly 0.5 electrodes on average that had been visually marked, and it included 2 electrodes on average that had not been visually marked. AUC ranged from 0.36 to 0.61 for the applied methods using the EEG conclusion of focal IED or not as the outcome. We found that these results regarding topography had too little value for further analysis. In our experience, the topography of a transient is an essential characteristic to rule out artifact waveforms or common physiological transients in visual analysis. A possible explanation for the lack of diagnostic value of an automated assessment could be that certain topographical characteristics have already been evaluated for all the IED candidates selected by an EEG reader, and therefore have little value for further classification of these IED candidates as either IEDs or non-epileptiform transients. Fábio A. Nascimento et al. [113] managed to quantify all six criteria listed in the definition of epileptiform discharges [7], covering morphological features (criterion 1-5) as well as topography or physiological field (criterion 6).

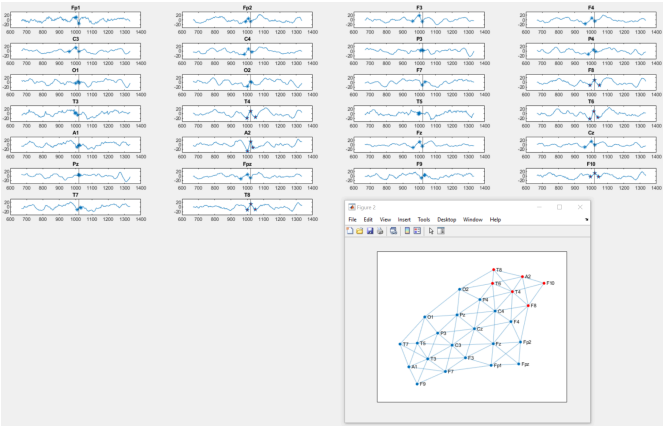

Figure 12. Quantification of topographical properties of an IED candidate. The EEG is shown as individual channels. The lower right window shows the head model as an undirected graph where the IED candidate topography is indicated by red dots.

EEG background activity

The background activity is most probably as important as the IED itself in spike detection. A core principle of epileptiform discharges is that they are distinct from the background activity [7]. Figure 13 is a mock illustration of how the exact same spike can become nearly invisible when surrounded by a higher amplitude background. We did include the power of the background activity immediately preceding the IED candidate as an independent variable in our preliminary logistic regression model for Paper II, but it was excluded during the process of stepwise elimination of non-significant predictor variables. The interindividual variation of EEG background activity is considerable [114], so lacking a variable for general background amplitude or power might be a weakness of our classification model and in defining BEMS. There is also intra-EEG background variability in various power frequency bands between patient states (awakeness and sleep stages) [115], and due to physiological transients [6] and reactive rhythms [116]. Our model did, however, include the variable “spike to background power”, which is a ratio for how much of the background activity power is within the frequency band

corresponding to the spiky component of the IED. This variable represents a fitting measure regarding IED criterion 2 (wave duration differs from the background), but it does not cover the frequency bands outside that of the spiky component, and may therefore be regarded as a limited assessment of the background activity. The age variable was included in the classification model since we showed that IED morphology depended on age in Paper I, and also because we showed that age was significant in the process of building the model. Some studies have demonstrated that the general background amplitude changes with age [35, 117].

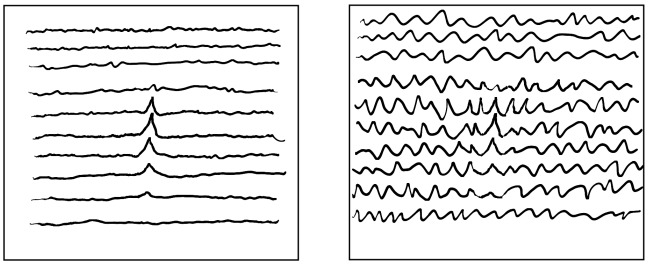

Figure 13. Left window: The transient is clearly distinguished from the background activity. Right window: The identical transient is not clearly distinguished from a different background activity.

Dependent variables and confounders

We established our prediction model by applying BEMS in Paper II, adhering as strictly as possible to sound scientific methods. However, we did not assume that the morphological features of IEDs are strictly independent. “What goes up must come down” applies to the ascending and descending part of the spike, and most probably to other morphological features as well. We tried to minimize this weakness of the model by applying a mix of well-accepted statistical concepts; keeping the number of predictor variables low, choosing variables not depending too much on each other, as well as applying a conservative approach for validation by splitting the datasets half and half for

training and validation. Age was included as a predictor variable in our model, but also represents a confounder as morphology (predictor) depends on age, as does the incidence of epilepsy (outcome) [103, 118]. Excluding the age variable as a predictor variable would on the other hand have been an example of omitted-variable bias, and it was therefore not done.

Categorical prediction model (BEMS)

We converted our multivariate logistic regression from several continuous numericals to the categorical model BEMS. This was done to make the model more accessible and easy to interpret, at the cost of losing statistical power because of the increased degrees of freedom.

Outcome variables

The classification models in Paper II-III used two outcomes; the diagnostic conclusion of focal epilepsy or non-epilepsy in the EEG report (EEG outcome) and the follow-up diagnosis of epilepsy (clinical outcome) according to the hospital records at least two years after the EEG recording. We built our models in Paper II-III using the EEG outcome as a reference standard because the clinical outcome was not available until the later stages of preparing Paper II. One could argue that the clinical outcome comes closer to the real outcome and would be the best target for model building. The clinical outcome was in part dependent on the result of the clinical EEG conclusion, which is a weakness of Paper II and III. The treating clinician would receive the EEG report, and if positive for IEDs, would be more likely to give the patient a diagnosis of epilepsy. An argument for using the EEG outcome as a primary outcome can be made as patients with epilepsy can have normal EEGs, and patients who do not develop epilepsy can still have IEDs in their EEG [119]. Ideally, the outcome should be independent of the diagnostic test. Theoretically we might have chosen for example the event of a recurring epileptic seizure during a 2 year follow-up period as an independent outcome.

Clinical and laboratory data

Our project did not include detailed clinical and other laboratory data, e.g. imaging data, use of anti-seizure medication, recurrent seizures, and IED candidate topography and localization. Imaging data (anatomical lesion location) has implications for IED morphology [120] and might have been a useful control variable for the regression model in Paper I, where we examined which factors had the strongest effect on IED morphology. It is reasonable to assume that anti-seizure medication can have an effect on IED morphology [121], but this has not yet been examined systematically. A follow-up assessment of recurrent epileptic seizures would have been a stronger outcome variable for Paper II-III, since it would have been independent of the clinical EEG conclusion. IED candidate topography and brain localization are core properties of IED candidates. Information about this variable might have improved the diagnostic accuracy of our classification models in Paper II-III.

Why we did not apply artificial intelligence

Artificial intelligence based classification of EEG abnormalities, including IEDs, has shown promising results recently [122]. However, the goals of our project were not limited to maximizing diagnostic performance, but also to improve the understanding of IED analysis, while at the same time ensuring transparent, robust and generalizable methods. Statistical regression models and hand-picked morphological features were in our opinion the appropriate analytical method to meet these goals [123]. This allowed us to identify reliable quantitative morphological features of sharp transients that were then applied to show morphological differences across age groups and in the classification of EEGs as either being epileptiform or non-epileptiform when applying the BEMS-score. Every step of the analytical process, from the morphological measurements to the construction and application of the classification model, is transparent and open for scrutiny. Although the diagnostic accuracy of BEMS and IED candidate count did not exceed that of visual IED analysis, generalizability was confirmed by validation on independent external datasets. We had no insight into the separate, independent and

closed commercial-focused work performed by Tveit [122], though some of the EEG data in our studies is likely present in some of Tveit's work though their description is not sufficient to estimate how much.

Conclusions

1. Morphological features of IED can be quantified with a high interrater reliability.

2. Quantified IED features depend on age. IEDs become blunter, wider, shorter, and have a smaller slow after-wave with increasing age.

3. Quantified IED features can classify EEGs as epileptiform or non-epileptiform with a high IRR, but with lower diagnostic accuracy than clinical visual analysis.

4. IED candidate count correlates with IED morphology as measured by BEMS. A higher IED candidate count was associated with a higher BEMS score.

5. IED candidate count represents a relevant variable in the classification of EEGs as epileptiform or non-epileptiform. Univariate classification of epileptiform EEGs in an external validation dataset had a sensitivity of 33% and specificity of 96%.

Future perspectives

A primary goal in this project was to develop a quantitative morphological score for focal IED candidates which relates to the standard visual evaluation of EEGs. The resulting score, BEMS, combined several morphological features (ascending slope, descending

amplitude, slow-wave area, spike to background, and age). We have validated BEMS applying internal and external data with regard to the clinical EEG conclusion and the diagnosis of epilepsy. The ideal study design to validate the BEMS score would be a prospective randomized controlled trial where the patients are assigned to either the standard workup with visual classification of EEGs as epileptiform or non-epileptiform, or to BEMS-assisted visual EEG analysis where the BEMS-score is used to classify the EEGs. A follow-up after 10 or more years regarding the presence or absence of recurrent, unprovoked epileptic seizures would be an excellent outcome measure to compare the predictive value of the two alternative methods for EEG evaluation.

An advantage of BEMS is that it places IED candidate morphology along a numerical scale where a higher BEMS score indicates a greater risk of receiving an epilepsy diagnosis according to our findings in Paper II. Such a score can easily be combined with other variables, e.g. a seizure semiology score and an MRI lesion score, into a composite risk score regarding recurring seizures after a single seizure. The risk assessment after a single epileptic seizure is subjective because of limited evidence in the literature, and with no established risk formula for the combined clinical and laboratory data that are used to predict this risk [124]. There is an urgent clinical need to predict this risk as accurately as possible because it guides decisions regarding anti-seizure drug therapy, non-drug precautions, and life-style changes.

We have demonstrated that quantitative IED features were excellent measures to identify morphological differences between age groups in Paper I. There is a potential role for quantitative IED features, BEMS and IED candidate count to help differentiate between other patient groups as well. We would expect differences in quantitative IED features and IED candidate counts between various epilepsy syndromes, various intracranial lesion localizations, and antiepileptic drugs. The measurements might be useful for assessing seizure control. A precise assessment of such differences might help optimizing therapy to prevent new seizures.

References:

1. Gibbs, F.A., H. Davis, and W.G. Lennox, *The electro-encephalogram in epilepsy and in conditions of impaired consciousness.* Archives of Neurology & Psychiatry, 1935. **34**(6): p. 1133-1148.
2. Williams, D., *The Nature of Transient Outbursts in the Electro-Encephalogram of Epileptics.* Brain, 1944. **67**(1): p. 10-37.
3. Lundervold, A., G. Henriksen, and L. Fegersten, *The spike and wave complex; a clinical correlation.* Electroencephalography and clinical neurophysiology, 1959. **11**(1): p. 13-22.
4. Benbadis, S.R. and P.W. Kaplan, *The Dangers of Over-Reading an EEG.* Journal of Clinical Neurophysiology, 2019. **36**(4): p. 249.
5. Nascimento, F.A., et al., *One EEG, one read - A manifesto towards reducing interrater variability among experts.* Clin Neurophysiol, 2021. **133**: p. 68-70.
6. Kang, J.Y. and G.L. Krauss, *Normal Variants Are Commonly Overread as Interictal Epileptiform Abnormalities.* Journal of Clinical Neurophysiology, 2019. **36**(4): p. 257-263.
7. Kane, N., 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**: p. 170-185.
8. Chatrian, G., et al., *A glossary of terms most commonly used by clinical electroencephalographers.* Electroencephalogr Clin Neurophysiol, 1974. **37**(5): p. 538-48.
9. Wüstenhagen, S., et al., *EEG normal variants: a prospective study using the SCORE system.* Clinical Neurophysiology Practice, 2022.
10. Holberg EEG AS, *SCORE EEG™ [software]. Available from* http://holbergeeg.com.
11. Beniczky, S., et al., *Standardized computer-based organized reporting of EEG: SCORE - Second version.* Clin Neurophysiol, 2017. **128**(11): p. 2334-2346.
12. MATLAB, *9.4.0.949201 (R2018a) Update 6*. 2018: The MathWorks Inc.
13. STATA, *StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.* 2017.
14. GitHub. *GitHub.com*. [Web page] [cited 2023; Available from: github.com.
15. Begley, C., et al., *The global cost of epilepsy: A systematic review and extrapolation.* Epilepsia, 2022. **63**(4): p. 892-903.
16. Thijs, R.D., et al., *Epilepsy in adults.* The Lancet, 2019. **393**(10172): p. 689-701.
17. Falco-Walter, J.J., I.E. Scheffer, and R.S. Fisher, *The new definition and classification of seizures and epilepsy.* Epilepsy Research, 2018. **139**: p. 73-79.
18. Blume, W.T., *The Progression of Epilepsy.* Epilepsia, 2006. **47**(s1): p. 71-78.
19. Shankar, R., et al., *Sudden unexpected death in epilepsy (SUDEP): what every neurologist should know.* Epileptic Disorders, 2017. **19**(1): p. 1-9.

20. Lv, R.-J., et al., *Status epilepticus-related etiology, incidence and mortality: A meta-analysis.* Epilepsy Research, 2017. **136**: p. 12-17.
21. de Boer, H.M., M. Mula, and J.W. Sander, *The global burden and stigma of epilepsy.* Epilepsy & Behavior, 2008. **12**(4): p. 540-546.
22. Fisher, R.S., et al., *ILAE Official Report: A practical clinical definition of epilepsy.* Epilepsia, 2014. **55**(4): p. 475-482.
23. Bast, T., et al., *Noninvasive Source Localization of Interictal EEG Spikes: Effects of Signal-to-Noise Ratio and Averaging.* Journal of Clinical Neurophysiology, 2006. **23**(6).
24. Oostenveld, R. and P. Praamstra, *The five percent electrode system for high-resolution EEG and ERP measurements.* Clinical Neurophysiology, 2001. **112**(4): p. 713-719.
25. Tao, J.X., et al., *Intracranial EEG Substrates of Scalp EEG Interictal Spikes.* Epilepsia, 2005. **46**(5): p. 669-676.
26. Peltola, M.E., et al., *Routine and sleep EEG: Minimum recording standards of the International Federation of Clinical Neurophysiology and the International League Against Epilepsy.* Clinical Neurophysiology, 2023. **147**: p. 108-120.
27. Brogger, J., et al., *Visual EEG reviewing times with SCORE EEG.* Clinical neurophysiology practice, 2018. **3**: p. 59-64.
28. Koutroumanidis, M., et al., *The role of EEG in the diagnosis and classification of the epilepsy syndromes: a tool for clinical practice by the ILAE Neurophysiology Task Force (Part 1).* Epileptic Disord, 2017. **19**(3): p. 233-298.
29. Alessi, N., P. Perucca, and A.M. McIntosh, *Missed, mistaken, stalled: Identifying components of delay to diagnosis in epilepsy.* Epilepsia, 2021.
30. Aanestad, E., N.E. Gilhus, and J. Brogger, *A New Score for Sharp Discharges in the EEG Predicts Epilepsy.* J Clin Neurophysiol, 2023. **40**(1): p. 9-16.
31. Beniczky, S., et al., *Standardized computer-based organized reporting of EEG: SCORE.* Epilepsia, 2013. **54**(6): p. 1112-24.
32. Noachtar, S., et al., *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**: p. 21-41.
33. Aurlien, H., et al., *Characteristics of generalised epileptiform activity.* Clin Neurophysiol, 2009. **120**(1): p. 3-10.
34. Aurlien, H., et al., *Focal epileptiform activity described by a large computerised EEG database.* Clin Neurophysiol, 2007. **118**(6): p. 1369-76.
35. Aurlien, H., et al., *EEG background activity described by a large computerized database.* Clin Neurophysiol, 2004. **115**(3): p. 665-73.
36. Tao, J.X., et al., *Cortical Substrates of Scalp EEG Epileptiform Discharges.* 2007. **24**(2): p. 96-100.
37. Ebus, S., et al., *Cognitive effects of interictal epileptiform discharges in children.* European Journal of Paediatric Neurology, 2012. **16**(6): p. 697-706.

38. Aldenkamp, A.P. and J. Arends, *Effects of epileptiform EEG discharges on cognitive function: Is the concept of "transient cognitive impairment" still valid?* Epilepsy & Behavior, 2004. **5**: p. 25-34.
39. Koutroumanidis, M., et al., *The role of EEG in the diagnosis and classification of the epilepsy syndromes: a tool for clinical practice by the ILAE Neurophysiology Task Force (Part 1).* Epileptic Disorders, 2017. **19**(3): p. 233-298.
40. Rodríguez, V., M.F. Rodden, and S.M. LaRoche, *Ictal–interictal continuum: A proposed treatment algorithm.* Clinical Neurophysiology, 2016. **127**(4): p. 2056-2064.
41. Berger, H., *Über das Elektrenkephalogramm des Menschen.* Archiv für Psychiatrie und Nervenkrankheiten, 1929. **87**(1): p. 527-570.
42. Walter, W.G., *Critical Review: The Technique and Application of Electro-Encephalography.* J Neurol Psychiatry, 1938. **1**(4): p. 359-85.
43. Erasmus, J.F.P., *Epilepsy : 1 : some physiological considerations*.* South African Medical Journal, 1949. **23**(36): p. 740-746.
44. Ralston, B.L., *The mechanism of transition of interictal spiking foci into ictal seizure discharges.* Electroencephalography and clinical neurophysiology, 1958. **10**(2): p. 217-232.
45. Sannit, T. and E. Lilienthal, *A Glossary for EEG Technicians.* American Journal of EEG Technology, 1962. **2**(4): p. 106-119.
46. Scheuer, M.L., A. Bagic, and S.B. Wilson, *Spike detection: Inter-reader agreement and a statistical Turing test on a large data set.* Clinical Neurophysiology, 2016. **128**: p. 243-250.
47. Tjepkema-Cloostermans, M.C., R.C.V. de Carvalho, and M. van Putten, *Deep learning for detection of focal epileptiform discharges from scalp EEG recordings.* Clin Neurophysiol, 2018. **129**(10): p. 2191-2196.
48. De Lucia, M., et al., *A novel method for automated classification of epileptiform activity in the human electroencephalogram-based on independent component analysis.* Medical & biological engineering & computing, 2008. **46**(3): p. 263-272.
49. Liu, Y.-C., et al., *Model-based spike detection of epileptic EEG data.* Sensors, 2013. **13**(9): p. 12536-12547.
50. İnan, Z.H. and M. Kuntalp, *A study on fuzzy C-means clustering-based systems in automatic spike detection.* Computers in biology and medicine, 2007. **37**(8): p. 1160-1166.
51. Acir, N., et al., *Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks.* IEEE Transactions on Biomedical Engineering, 2004. **52**(1): p. 30-40.
52. Flanagan, D., et al., *Improvement in the performance of automated spike detection using dipole source features for artefact rejection.* Clinical Neurophysiology, 2003. **114**(1): p. 38-49.
53. Black, M.A., et al., *Real-time detection of epileptiform activity in the EEG: a blinded clinical trial.* Clinical Electroencephalography, 2000. **31**(3): p. 122-130.

54. Wilson, S.B., et al., *Spike detection II: automatic, perception-based detection and clustering.* Clinical neurophysiology, 1999. **110**(3): p. 404-411.
55. James, C.J., et al., *Detection of epileptiform discharges in the EEG by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages.* Clinical Neurophysiology, 1999. **110**(12): p. 2049-2063.
56. Benlamri, R., et al., *An automated system for analysis and interpretation of epileptiform activity in the EEG.* Comput Biol Med, 1997. **27**(2): p. 129-39.
57. Webber, W., et al., *Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural network: a comparison of raw and parameterized EEG data.* Electroencephalography and clinical Neurophysiology, 1994. **91**(3): p. 194-204.
58. Pietilä, T., et al., *Evaluation of a computerized system for recognition of epileptic activity during long-term EEG recording.* Electroencephalography and clinical neurophysiology, 1994. **90**(6): p. 438-443.
59. Dingle, A.A., et al., *A multistage system to detect epileptiform activity in the EEG.* IEEE Transactions on Biomedical Engineering, 1993. **40**(12): p. 1260-1268.
60. Hostetler, W.E., H.J. Doller, and R.W. Homan, *Assessment of a computer program to detect epileptiform spikes.* Electroencephalogr Clin Neurophysiol, 1992. **83**(1): p. 1-11.
61. Gotman, J. and L.-Y. Wang, *State dependent spike detection: validation.* Electroencephalography and clinical neurophysiology, 1992. **83**(1): p. 12-18.
62. De Oliveira, P.G., C. Queiroz, and F.L. Da Silva, *Spike detection based on a pattern recognition approach using a microcomputer.* Electroencephalography and clinical neurophysiology, 1983. **56**(1): p. 97-103.
63. Gotman, J., J.R. Ives, and P. Gloor, *Automatic recognition of inter-ictal epileptic activity in prolonged EEG recordings.* Electroencephalography and Clinical Neurophysiology, 1979. **46**: p. 510-520.
64. Gotman, J. and P. Gloor, *Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG.* Electroencephalography and Clinical Neurophysiology, 1976. **41**: p. 513-529.
65. Chinappen, D.M., et al., *Spike height improves prediction of future seizure risk.* Clinical Neurophysiology, 2023.
66. Bagheri, E., et al., *Interictal epileptiform discharge characteristics underlying expert interrater agreement.* Clin Neurophysiol, 2017. **128**(10): p. 1994-2005.
67. Jing, J., et al., *Interrater Reliability of Experts in Identifying Interictal Epileptiform Discharges in Electroencephalograms.* JAMA Neurol, 2020. **77**(1): p. 49-57.
68. Terney, D., et al., *The slow-wave component of the interictal epileptiform EEG discharges.* Epilepsy Res, 2010. **90**(3): p. 228-33.
69. Seneviratne, U., M. Cook, and W. D'Souza, *Consistent topography and amplitude symmetry are more typical than morphology of epileptiform discharges in genetic generalized epilepsy.* Clinical Neurophysiology, 2016. **127**(2): p. 1138-1146.

70. Fürbass, F., et al., *An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: Validation against the diagnostic gold standard.* Clinical Neurophysiology, 2020.
71. Bagheri, E., et al., *A Fast Machine Learning Approach to Facilitate the Detection of Interictal Epileptiform Discharges in the Scalp Electroencephalogram.* Journal of Neuroscience Methods, 2019.
72. Sartoretto, F. and M. Ermani, *Automatic detection of epileptiform activity by single-level wavelet analysis.* Clinical Neurophysiology, 1999. **110**(2): p. 239-249.
73. Lodder, S.S., J. Askamp, and M.J. van Putten, *Inter-ictal spike detection using a database of smart templates.* Clin Neurophysiol, 2013. **124**(12): p. 2328-35.
74. Eusebi, P., *Diagnostic Accuracy Measures.* Cerebrovascular Diseases, 2013. **36**(4): p. 267-272.
75. Benbadis, S.R. and P. Thomas, *When EEG is bad for you.* Clinical Neurophysiology, 2017. **128**(4): p. 656-657.
76. Bouma, H.K., et al., *The diagnostic accuracy of routine electroencephalography after a first unprovoked seizure.* Eur J Neurol, 2016. **23**(3): p. 455-63.
77. Kural, M.A., et al., *Optimized set of criteria for defining interictal epileptiform EEG discharges.* Clinical Neurophysiology, 2020. **131**(9): p. 2250-2254.
78. Kural, M.A., et al., *The operational definition of epileptiform discharges significantly improves diagnostic accuracy and inter-rater agreement of trainees in EEG reading.* 2022. **24**(2): p. 353-358.
79. Obuchowski, N.A. and J. Bullen, *Multireader diagnostic accuracy imaging studies: fundamentals of design and analysis.* Radiology, 2022. **303**(1): p. 26-34.
80. Cohen, J.J.E. and p. measurement, *A coefficient of agreement for nominal scales.* 1960. **20**(1): p. 37-46.
81. Wongpakaran, N., et al., *A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples.* BMC Medical Research Methodology, 2013. **13**(1): p. 61.
82. McHugh, M.L., *Interrater reliability: the kappa statistic.* Biochem Med (Zagreb), 2012. **22**(3): p. 276-82.
83. Mani, R., et al., *Interrater reliability of ICU EEG research terminology.* J Clin Neurophysiol, 2012. **29**(3): p. 203-12.
84. Foldvary, N., et al., *The localizing value of ictal EEG in focal epilepsy.* Neurology, 2001. **57**(11): p. 2022-2028.
85. Mégevand, P. and M. Seeck, *Electroencephalography, magnetoencephalography and source localization: their value in epilepsy.* Current opinion in neurology, 2018. **31**(2): p. 176-183.
86. Hillebrand, A., et al., *Non-invasive measurements of ictal and interictal epileptiform activity using optically pumped magnetometers.* Scientific Reports, 2023. **13**(1): p. 4623.
87. Guth, T.A., et al., *Interictal spikes with and without high-frequency oscillation have different single-neuron correlates.* Brain, 2021. **144**(10): p. 3078-3088.

88. Matos, J., et al., *Diagnosis of Epilepsy with Functional Connectivity in EEG after a Suspected First Seizure.* Bioengineering, 2022. **9**(11): p. 690.
89. Kural, M.A., et al., *Criteria for defining interictal epileptiform discharges in EEG.* A clinical validation study, 2020: p. 10.1212/WNL.0000000000009439.
90. Aykut Kural, M., et al., *The influence of the abundance and morphology of epileptiform discharges on diagnostic accuracy: How many spikes you need to spot in an EEG.* Clinical Neurophysiology, 2021.
91. Halford, J.J., et al., *Characteristics of EEG Interpreters Associated With Higher Interrater Agreement.* J Clin Neurophysiol, 2017. **34**(2): p. 168-173.
92. Arends, J.B., et al., *Value of re-interpretation of controversial EEGs in a tertiary epilepsy clinic.* Clinical Neurophysiology, 2017. **128**(4): p. 661-666.
93. Aanestad, E. and J.C. Brogger, *EpiOneClick*. 2019, GitHub.
94. van Smeden, M., et al., *Sample size for binary logistic prediction models: Beyond events per variable criteria.* 2019. **28**(8): p. 2455-2474.
95. McKight, P.E. and J. Najab, *Kruskal-Wallis Test*, in *The Corsini Encyclopedia of Psychology*. p. 1-1.
96. Hallgren, K.A., *Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial.* Tutor Quant Methods Psychol, 2012. **8**(1): p. 23-34.
97. Hughes, J., *EEG epileptiform abnormalities at different ages.* Epilepsia, 1967. **8**(2): p. 93-104.
98. Koufen, H. and C. Gast, *Zur Frage der Alters- und Diagnoseabhängigkeit der Links-Lateralisation und Lokalisation von EEG-Herden.* Arch Psychiatr Nervenkr (1970), 1981. **229**(3): p. 227-37.
99. Konishi, T., et al., *Changes in EEG foci with age in childhood partial epilepsies.* Clin Electroencephalogr, 1994. **25**(3): p. 104-9.
100. Sadleir, L.G., et al., *EEG features of absence seizures in idiopathic generalized epilepsy: Impact of syndrome, age, and state.* Epilepsia, 2009. **50**(6): p. 1572-1578.
101. Lee, I.-C., Y.-J. Chen, and H.-S. Lee, *Migration of epileptic spike foci in encephalograms may correlate with a better outcome in pediatric epilepsy.* Brain and Development, 2010. **32**(10): p. 821-828.
102. Beghi, E. and G. Giussani, *Aging and the Epidemiology of Epilepsy.* Neuroepidemiology, 2018. **51**(3-4): p. 216-223.
103. Fiest, K.M., et al., *Prevalence and incidence of epilepsy.* A systematic review and meta-analysis of international studies, 2017. **88**(3): p. 296-303.
104. Nicolson, A., D.W. Chadwick, and D.F. Smith, *A comparison of adult onset and "classical" idiopathic generalised epilepsy.* Journal of Neurology, Neurosurgery & Psychiatry, 2004. **75**(1): p. 72-74.
105. Gotman, J., P. Gloor, and N. Schaul, *Comparison of traditional reading of the EEG and automatic recognition of interictal epileptic activity.* Electroencephalography and clinical Neurophysiology, 1978. **44**(1): p. 48-60.

106. Gotman, J., *Relationships Between Interictal Spiking and Seizures: Human and Experimental Evidence.* Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques, 1991. **18**(S4): p. 573-576.
107. Houfek, E.E. and R.J. Ellingson, *On the reliability of clinical EEG interpretation.* The Journal of nervous and mental disease, 1959. **128**(5): p. 425-437.
108. Webber, W.R., et al., *Automatic EEG spike detection: what should the computer imitate?* Electroencephalogr Clin Neurophysiol, 1993. **87**(6): p. 364-73.
109. Delorme, A. and S. Makeig, *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.* Journal of neuroscience methods, 2004. **134**(1): p. 9-21.
110. Stefansky, W., *Rejecting Outliers by Maximum Normed Residual.* The Annals of Mathematical Statistics, 1971. **42**(1): p. 35-45.
111. Fitrianto, A., et al., *Comparing outlier detection methods using boxplot generalized extreme studentized deviate and sequential fences.* Aceh International Journal of Science and Technology, 2022. **11**(1): p. 38-45.
112. Jing, J., et al., *Rapid annotation of interictal epileptiform discharges via template matching under Dynamic Time Warping.* J Neurosci Methods, 2016. **274**: p. 179-190.
113. Nascimento, F.A., et al., *A quantitative approach to evaluating interictal epileptiform discharges based on interpretable quantitative criteria.* Clinical Neurophysiology, 2022.
114. Smit, D.J.A., et al., *Heritability of background EEG across the power spectrum.* Psychophysiology, 2005. **42**(6): p. 691-697.
115. Šušmáková, K. and A. Krakovská, *Discrimination ability of individual measures used in sleep stages classification.* Artificial Intelligence in Medicine, 2008. **44**(3): p. 261-277.
116. Schomer, D.L., *The normal EEG in an adult.* The clinical neurophysiology primer, 2007: p. 57-71.
117. Dustman, R.E., D.E. Shearer, and R.Y. Emmerson, *Life-span changes in EEG spectral amplitude, amplitude variability and mean frequency.* Clinical Neurophysiology, 1999. **110**(8): p. 1399-1409.
118. Aanestad, E., N.E. Gilhus, and J. Brogger, *Interictal epileptiform discharges vary across age groups.* Clin Neurophysiol, 2020. **131**(1): p. 25-33.
119. Noachtar, S. and J. Rémi, *The role of EEG in epilepsy: A critical review.* Epilepsy & Behavior, 2009. **15**(1): p. 22-33.
120. Cuello-Oderiz, C., et al., *Influence of the location and type of epileptogenic lesion on scalp interictal epileptiform discharges and high-frequency oscillations.* Epilepsia, 2017.
121. Larsson, P.G., et al., *Decrease in propagation of interictal epileptiform activity after introduction of levetiracetam visualized with electric source imaging.* 2010. **23**: p. 269-278.
122. Tveit, J., et al., *Automated Interpretation of Clinical Electroencephalograms Using Artificial Intelligence.* JAMA Neurology, 2023.

123. Hunter, D.J. and C. Holmes, *Where Medical Statistics Meets Artificial Intelligence.* N Engl J Med, 2023. **389**(13): p. 1211-1219.
124. Fisher, R.S., et al., *ILAE Official Report: A practical clinical definition of epilepsy.* Epilepsia, 2014. **55**(4): p. 475-482.

Papers I-III:

Clinical Neurophysiology 131 (2020) 25–33

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Interictal epileptiform discharges vary across age groups

Eivind Aanestad [a,*], Nils Erik Gilhus [b,c], Jan Brogger [a]

[a] Section for Clinical Neurophysiology, Department of Neurology, Haukeland University Hospital, 5021 Bergen, Norway
[b] Department of Neurology, Haukeland University Hospital, 5021 Bergen, Norway
[c] Department of Clinical Medicine, University of Bergen, Bergen, Norway

ARTICLE INFO

Article history:
Accepted 26 September 2019
Available online 4 November 2019

Keywords:
Interictal epileptiform discharges
SCORE
EEG
Quantitative EEG
Epilepsy
Ageing

HIGHLIGHTS

- Focal interictal epileptiform discharge (IED) morphology changes with age.
- The distribution of quantitative IED measures by age needs to be considered in EEG interpretation.
- IEDs are consistently asymmetric across all age groups.

ABSTRACT

Objective: To investigate whether the occurrence and morphology of interictal epileptiform discharges (IEDs) in scalp-EEG change by age.
Methods: 10,547 patients who had a standard or sleep deprived EEG recording reported using the SCORE standard were included. 875 patients had at least one EEG with focal IEDs. Focal IED morphology was analyzed by age using quantitative measures in EEGLAB and by visual classification based on the SCORE standard. We present distributions of IED measures by age group, with medians, interquartiles, 5th and 95th percentiles.
Results: Focal IEDs occurred most frequently in children and elderly. IED morphology and localization depended on age ($p < 0.001$). IEDs had higher amplitudes, sharper peaks, larger slopes, shorter durations, larger slow-wave areas and wider distributions in children. These morphological characteristics diminished and the IEDs became more lateralized with increasing age. Spike asymmetry was stable across all age groups.
Conclusions: IEDs have age-dependent characteristics. A spike detector, human or computer, should not operate with the same set of thresholds for patients at various age. With increasing age, focal IEDs are less sharp, have lower amplitudes, have less prominent slow-waves and they become more lateralized. Our findings can help EEG readers in detecting and correctly describing IEDs in patients of various age.
Significance: EEG readers should always consider patient age when interpreting interictal epileptiform discharges.

1. Introduction

Interictal epileptiform discharges (IEDs) represent a highly relevant finding in EEG (Bouma et al., 2016; Krumholz et al., 2015; Koutroumanidis et al., 2017). IEDs are usually described in general and descriptive terms without any quantitative definitions (Sannit and Lilienthal, 1962; Chatrian et al., 1974; Noachtar et al., 1999; Kane et al., 2017). Simple metrics such as amplitude, duration and sharpness could be easily obtained, but are not utilized in clinical practice. Clinical experience indicates that IEDs tend to be blunter in the elderly, but there is little data to corroborate this. No studies have looked at quantitative measures of focal IEDs by age.

Localization and frequency of occurrence of IEDs and other focal EEG abnormalities depend on age (Hughes, 1967; Koufen and Gast, 1981). IED morphology and localization may also be influenced by age for well-defined epilepsy types or specific syndromes (Aurlien et al., 2007; Konishi et al., 1994; Sadleir et al., 2009; Lee et al., 2010), but this has not been systematically examined. Aurlien et al. (2009) showed that the amplitude of generalized epileptiform activity changed with age, while the frequency of the discharges did not.

* Corresponding author at: Helse Bergen, Haukeland universitetssykehus, Nevrologisk avdeling, Postboks 1400, 5021 Bergen, Norway.
E-mail addresses: eivind.aanestad@helse-bergen.no (E. Aanestad), nils.gilhus@uib.no (N.E. Gilhus), jan.brogger@helse-bergen.no (J. Brogger).

https://doi.org/10.1016/j.clinph.2019.09.017
1388-2457/

Analytical methods and classification of EEG activity vary in the previous studies. The SCORE standard has been adopted to reduce EEG interpretation variability, and to improve clinical and scientific studies (Beniczky et al., 2017), but so far only sparse data has been published on the basis of SCORE. Detection of IEDs by ordinary visual analysis has less than optimal reproducibility (van Donselaar et al., 1992; Stroink et al., 2006). IED detection is also subject to overinterpretation of non-epileptiform EEG graphoelements as IEDs (Benbadis and Thomas, 2017), and depends on only a few aspects of quantitative IED morphology (Bagheri et al., 2017). More detailed knowledge of how IEDs change by age should improve clinical EEG interpretation in the individual patient and lead to more precise age-dependent spike detection algorithms.

IEDs are defined as transient activity distinguishable from the background activity and with a characteristic morphology typically, but neither exclusively nor invariably, found in interictal EEGs of people with epilepsy (Kane et al., 2017). Six morphological criteria are given, out of which four have to be met in order to classify a graphoelement as IED. IEDs should contain:

(1) Di- or tri-phasic waves with sharp or spiky morphology (i.e. pointed peak).
(2) Wave-duration different from the ongoing background activity, either shorter or longer.
(3) Waveform asymmetry: a sharply rising ascending phase and a more slowly decaying descending phase, or vice versa.
(4) The transient should be followed by an associated slow after-wave.
(5) The background activity surrounding epileptiform discharges should be disrupted.
(6) Distribution of the negative and positive potentials on the scalp should suggest a brain source of the signal, corresponding to a radial, oblique or tangential orientation.

The exact sensitivity and specificity of each criterion for IEDs is not known. Descriptors for criterion 1 (spike and sharp wave morphology) and 4 (presence of slow-wave), are included in the SCORE terminology.

The aim of this study was to investigate how the occurrence and morphology of IEDs change with age in a large and unselected cohort of patients with epilepsy. Precise, age-dependent criteria for IED should improve diagnostic specificity and individualized treatment in epilepsy.

2. Methods

2.1. Patients

We included all consecutive patients who had standard EEGs or sleep deprived EEGs recorded at Haukeland University Hospital during the period March 4th, 2013 - October 29th, 2017, and which were reported in SCORE EEG (13143 EEGs, 10,547 patients, Fig. 1). One EEG was selected for each patient. For the 2596 patients who had recorded two or more EEGs the first EEG with a diagnostic conclusion of epilepsy and epileptiform findings was chosen. If the patient had no such EEG the last EEG recording was chosen. We used the diagnostic conclusion in the clinical EEG report produced in SCORE to further categorize patients. This diagnostic conclusion was drawn by the EEG interpreter from the EEG findings together with available clinical and paraclinical information. A diagnostic conclusion of epilepsy required a clinical suspicion of epilepsy written on the referral and epileptiform activity in the EEG. 9238 patients had a diagnostic conclusion other than epilepsy in their EEG report and served as a control group when assessing demographic characteristics.

1309 patients had at least one EEG with epileptiform findings and a diagnostic conclusion of epilepsy. The first EEG with epileptiform activity for each patient was selected for analysis. The groups with *focal* ($N = 875$), *generalized* ($N = 207$) and *unspecified* ($N = 227$) epilepsy type were analyzed separately, compared to controls, and analyzed by age. The *unspecified* group consisted of patients who had a diagnostic conclusion of either "epilepsy not further specified" ($N = 152$), an epileptic seizure during the EEG recording ($N = 27$), a hypsarrhythmia pattern ($N = 2$), or where the EEG report had conflicting data, e.g. included both focal and generalized epilepsy in the diagnostic conclusion ($N = 46$). The diagnostic conclusion included a suggestion of probable *etiology* of the patient's epilepsy, whether *symptomatic*, *idiopathic* or *undetermined*. Etiology was used as a control variable in multiple linear regression described in Section 2.5. IED morphology was visually and quantitatively analyzed by age and IED localization was visually analyzed in the focal epilepsy group. The patients were grouped by age in years into ten groups: <1 year, 1–9 years, 10–19 years, 20–29 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years, 70–79 years and 80–101 years.

2.2. EEG recording

Electrodes were applied according to the 10–20 system with a minimum of 21 and a maximum of 25 electrodes. When possible, EEGs were recorded with the patients in a supine, relaxed position with their eyes closed. For sleep deprived EEGs, adults were deprived of a whole night's sleep before the recording, while children were kept awake since 3am. the same morning. Patients were encouraged to sleep during the recording. Provocation by hyperventilation and photic stimulation was carried out unless contraindicated. Nicolet™ EEG system was used to record and display EEGs. Average montage with paper speed 3 cm per second and 1 cm per 100 μV was the default setup for review. Montage, sensitivity and paper speed could be adjusted freely by the EEG reader.

2.3. IED morphology

2.3.1. Visual analysis

Morphological categories for IEDs were determined according to the SCORE standard as spikes, spike-and-slow-waves, sharp-waves, sharp-and-slow-waves, polyspikes, polyspike-and-slow-waves, and slow-sharp-waves (Beniczky et al., 2017). Multiple morphologies could be selected for each finding. The patients were divided into two groups depending on whether their EEG contained any IEDs with spike morphology or not. Patients were also divided into two groups depending on whether their EEG contained any IEDs with a slow-wave or not.

2.3.2. Quantitative analysis

The first IED finding in each EEG was chosen for quantitative analysis. All EEGs were digitally filtered using the EEGLAB function pop_eegfilt with a high pass filter at 1 Hz, low pass filter at 70 Hz and a notch band filter spanning 48–52 Hz (Delorme and Makeig, 2004). The electrode or channel in average montage where the IED was most convincingly epileptiform was selected. Spike start, spike peak, spike end and slow-wave end was manually annotated by the first author for all EEGs using custom software built on EEGLAB (ScorePipeline, available from the authors on GitHub (Brøgger, 2019), Fig. 2). Voltage sensitivity and time axis could be adjusted freely for optimal annotation placement. Spike start was marked at the maximal positive time point in the trough leading into the spike. Spike peak was marked at the negative maximum following spike start. Spike end was marked at the maximal positive component following spike peak. If a slow-wave followed, spike end marked its start, and another mark was set at the end

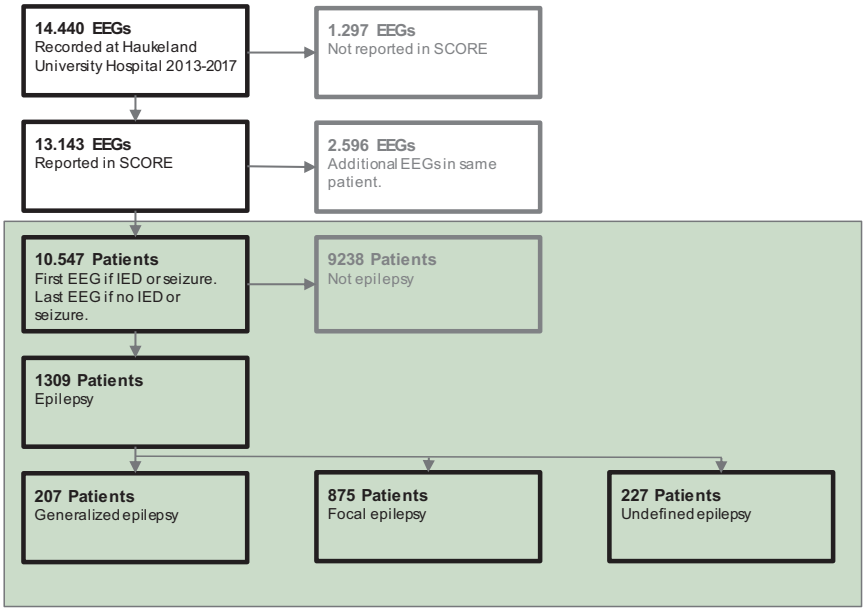

Fig. 1. Patient and EEG flow chart with included and excluded groups.

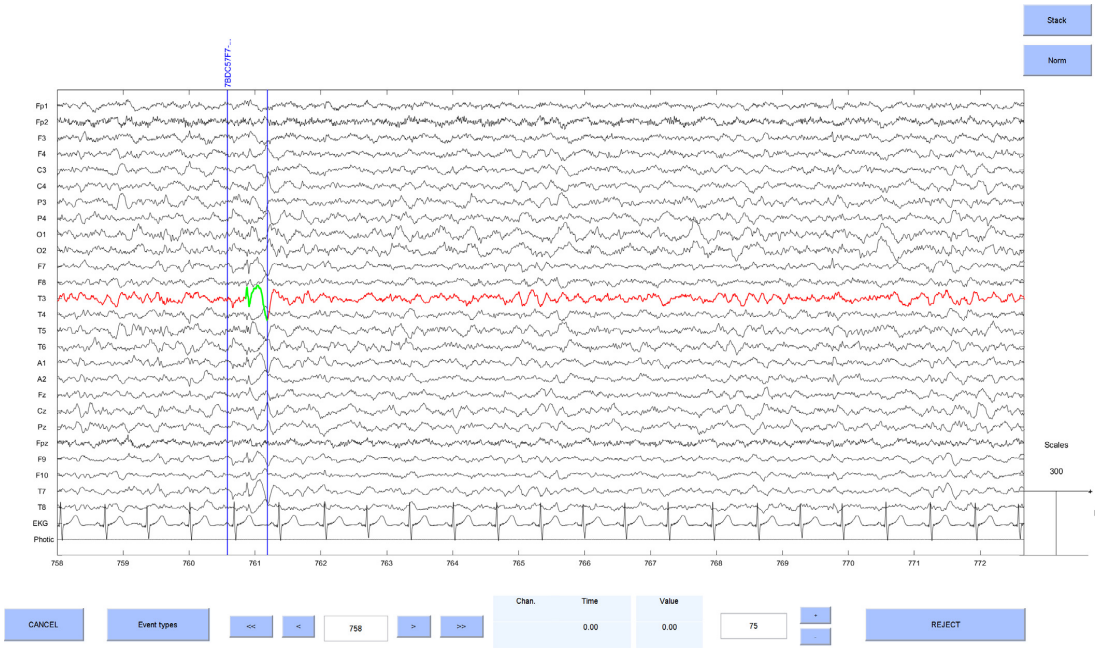

Fig. 2. Manual annotation of an IED in ScorePipeline. Green: Annotation in progress. Leftmost vertical blue line: An EEGLAB event marker. Rightmost vertical blue line: Mouse cursor position on the x-axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the slow-wave. The first half-wave was defined as the sample points between spike start and spike peak. The second half-wave was defined as the sample points between spike peak and spike end. The following quantitative measures were derived from these time points.

2.3.2.1. Sharpness. Approximation D of the $d^2\mu V/dt^2$ value around spike peak according to Frost (1979). Five sample points were obtained at fixed distances from spike peak ($t = 0$ ms), where $N_1 = t - 8$ ms, $N_2 = t - 4$ ms, $N_3 = t - 0$ ms, $N_4 = t + 4$ ms, $N_5 = t + 8$ ms. Then $D = (N_5 - 2^*N_3 + N_1)/2$. Values are greater than 0. A greater value corresponds to a sharper spike peak.

2.3.2.2. Ascending slope. Ascending amplitude (μV) divided by the duration of the first half-wave. Values are greater than 0. A greater value corresponds to a steeper ascending slope.

2.3.2.3. Descending slope. Descending amplitude (μV) divided by the duration of the second half-wave. Values are greater than 0. A greater value corresponds to a steeper descending slope.

2.3.2.4. Ascending amplitude. Voltage difference (μV) between spike peak and spike start. Values are greater than 0. A greater value corresponds to a larger amplitude of the first half-wave.

2.3.2.5. Descending amplitude. Voltage difference (μV) between spike peak and spike end. Values are greater than 0. A greater value corresponds to a larger amplitude of the second half-wave.

2.3.2.6. Duration. Milliseconds from spike start to spike end. Values are greater than 0. A greater value corresponds to a broader spike component of the IED.

2.3.2.7. Area of slow-wave. A Gaussian wave was fitted to the time series segment defined as the slow-wave using MATLAB's *fit* function (MATLAB). The signal was shifted so that the positive maximum was at baseline (0 μV). The trapezoidal integral, defined by the start and end point of the slow-wave, was then subtracted from the area of the fitted Gaussian to give the estimated area of the slow-wave in μV*second (weber, a unit derived from the International System of Units). Values are positive, zero, or negative. A greater positive value corresponds to a larger slow-wave area. Negative values can result from a poor model fit.

2.3.2.8. Spike asymmetry. Duration of the first half-wave divided by the duration of the second half-wave as defined by Henze et al. (2002). Values are real numbers greater than 0. A value less than 1 corresponds to a shorter duration of the first half-wave compared to the second half-wave. A value greater than 1 corresponds to a longer duration of the first half-wave compared to the second half-wave.

To provide a visual reference for EEGers, an average IED was calculated from the raw EEG signal for each age group. The IEDs were averaged centered on the spike peaks as time zero. 95% confidence intervals were calculated for each age group from 10,000 bootstrap samples.

2.4. IED localization

IED visual localization was examined on a regional level where each electrode containing IEDs was assigned to one out of thirteen topographical brain regions *frontal (left/midline/right)*, *central (left/midline/right)*, *temporal (left/right)*, *parietal (left/midline/right)* and *occipital (right/left)*. Ordinal categories were used for multiple linear regression, with regions *frontal, temporal, central, parietal and occipital*, and laterality *left*, *right* and *other*. IEDs were classified as *frontal* if regional localization included the frontal region. Remaining observations were successively classified as *temporal*, *central*, *parietal* and *occipital*. IEDs were classified as *left* if localized strictly to the left hemisphere and/or midline, as *right* if localized strictly to the right hemisphere and/or midline, or as *other*.

2.5. Statistics

Pearson's chi-squared test was used to examine the association between age and epilepsy type and IED morphology, and multiple logistic regression to test age dependency for IED localization. The non-parametric Kruskal-Wallis-test was used to examine age dependency for quantitative IED measures (sharpness, ascending slope, ascending amplitude, duration, spike asymmetry, area of slow-wave). We performed linear regression of the IED quantitative measures (sharpness, ascending slope, ascending amplitude, duration, slow-wave area) as the dependent variable and age, etiology, region and laterality as independent variables, in order to control for the possible effect of location. A p-value threshold of $p < 0.01$ was chosen due to the numerous comparisons that were undertaken.

2.6. Software

Nicolet™ EEG system was used to record and display EEGs for visual analysis. Clinical EEG reports were made with SCORE EEG (versions 1.0.9.4012 to 2.9.16.24). All EEG reports were stored in the SCORE database, a structured SQL database. Quantitative annotation was implemented in custom software built on EEGLAB. All statistics were handled in Stata. Scripts will be made available on GitHub (Brøgger and Aanestad, 2019).

2.7. Ethical approval

The study was approved by the Regional Committees for Medical and Health Research Ethics (reference code 2017/1512/REK vest).

3. Results

3.1. Demography

The mean age of all 10,547 included patients was 35 years, and 48.9% were females (Table 1). The occurrence of both epilepsy and epilepsy type depended on gender ($p < 0.01$ and $p < 0.001$, respectively).

3.2. Epilepsy type

Epilepsy type depended on age ($p < 0.001$, Fig. 3). Focal epilepsy was the most common type in all age groups, with the highest occurrence in children and elderly people. Generalized epilepsy had its peak in adolescence. The age groups 20–29 and 30–39 years had the lowest occurrence of epilepsy.

Table 1
Demographic characteristics of patients included in the study.

	N=	Female % (95% CI)	Age in years mean (SD)
Diagnosis			
Epilepsy	1270	52.5 (49.8–55.3)	35.3 (28.3)
No epilepsy	8971	48.4 (47.4–49.4)	34.8 (25.3)
Total	10,241	48.9 (47.9–49.9)	34.9 (25.7)
Epilepsy type			
Generalized	198	65.2 (58.2–71.5)	22.8 (16.3)
Focal	850	49.8 (46.4–53.1)	40.1 (30.3)
Other	222	51.8 (45.2–58.3)	28.1 (28.3)

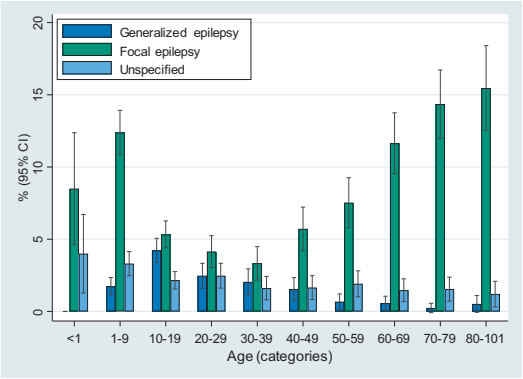

Fig. 3. Occurence of generalized, focal and unspecified epilepsy, by age, in patients referred for standard or sleep-deprived EEG (N = 10547).

3.3. IED morphology by age

The average time series by age illustrate trends in quantitative measures (Fig. 4 and Supplementary Figure 1). The various measures of IED morphology showed that IEDs became blunter with increasing age, and also that slow-waves became less pronounced (Fig. 5A–F). Spike sharpness, slope, amplitude, duration and slow-wave area all depended on age ($p < 0.001$, Table 2).

3.3.1. Age trends by visual classification

The occurrence of spikes and slow-waves both depended on age when examined by visual analysis ($p < 0.001$ and $p < 0.001$. Fig. 5A). Spikes and slow-waves were more often reported in the younger age groups (Fig. 5A). In the age groups 1–19 years, 70% of focal spikes were classified as having a following slow-wave, whereas this was the case for only 30% above age 80 years.

3.3.2. Spike sharpness

Spike sharpness, measured as the approximation $D = d^2V/dt^2$ around the peak, had its maximum in infancy with median D = 4.5. It then declined slowly with increasing age to a minimum with median D = 2.0 above age 80 years. A sharpness of D = 1 was in the 10th percentile below age 70 years. A sharpness of D = 6 was in the 75th percentile at age 0–9 years, but in the 90th percentile or above in patients older than 9 years. The distribution of sharpness was wider below age 10 than above.

3.3.3. Spike slope

Ascending and descending slopes were increasingly steep during the first decade of life, up to a median 2.5 μV/ms in the age group 1–9 years, but then gradually less steep with a further increase in age, down to a median of 1 μV/ms. Slopes of 1 μV/ms were in the 5th percentile in age groups 1–19 years, while in the 25th percentile or higher for other age groups. Slopes of 3 μV/ms were in the 75th percentile in age groups 0–19 years, and in the 90th percentile or higher at ages above 20 years. The distribution of spike slopes was wider below age 10 than above, and declining with age.

3.3.4. Spike amplitude

Spike ascending amplitude had a similar age distribution as spike slopes, with an increase from infancy to early childhood up to a median of 100 μV in age group 1–9 years, and then a gradual decline to a median of 60 μV. An ascending amplitude of 40 μV was in the 25th percentile or lower for all age groups. 100 μV was in the 75th percentile at age 0–19 years, while in the 90th per-

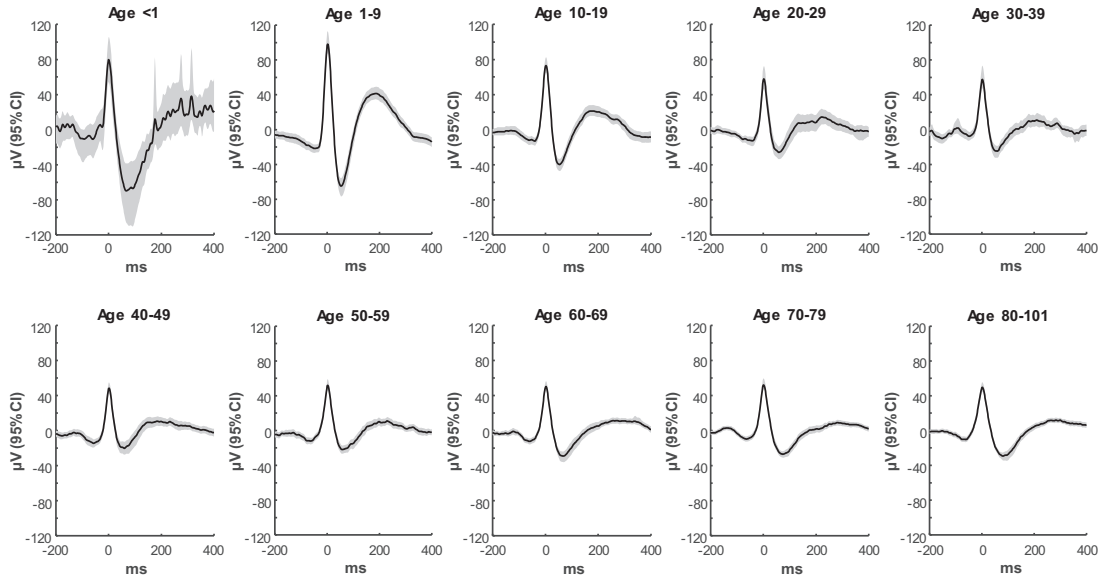

Fig. 4. Average IED time series in focal epilepsy by age categories in years. 95% confidence levels are shown by shaded grey area (barely visible for most age groups). The average was calculated with spike peak defined as time = 0 ms, from 200 ms before until 400 ms after the spike peak, at the electrode where the IED was most convincingly epileptiform (N = 868).

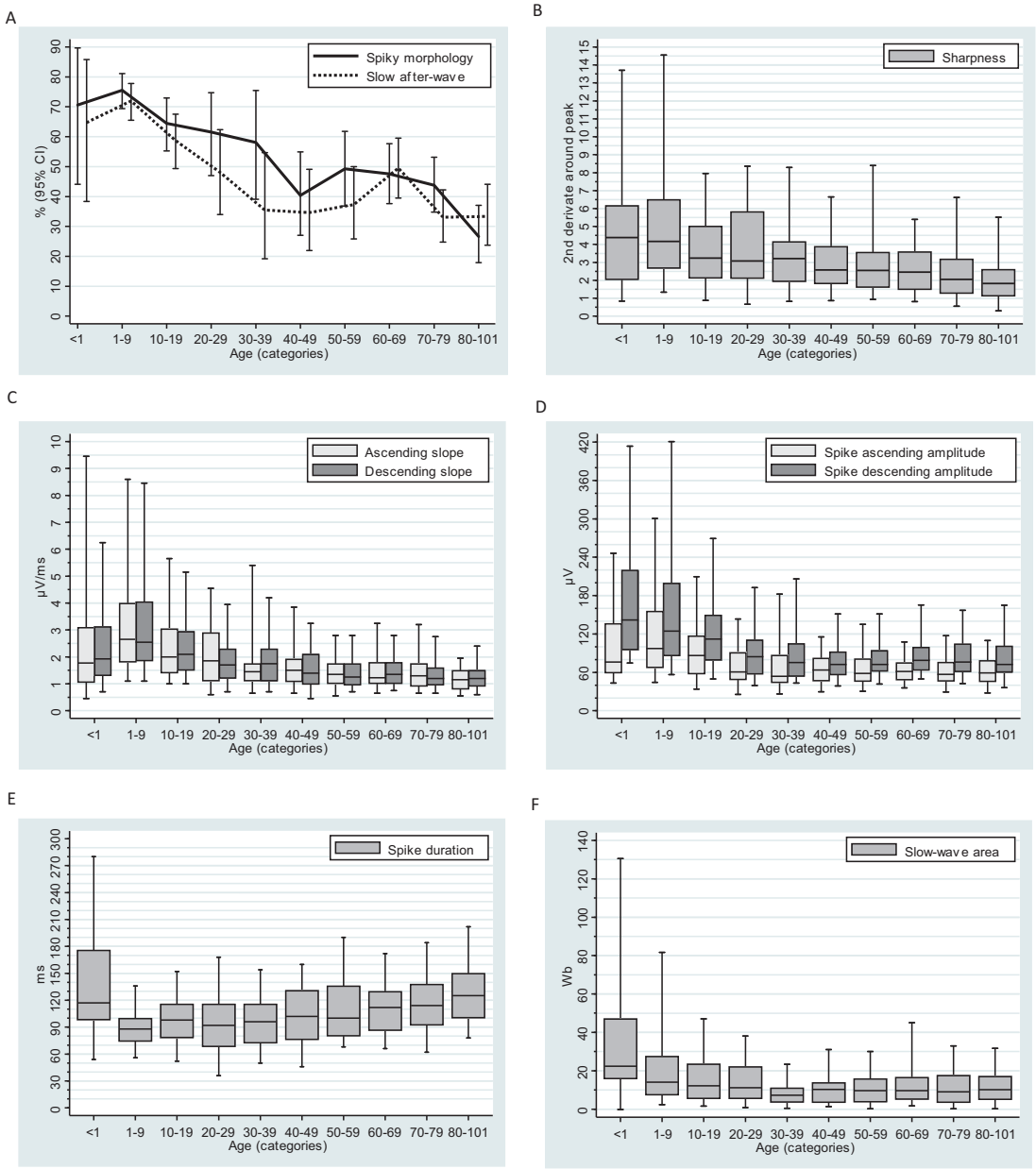

Fig. 5. Morphology and quantitative measures of IEDs by age categories (years). Occurrence is given by percentage with a 95% confidence interval for morphology classified by visual interpretation (A). Median, interquartile range, 5th percentile and 95th percentile is given for the quantitative measures (B–F). A: Occurrence of spikes and slow-waves components in morphological descriptors of IED according to visual analysis with SCORE in focal epilepsy by age. (N = 875). B: Sharpness of the IED spike component around peak (N = 868). C: Ascending and descending slope of the IED spike component in µVolt per ms (N = 868). D: Ascending and descending amplitudes of the IED first and second half waves (N = 868). E: Duration of the IED spike component in milliseconds (N = 868). F: Area of the slow-wave in weber (N = 868).

centile or higher above age 20. Spike descending amplitude was larger compared to the ascending amplitude for all age groups. It had a maximum of 140 µV in infants, then a gradual decrease, before stabilizing at 80 µV for all age groups 20–101 years. The distribution of spike amplitudes was wider below age 10 than above.

3.3.5. Spike duration

Spike duration decreased from early infancy to a minimum median of 90 ms in early childhood. Then it gradually increased with age to a median of 130 ms. Spike duration < 60 ms was in the 5th percentile above age 50 years. A spike duration of > 200 ms

Table 2
Regression model for quantitative IED measures by EEG region, laterality and age category (years).

	Sharpness		Slope		Amplitude		Duration (ms)		Slow-wave area	
	Coef. (μV/ms2)	*p*-value	Coef. (dμV/ms)	*p*-value	Coef. (μV)	*p*-value	Coef. (ms)	*p*-value	Coef. (weber)	p-value
Base value		<0.001		<0.001		<0.001		<0.001		<0.001
Constant	5,8		7,3		141		98,6		32,3	
Region		0,02		0,08		0,29		0,94		<0.01
Frontal	(Reference category)									
Temporal	−0,8		−0,9		−7,6		0,7		−4,5	
Central	−1		−0,9		−18,8		2,7		−8,3	
Parietal	−0,1		−0,1		0		1,9		−4,1	
Occipital	0,9		0,6		1,8		−6,9		−7,2	
Laterality		<0.01		<0.001		0,01		<0.001		0,02
Left	(Reference category)									
Right	0,5		0,8		9,6		2,8		2,3	
Other	0,8		1,2		17		−8,7		4,4	
Etiology		0,04		<0.01		<0.001		<0.01		<0,001
Idiopathic	−1,1		−1,2		−16,2		−4,3		−13,2	
Symptomatic	(Reference category)									
Undetermined	−0,4		−0,9		−22		−9,6		−9,4	
Age category		<0.001		<0,001		<0,001		<0,001		<0,001
<1	−1,6		−2,7		−35,3		55,6		6,6	
1–9	(Reference category)									
10–19	−1,7		−2,2		−40,5		7,6		−10,9	
20–29	−1,9		−3,2		−66,3		−1,1		−14,8	
30–39	−2,8		−4,4		−76,5		4,8		−22,3	
40–49	−2,8		−4,2		−75,6		8,0		−20,5	
50–59	−2,6		−4,4		−71,5		16,9		−19,9	
60–69	−3		−4,4		−77,2		15,4		−16,8	
70–79	−3,1		−4,5		−76,9		19,1		−19,2	
80–101	−3,9		−5,2		−79,7		35,3		−19,9	
Adjusted R-squared	0,14		0,19		0,16		0,14		0,15	

was in the 90th percentile in infants < 1 year old, and in the 95th percentile or above for all ages 1–101 years. The distribution of spike durations was wider below age 1 years than any other age, then reached a minimum at ages 1–9 years.

3.3.6. Slow-wave

The slow-wave area was largest and most variable in infancy with median 20 weber. It decreased during early childhood to a median of 15 weber, but with substantial variability for age group 1–9 years. It then stabilized at a median of around 10 weber for the age groups 10–101 years. Slow-waves with an area of 50 weber were in the 90th percentile at age 0–9 years, and in the 99th percentile or higher for ages above 10 years.

3.3.7. Spike asymmetry

Spike asymmetry had a median around 0.8 for all groups and did not depend on age (results not shown in Fig. 5).

3.3.8. Regression model

Multiple linear regression models for quantitative IED measures by brain region, laterality, etiology and age, showed that age had the strongest effect on all of these measures (Table 2). Idiopathic and undetermined etiology had a subtractive effect on spike slope, spike amplitude, spike duration and slow-wave. For ease of interpretation the percent changes in age coefficients are shown in Fig. 6. The oldest patient group had the most pronounced change for all quantitative measures, and the coefficient change from the base value ranged from 36% to 72%. The only exception was for slow-wave area.

3.4. Localization of focal IEDs

IED localization depended on age for most brain regions ($p < 0.001$). The exceptions were the right frontal, right temporal and left parietal regions (Fig. 7). IEDs became more lateralized with increasing age. The occipital and central regions rarely showed

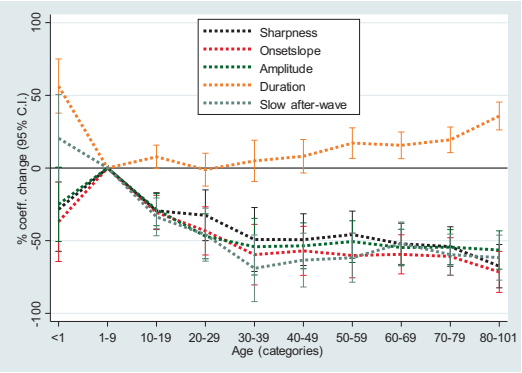

Fig. 6. Percent change in the coefficient size of age (years) from a linear regression of quantitative IED measures by age adjusted for brain region, laterality and etiology. The age group 1–9 years was used as reference in the regression model.

IEDs in elderly patients. IEDs were increasingly common over the left hemisphere with increasing age.

4. Discussion

We have shown that the morphology of IEDs depends on age. Focal IEDs become more common with age, and their quantitative characteristics change. With increasing age IEDs appear less sharp, have lower amplitudes, have less prominent slow-waves, and their scalp localization becomes more lateralized. They also occur more frequently over the left hemisphere. Spike asymmetry was our only IED measure that did not vary by age, and this IED criterion applies evenly for all ages. Our findings can help EEG readers in detecting and correctly describing IEDs in patients of various age.

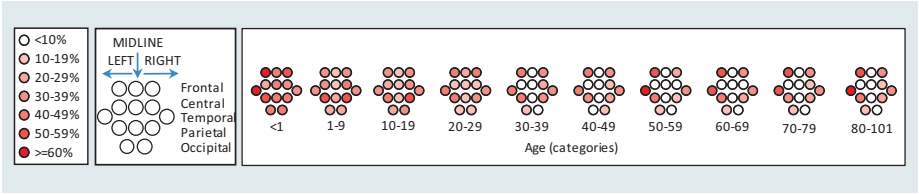

Fig. 7. Regional occurrence of IEDs by age (years) as scored by visual analysis (N = 844). One head model is shown for each age group. Circles represent EEG regions. Intensity of red indicates the percentage of IEDs that occurred in the corresponding region and age group.

Guidelines for classifying IEDs do not include quantitative criteria, and no provision has been made for changes in the criteria with age. Our data provide such quantitative characteristics to guide the EEGer in IED detection and classification. All quantitative measures falling outside the 5th or 95th percentile of these age distributions should be interpreted cautiously to avoid false positive IED detection. For example, if an EEG reader is contemplating whether a sharp wave with ascending slope of less than 1 μV/ms represents an IED in a 15 year old patient, this would be a rare IED event. The possibility that the wave instead represents an artifact or a physiological waveform should be examined more closely. On the other hand, if the patient was 90 years old, the ascending slope would be well within the interquartile range and even close to the median value for the corresponding age group. It is apparent from our findings that the spike detector, human or computer, cannot operate with the same set of thresholds for patients of various ages. Furthermore, the observed differences imply that the sensitivity and specificity in IED detection might not be the same across age groups.

IEDs with a duration of less than 35 ms were infrequent in our material, while IEDs with a duration greater than 200 ms were seen occasionally, mostly in the very young and very elderly. No limits for the duration of IEDs are given in the definition of epileptiform activity. Spikes and sharp waves are defined separately as epileptiform transients with a duration of 20 to less than 70 ms and 70 ms to 200 ms respectively, effectively rendering a portion of the IEDs in our material unnamed. The glossary of terms should not exclude IEDs by arbitrary limitations.

Despite blunted and low amplitude IEDs, the oldest age group had the highest occurrence of focal IEDs. The prevalence of epilepsy is higher in the elderly (Beghi and Giussani, 2018), but the sensitivity of EEG decreases with age as the occurrence of IEDs in elderly with epilepsy is less frequent than in younger patients (Drury and Beydoun, 1998). A higher signal-to-noise ratio due to lower background activity power in older patients (Dustman et al., 1999) makes their IEDs stand out more. Life expectancy is increasing and the oldest segment of the population is expanding (Christensen et al., 2009). To diagnose epilepsy and detect IEDs with sufficient sensitivity and specificity in elderly patients will be of even greater importance in the future and needs special attention in diagnostic workup.

The quantitative IED measures are not independent variables. They are mostly measures of the same triangular-like shape that constitutes the spike. Still, each measure represents visually distinct and meaningful properties of the IED. The measures display a similar percent wise change by age when controlled for IED localization, laterality and etiology. No attempt was made to explain if, and how, one quantitative IED measure predicts another. The aim of this study was rather to test the hypotheses that IED morphology and IED occurrence change with age, which were confirmed for both.

We included all EEGs examined at our department for this study, not excluding those with referral reasons other than a suspicion of epilepsy. 63% of the patients were referred with an indication related to epilepsy (data not shown). This included patients for whom the epilepsy diagnosis had not yet been established, and those with an established diagnosis where monitoring or follow-up was requested to guide therapy. The occurrence of epilepsy had a bimodal distribution with two peaks, one at age 1–9 years and another at 80–101 years. Our laboratory is the only EEG provider in our region, and so our material consisted of an unselected and complete EEG patient population referred from a wide range of general practitioners and specialists.

This is a cross-sectional study. We annotated only the first IED of the first epileptiform EEG for each patient, which may not have been the most prominent or informative. Variability of IEDs within the same EEG occurs, but was not examined in this study. Our regression model might have better explained the quantitative IED measures with access to variables such as background activity power, patient medication, intracranial imaging data and seizure frequency. Decreasing general EEG amplitude with age is a known phenomenon (Dustman et al., 1999), and this may influence the amplitude also of IEDs.

5. Conclusions

Focal IEDs occurred most frequently in children and elderly. IEDs have age-dependent characteristics. With increasing age, focal IEDs appeared less sharp, had lower amplitudes, and had less prominent slow-waves. With increasing life expectancy these changes in IED morphology have increasing relevance. A spike detector, human or computer, should not operate with the same set of thresholds for patients at various ages.

Acknowledgements

Harald Aurlien and Tom Eichele gave valuable advice during this project.

Declaration of Competing Interest

Jan Brøgger and Eivind Aanestad are minority shareholders in Holberg EEG AS, the providers of the SCORE EEG software used in this study. Nils Erik Gilhus declares no conflicts of interest.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.clinph.2019.09.017.

References

Aurlien H, Aarseth JH, Gjerde IO, Karlsen B, Skeidsvoll H, Gilhus NE. Focal epileptiform activity described by a large computerised EEG database. Clin Neurophysiol 2007;118(6):1369–76.

Aurlien H, Gjerde IO, Eide GE, Brogger JC, Gilhus NE. Characteristics of generalised epileptiform activity. Clin Neurophysiol 2009;120(1):3–10.
Bagheri E, Dauwels J, Dean BC, Waters CG, Westover MB, Halford JJ. Interictal epileptiform discharge characteristics underlying expert interrater agreement. Clin Neurophysiol 2017;128(10):1994–2005.
Beghi E, Giussani G. Aging and the epidemiology of epilepsy. Neuroepidemiology 2018;51(3–4):216–23.
Benbadis SR, Thomas P. When EEG is bad for you. Clin Neurophysiol 2017;128(4):656–7.
Beniczky S, Aurlien H, Brogger JC, Hirsch LJ, Schomer DL, Trinka E, et al. Standardized computer-based organized reporting of EEG: SCORE - Second version. Clin Neurophysiol 2017;128(11):2334–46.
Bouma HK, Labos C, Gore GC, Wolfson C, Keezer MR. The diagnostic accuracy of routine electroencephalography after a first unprovoked seizure. Eur J Neurol 2016;23(3):455–63.
Brøgger JC. ScorePipeline, <https://github.com/janbrogger/ScorePipeline.git>; 2019.
Brøgger JC, Aanestad E. Scripts, <https://github.com/janbrogger/Art1EpiMorphLoc>; 2019.
Chatrian G, Bergamini L, Dondey M, Klass D, Lennox-Buchtal M, Petersen I. A glossary of terms most commonly used by clinical electroencephalographers. Electroencephalogr Clin Neurophysiol 1974;37(5):538–48.
Christensen K, Doblhammer G, Rau R, Vaupel JW. Ageing populations: the challenges ahead. The Lancet 2009;374(9696):1196–208.
Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134(1):9–21.
Drury I, Beydoun A. Interictal epileptiform activity in elderly patients with epilepsy. Electroencephalogr Clin Neurophysiol 1998;106(4):369–73.
Dustman RE, Shearer DE, Emmerson RY. Life-span changes in EEG spectral amplitude, amplitude variability and mean frequency. Clin Neurophysiol 1999;110(8):1399–409.
Frost Jr JD. Microprocessor-based EEG spike detection and quantification. Int J Biomed Comput 1979;10(5):357–73.
Henze DA, Wittner L, Buzsáki G. Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. Nat Neurosci 2002;5(8):790–5.
Hughes J. EEG epileptiform abnormalities at different ages. Epilepsia 1967;8(2):93–104.
Kane N, Acharya J, Benickzy 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. Clin Neurophysiol Pract 2017.
Konishi T, Naganuma Y, Hongou K, Murakami M, Yamatani M, Yagi S. Changes in EEG foci with age in childhood partial epilepsies. Clin Electroencephalogr 1994;25(3):104–9.
Koufen H, Gast C. Zur Frage der Alters- und Diagnoseabhängigkeit der Links-Lateralisation und Lokalisation von EEG-Herden. Arch Psychiatr Nervenkr (1970) 1981;229(3):227-37.
Koutroumanidis M, Arzimanoglou A, Caraballo R, Goyal S, Kaminska A, Laoprasert P, et al. The role of EEG in the diagnosis and classification of the epilepsy syndromes: a tool for clinical practice by the ILAE Neurophysiology Task Force (Part 1). Epileptic Disord 2017;19(3):233–98.
Krumholz A, Shinnar S, French J, Gronseth G, Wiebe S. Evidence-based guideline: Management of an unprovoked first seizure in adults: report of the Guideline Development Subcommittee of the American Academy of Neurology and the American Epilepsy Society. Neurology 2015;85(17):1526–7.
Lee I-C, Chen Y-J, Lee H-S. Migration of epileptic spike foci in encephalograms may correlate with a better outcome in pediatric epilepsy. Brain Dev 2010;32(10):821–8.
MATLAB. fit, <https://se.mathworks.com/help/curvefit/fit.html>.
Noachtar S, Binnie C, Ebersole J, Mauguiere 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.
Sadleir LG, Scheffer IE, Smith S, Carstensen B, Farrell K, Connolly MB. EEG features of absence seizures in idiopathic generalized epilepsy: impact of syndrome, age, and state. Epilepsia 2009;50(6):1572–8.
Sannit T, Lilienthal E. A Glossary for EEG Technicians. Am J EEG Technol 1962;2(4):106–19.
Stroink H, Schimsheimer RJ, de Weerd AW, Geerts AT, Arts WF, Peeters EA, et al. Interobserver reliability of visual interpretation of electroencephalograms in children with newly diagnosed seizures. Dev Med Child Neurol 2006;48(5):374–7.
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(3):231–7.

ORIGINAL RESEARCH

A New Score for Sharp Discharges in the EEG Predicts Epilepsy

OPEN

Eivind Aanestad,*† Nils E. Gilhus,‡† and Jan Brogger*†

*Section for Clinical Neurophysiology, Department of Neurology, Haukeland University Hospital, Bergen, Norway; ‡Department of Neurology, Haukeland University Hospital, Bergen, Norway; and †Department of Clinical Medicine, University of Bergen, Bergen, Norway.

Purpose: A challenge in EEG interpretation is to correctly classify suspicious focal sharp activity as epileptiform or not. A predictive score was developed from morphologic features of the first focal sharp discharge, which can help in this decision.

Methods: From a clinical standard EEG database, the authors identified 2,063 patients without a previous epilepsy diagnosis who had a focal sharp discharge in their EEG. Morphologic features (amplitude, area of slow wave, etc.) were extracted using an open source one-click algorithm in EEGLAB, masked to clinical classification. A score was developed from these features and validated with the clinical diagnosis of epilepsy over 2 to 6 years of follow-up. Independent external validation was performed in Kural long-term video-EEG monitoring dataset.

Results: The score for the first focal sharp discharge had a moderate predictive performance for the clinical designation as the EEG being epileptiform (area under the receiver operating characteristics curve = 0.86). Best specificity was 91% and sensitivity 55%. The score also predicted a future epilepsy diagnosis (area under the receiver operating characteristics curve = 0.70). Best specificity was 86% and sensitivity 38%. Validation on the external dataset had an area under the receiver operating characteristics curve = 0.80. Clinical EEG identification of focal interictal epileptiform discharges had an area under the receiver operating characteristics curve = 0.73 for prediction of epilepsy. The score was based on amplitude, slope, difference from background, slow after-wave area, and age. Interrater reproducibility was high (ICC = 0.91).

Conclusions: The designation of the first focal sharp discharge as epileptiform depends on reproducible morphologic features. Characteristic features were amplitude, slope, slow after-wave area, and difference from background. The score was predictive of future epilepsy. Halford semiquantitative scale had similar diagnostic performance but lower reproducibility.

Key Words: Epileptiform, Morphology, Quantitative, SCORE, Validation, Feature.

(J Clin Neurophysiol 2021;00: 1–8)

Detection of interictal epileptiform discharges (IEDs) is a major task in the clinical review of EEG.[1–3] A common challenge in EEG interpretation is to classify focal sharp-appearing activity as epileptiform or not. Visual analysis is the current gold standard for this classification, but interrater agreement is only moderate.[4–7]

No quantitative guidelines exist to help in classifying sharp-appearing activity as epileptiform or not. The score in the study by Halford et al.[8] classified discharges for likelihood of being epileptiform on a 1 to 5 Likert scale, but this score has limited use and variable reproducibility. The 2017 EEG glossary[9] introduced a criterion-based scoring of epileptiform activity, but with limited data to support the criteria. These new criteria apply qualitative terms for morphologic properties of IEDs and do not give any quantitative definitions. The EEG reader has to rely on experience and training when evaluating whether a transient fulfills qualitative IED criteria. Kural et al.[10] examined the new criteria and concluded that five of six criteria should be fulfilled for optimal visual acceptance of epileptiform discharges.

Misdiagnosis of epilepsy is common.[11] Missed epileptiform activity in the EEG occurs on occasion.[12] Specialized epilepsy centers report a high prevalence of false-positive EEGs[12–14] and urge a conservative approach in IED assessment. While a higher specificity may be sensible for such centres, meta-analyses of EEG interpretation after a first seizure show the need to balance sensitivity and specificity.[1,15]

Most studies involving IED quantification have included development and application of automated spike detectors.[16–19] A few studies have examined specific quantitative IED features and their relation to a future epilepsy diagnosis,[20] their correlation with human IED detection,[21] their reproducibility,[20] age dependency,[22,23] and how they can contribute to epilepsy syndrome classification.[24,25]

The aim of this article was to develop a publicly available predictive score, the Bergen Epileptiform Morphology Score (BEMS), from morphologic features of the first suspicious sharp discharge, which can help in the classification of sharp-appearing as epileptiform or not. We improved an existing freely available algorithm[22] to measure morphology of sharp-appearing activity with one click on the pointed peak. The best morphologic features were combined into a predictive score.

METHODS

Patients and EEGs

We included all consecutive inpatients and outpatients who had standard EEGs or sleep-deprived EEGs recorded in our EEG

J. Brøgger and E. Aanestad are minority shareholders in Holberg EEG AS, the providers of the SCORE EEG software used in this study. The remaining author has no funding or conflicts of interest to disclose.

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 Eivind Aanestad, MD, Nevrologisk avdeling, Haukeland universitetssykehus, Helse Bergen, Postboks 1400, 5021 Bergen, Norway; e-mail: eivind.aanestad@helse-bergen.no.

ISSN: 0736-0258/21/0000-0001
DOI 10.1097/WNP.0000000000000849

laboratory facilities at Haukeland University Hospital, during the period March 4, 2013 to October 29, 2017, that were reported in SCORE EEG,[26] and that had nonepileptiform sharp transients or focal epileptiform activity (Fig. 1).

We excluded all patients who had an EEG before the inclusion period, a prior clinical diagnosis of epilepsy (since January 1, 1999), or a nonfocal epilepsy finding on EEG. For each of the remaining patients, we selected their first EEG with either (1) interictal epileptiform discharges (IEDs) and an EEG conclusion of focal epilepsy, hereafter simplified as "focal IEDs," or (2) sharp transients, wicket spikes, small sharp spikes (benign epileptiform transients of sleep), 6-Hz spike and slow wave, rudimentary spike–wave complex, hereafter simplified as "sharp transients," without an EEG conclusion of epilepsy. The EEG conclusion was drawn by the EEG interpreter during the initial clinical evaluation of the EEG and based on the EEG findings together with available clinical and paraclinical information, similar to the recommended "clinical correlation" section of the American Clinical Neurophysiology Society Guideline report template.[27] Patients were categorized into clinical outcomes of (1) epilepsy or (2) not epilepsy according to whether they had received a clinical diagnosis of epilepsy in the hospital records during follow-up until November 27, 2019. We selected for each patient the first EEG that contained sharp activity, and in that EEG,

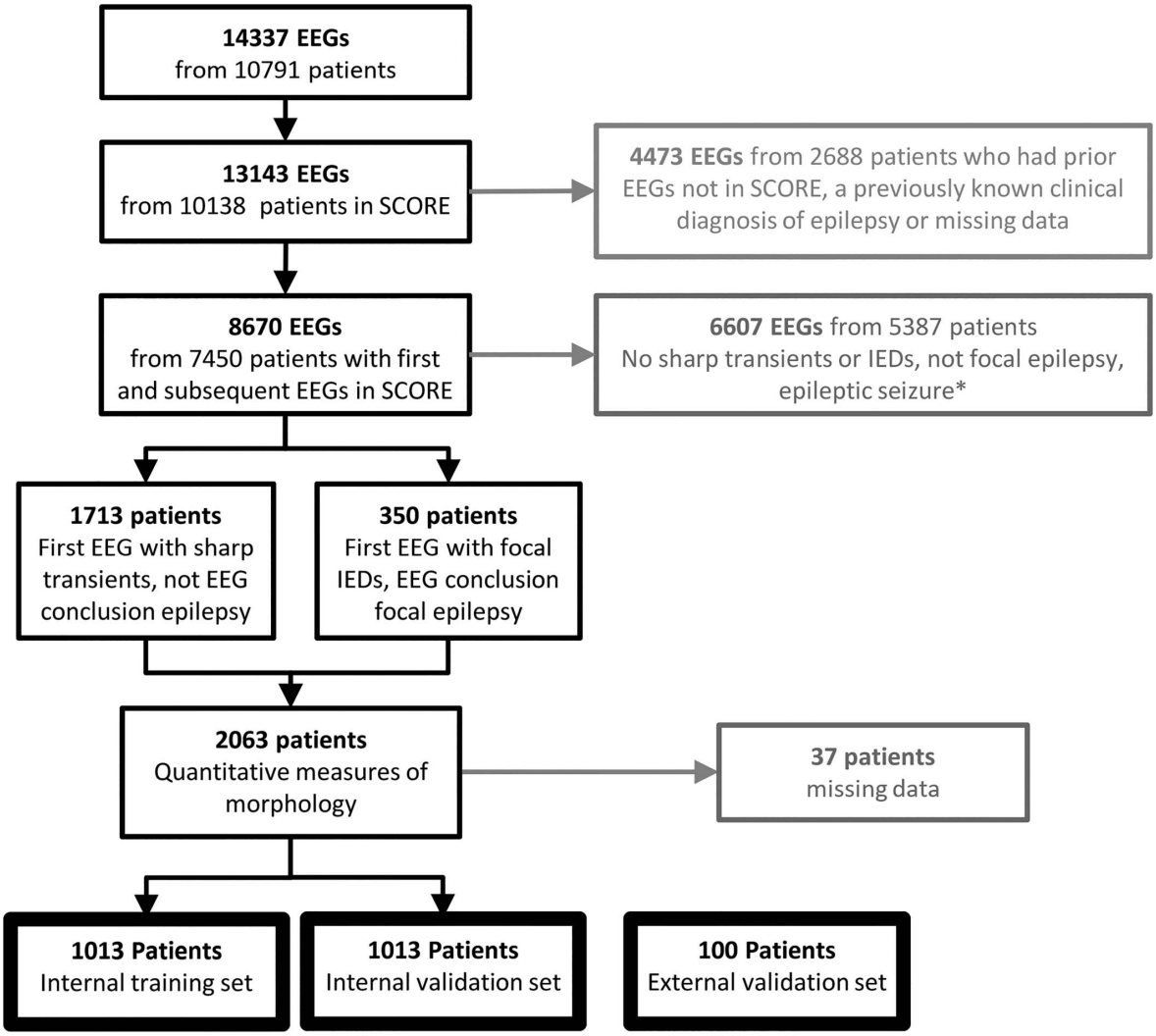

FIG. 1. Flow chart of patients. *EEG recordings with epileptic seizures were excluded.

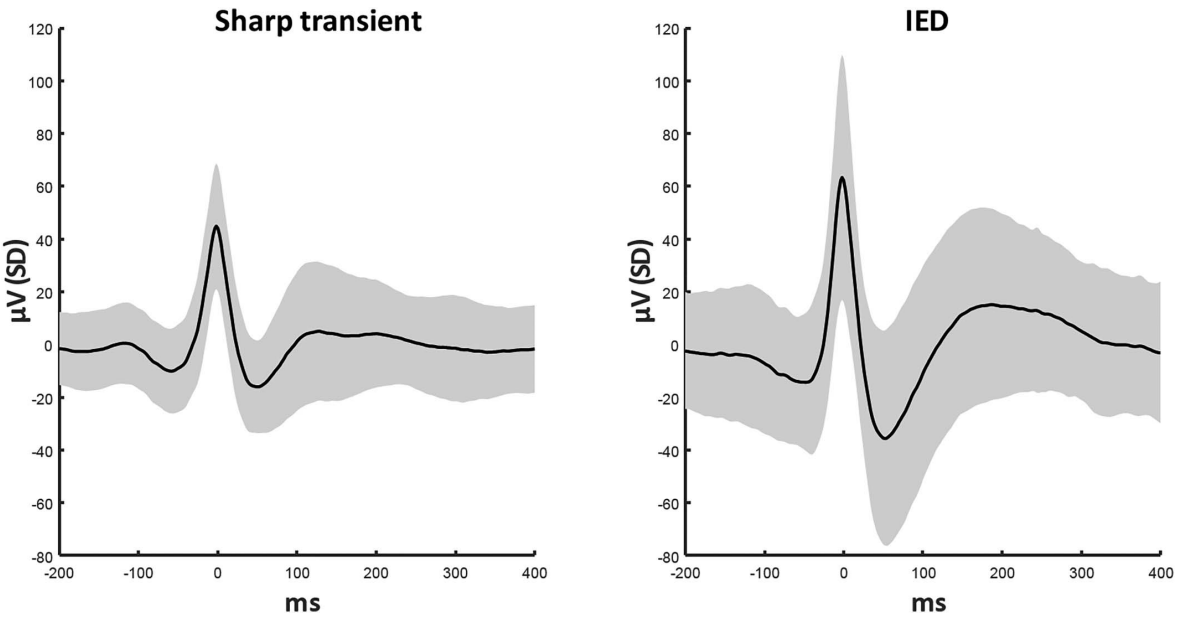

FIG. 2. Morphologic averages of sharp transients and focal IEDs in standard EEG. Average time series of sharp transients ($N = 1{,}677$) and IEDs ($N = 349$). One SD is shown by shaded grey areas. The average was calculated with the spiky component peak defined as time = 0 ms, from 200 ms before until 400 ms after the peak, and from the electrode where the sharp transient or IED was most convincingly epileptiform. IED, interictal epileptiform discharge.

we measured morphologic features of the first sharp discharge. The EEGs were randomized into two equally sized groups for multivariate analysis; one training set for dose–response modeling and elimination of similar features and another validation set.

EEG Recordings

Electrodes were applied according to the 10 to 20 system with a minimum of 21 and a maximum of 26 electrodes. Recording length was 20 minutes for standard EEGs and 60 minutes for sleep-deprived EEGs. NicoletOne EEG system was used to record and display EEGs.

IED Features

We selected the first sharp discharge marked as a sharp transient or an IED during the clinical EEG review. If several sharp appearing waves were present on the same EEG page, the most convincing epileptiform wave was chosen. Blinding was done for the clinical description as either a sharp transient or a focal IED. The first author of this article evaluated all sharp discharges, and the last author evaluated a random subset of 244 sharp discharges. All discharges were also scored according to the 5-point Likert scale by Halford et al.[8] We assessed the interrater reproducibility of Halford scale on the 244 sharp discharges evaluated by both authors.

Morphologic features of the sharp discharges were obtained using a custom-built tool in MATLAB[28]/EEGLAB[29] described previously.[22] The code is freely available (https://github.com/janbrogger/EpiOneClick). The software automatically determines 11 features of the sharp discharge. These were ascending and descending spike amplitudes, ascending and descending spike slopes, spike sharpness according to Frost,[30] spike duration, spike asymmetry according to Henze et al.,[31] spike to background power, slow after-wave area, background amplitude, and spatial extent of the negative spike pole. A detailed description of the algorithm is given in **Supplemental Digital Content 1** (see **S1**, http://links.lww.com/JCNP/A155), interrater reproducibility data for two authors using it in **Supplemental Digital Content 1** (see **S2**, http://links.lww.com/JCNP/A155), and inter-method data for sharp discharges that were annotated in a previous study[22] using four mouse clicks in **Supplemental Digital Content 1** (see **S3**, http://links.lww.com/JCNP/A155).

Multivariate Predictive Modeling

The model included spike descending amplitude, spike onset slope, slow after-wave area, spike to background power, and patient age as predictor variables. This model was applied on the validation set and an independent external dataset.[10] The BEMS was calculated by multiplying the model coefficients by 10, rounding and centering the score. Model probabilities by this score were calculated by averaging probabilities from the logistic regression model.

Validation on an External Dataset

The developed BEMS was applied on an external and independent dataset[10] to assess generalizability. This dataset consists of one short EEG segment per patient. Each EEG

segment contains a sharp transient (epileptiform or nonepileptiform). Characteristics that differed from our primary dataset were patient selection (long-term video-EEG monitoring), greater pretest probability of epilepsy (54%), and an outcome defined by recorded seizures as either epileptic or nonepileptic on long-term video-EEG monitoring.

Statistics

The diagnostic performance of the multivariate predictive model and of each sharp discharge feature for both EEG and clinical outcome was quantified as sensitivity, specificity, and area under the receiver operating characteristics curve (AUC) for included EEGs. Accuracy, the percentage of correctly classified observations, was calculated for the BEMS performance in the external dataset. The multivariate logistic model building was performed with univariate dose–response estimation using locally weighted regression[32] and quartiles to guide the selection of multiple cut points for each feature. We used logistic regression with EEG outcome as the dependent variable to exclude nonsignificant features by using Wald test. The cumulative incidence of a diagnosis of epilepsy was estimated with Stata stcrreg,[33] accounting for death as a competing risk.[34] The diagnostic performance of the clinical EEG conclusion and the BEMS score in predicting clinical outcome (future epilepsy) was estimated from the cumulative incidence of epilepsy at up to 6 years of follow-up. Interrater agreement for the BEMS score was calculated using intraclass correlation. Interrater agreement for Halford semiquantitative score was calculated with Cohen[35] kappa.

A formal sample size calculation was not performed for the main study as we used all the available data. However, we decided that events per variable should be >10.[36] We selected 50% of the EEGs for the prediction model development and 50% for validation to have sufficient statistical power for all categorical variables in the final validation model.

Software

Nicolet EEG system was used to record and display EEGs for clinical visual analysis. Clinical EEG reports were made with SCORE EEG (versions 1.0.9.4012 to 2.9.16.24). All EEG reports were stored in the SCORE EEG database, a structured SQL database. Quantitative annotation was implemented in custom software built on EEGLAB. All statistics were handled in Stata.[33]

Ethical Approval

The study was approved by the Regional Committees for Medical and Health Research Ethics (reference code 2017/1512/REK vest).

RESULTS

Demography

Two thousand twenty-six patients were included after excluding 653 patients that were not reported in SCORE EEG,[26] 2,688 patients who had an EEG before the inclusion period, a previously known clinical diagnosis of epilepsy or missing data, 5,387 patients who did not have sharp transients or focal IEDs in their EEGs or who had an epileptic seizure during their EEG recordings, and 37 patients because of missing data with regard to morphologic measurements. Included patients had a wide age range (see **Table S4**, **Supplemental Digital Content**

TABLE 1. BEMS for Classifying a Sharp Discharge as Focal IED or as a Sharp Transient

Predictor	Corresponding Epileptiform Criterion	Category	Points
Spike descending amplitude (μV)	Criterion 1	0–69	1
		70–89	0
		90–119	7
		>119	17
Spike onset slope (μV/ms)	Criterion 1	0–0.9	0
		1.0–1.4	4
		1.5–1.9	5
		>1.9	11
Spike to background power (%)	Criterion 3	>8.5	0
		4.7–8.5	9
		2.6–4.6	6
		0–2.5	14
Slow after-wave area (weber)	Criterion 4	0–4	0
		5–9	6
		10–19	11
		>19	19
Age (years)	None	0–9	16
		10–19	0
		20–59	12
		>59	25

BEMS is calculated by summing the individual scores for spike amplitude, spike onset slope, spike to background power, slow after-wave area, and age. See Fig. 3 for BEMS-to-probability translation.

BEMS, Bergen Epileptiform Morphology Score; IED, interictal epileptiform discharge.

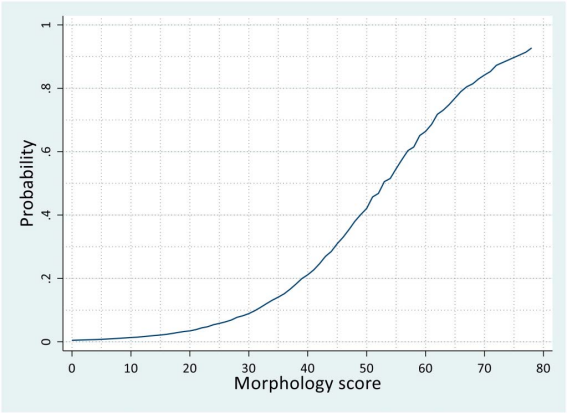

FIG. 3. Mean model-predicted probability of IED by morphology score in the validation set. IED, interictal epileptiform discharge.

1, http://links.lww.com/JCNP/A155), with a mean age of 39 years (SD = 27) and a female overrepresentation (56%). Patients with focal IEDs (n = 350) were 5 years older on average than patients with sharp transients (n = 1,713). Ninety percent of included EEGs were the patient's first EEG, and 93% were standard EEGs, reflecting the aim to examine the earliest EEG containing sharp activity. Seventy-six percent of patients were referred with a clinical suspicion of epilepsy. The time interval between the date of the included EEG and follow-up ranged from 769 to 2,447 days. Sixty-five percent of the patients who was diagnosed with epilepsy during follow-up had one or several acute emergency hospital admissions for epilepsy.

Morphologic Features

The distribution of morphologic features, except for spike duration, differed significantly between the two EEG outcome categories (see **Table S1**, **Supplemental Digital Content 1**, http://links.lww.com/JCNP/A155). However, there was also considerable overlap between them (see **Table S2**, **Supplemental Digital Content 1**, http://links.lww.com/JCNP/A155; Fig. 2). Descending amplitude, slow after-wave area and preceding background power had an AUC >0.7. Spike to background power, spike sharpness, duration, onset slope, descending slope, ascending amplitude, Henze asymmetry, and number of channels had an AUC ≤ 0.7. The BEMS score had the same performance as the visually assessed Halford score; AUC = 0.84 for both. The AUC was 0.70 for both BEMS and Halford score for the clinical outcome of epilepsy and was ≤0.64 for all univariate quantitative features.

Prediction of EEG Outcome in Validation Set

Table 1 shows how the IED features contributed to the BEMS score in the validation set (N = 1,013). The AUC for the BEMS score and in the multivariate logistic regression model was the same, with AUC = 0.86. A cut point of 46 on BEMS had a specificity of 91% for a clinical EEG conclusion of epileptiform activity, with a sensitivity of 55%. A cut point of 29 had a specificity of 57% with a sensitivity of 90%. The translation from the BEMS score to probability based on this model is given in Fig. 3. Odds ratios for the BEMS categories in the validation dataset are shown in **Supplemental Digital Content 1** (see **Figure S3**, http://links.lww.com/JCNP/A155).

Prediction of Epilepsy

A higher BEMS was associated with a higher risk of epilepsy (Fig. 4A; N = 2026). The cumulative incidence of a clinical diagnosis of epilepsy in up to 6 years of follow-up was 10% in patients with a BEMS of 0 to 16 points, 14% with a BEMS of 17 to 23 points, 23% with a BEMS of 24 to 32 points, 34% with a BEMS of 33 to 43 points, and 50% with a BEMS of 44 to 79 points. The receiver operating AUC was 0.70 for both Halford score and for BEMS in five quantiles with clinical

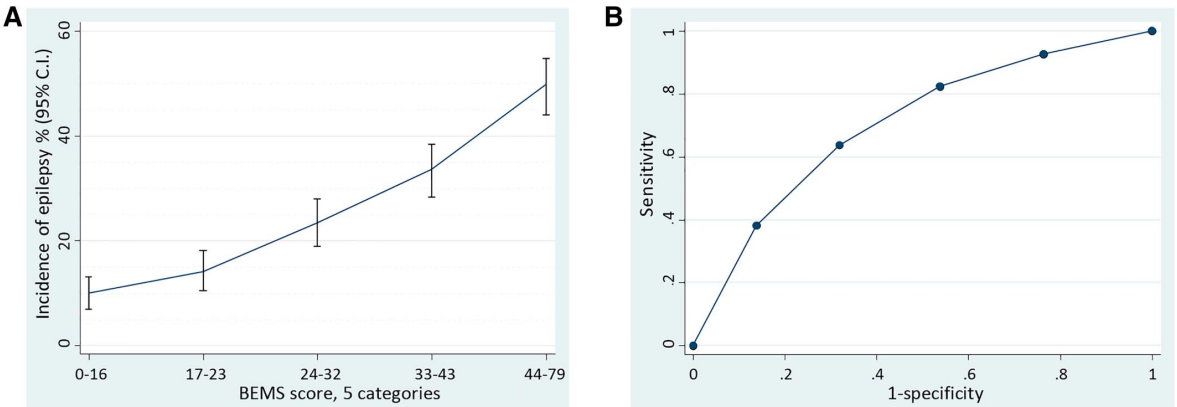

FIG. 4. **A**, Cumulative incidence of epilepsy after 6 years of follow-up according to BEMS in five quantiles. **B**, Receiver operating characteristic curve for cut points of the BEMS score in five quantiles. The area under the receiver operating characteristics curve is 0.7. BEMS, Bergen Epileptiform Morphology Score.

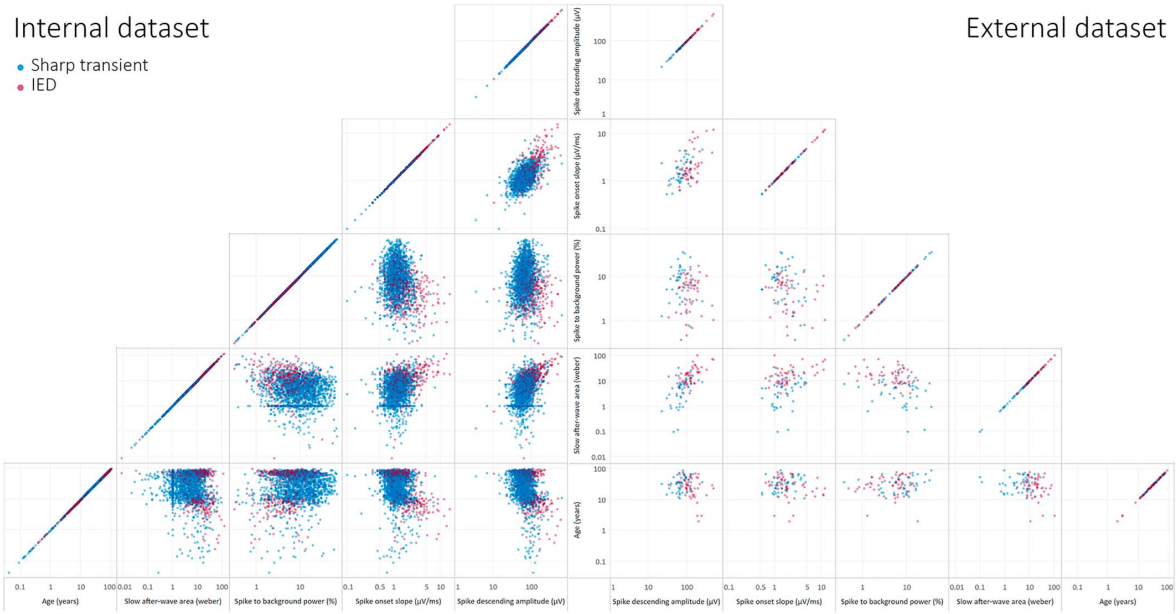

FIG. 5. Scatter plot matrix of the predictors for the internal dataset (left half) and the external dataset (right half).

epilepsy diagnosis as the outcome (Fig. 4B). The cumulative incidence of epilepsy was 60% with a BEMS score of 54 or greater.

Compared with BEMS, the clinical EEG identification of focal IEDs had better diagnostic prediction of epilepsy (see **Figure S4**, **Supplemental Digital Content 1**, http://links.lww.com/JCNP/A155). Eighty-nine percent of patients with epilepsy were diagnosed within 1 year after their EEG. The cumulative incidence of epilepsy at 6 years of follow-up was 16% in those with EEGs containing sharp transients only and 78% in those with focal IEDs. This corresponds to a sensitivity of 52%, a specificity of 95%, and an AUC of 0.73.

Validation on an External Dataset

The external and independent dataset[10] contained 100 short EEG segments from 100 patients, out of which 54 had epilepsy. The receiver operating AUC for BEMS was 0.80. A cut point of 53 on BEMS gave a specificity of 91%, sensitivity of 41%, and accuracy of 64% for a clinical and EEG-based diagnosis of epilepsy. A cut point of 29 on BEMS had a specificity of 50%, sensitivity of 98%, and accuracy of 76%. A cut point at 46, equal to the best performing cut point in our internal validation set, had a specificity of 83%, sensitivity of 57%, and accuracy of 69%. The distributions of morphologic features are shown in Fig. 5 alongside the internal dataset for comparison.

DISCUSSION

We have shown that distinct morphologic features of the first suspicious sharp discharge in an EEG can be combined into a simple score. This score predicts classification as epileptiform activity with a value similar to that of a visually assessed semiquantitative Halford score but with higher reproducibility. Application of the BEMS annotation tool on sharp EEG activity is fast (less than 1 second) and straightforward (one click on the peak). The score should be of interest to treating physicians as a higher score carries a higher risk of epilepsy. It can provide instant feedback to EEG readers in training by the score and its contributing features. Three of the criteria for epileptiform activity[9] (spike sharpness, different wave duration, and slow after-wave) are included in our new score. They were all shown to be important predictors of IEDs and for a clinical diagnosis of epilepsy. The same criteria had the highest interrater reproducibility among seven raters in a recent article by Kural et al.[10,37] and also among three raters in their successive article. A combination of spike sharpness, slow after-wave, and a dipole suggesting a source within the brain gave the highest accuracy regarding a clinical diagnosis of epilepsy (93%) in their latter paper.[37]

The score had a similar and good performance when applied to an external and independent dataset, demonstrating generalizability of BEMS. This result confirms that the included morphologic features are relevant not only locally but also where the sharp discharge selection, patient population, pretest probability, and outcome assessment differ.

Four of the features included in the score correspond to criteria in the definition of epileptiform EEG activity.[9] Whether these criteria capture the essence of IED morphology has not been proven. However, we have now shown that the first, second, and fourth criteria represent important and reproducible predictors for epileptiform activity. Spike descending amplitude and

spike onset slope are both features that correspond to the visual perception of spikiness (criterion 1). Spike to background power associates with the amount of background activity having a similar wave duration as the spike (criterion 2). Slow after-wave area is a relevant measure for slow wave prominence (criterion 4). While no single measure is able to cover all subjective interpretations of qualitative criteria, we consider these BEMS features to fit criterion 1, 2, and 4 well. Applying the score elucidates the visual features that contribute to the evaluation of sharp discharges. Such transparency should benefit EEG readers and also in their education.

Only those features that contributed to the classification of EEG outcome in the multivariate model were included in BEMS. The omitted features might be considered in future attempts to further improve the epileptiform criteria. Spike asymmetry according to Henze[31] was not statistically significant in the multivariate model. This was surprising as spike asymmetry is included in the definition of epileptiform activity. Findings from our previous study also suggested that spike asymmetry was a prominent feature in all age groups.[22] Jing et al.[38] noted that IED candidates were more likely to be scored as IEDs on visual inspection if they were asymmetric. Asymmetry seems therefore to be a characteristic feature of all IED candidates but not important to distinguish between focal IED associated with epilepsy and sharp transients.

An advantage of BEMS was its high reproducibility. The new annotation tool reported in this article is an improvement of our previous tool,[22] requiring only one mouse click instead of four. This reduces the workload for research studies and clinical use. Some of the features in the score can be assessed using any clinical EEG software. Our univariate reference ranges provide guidance for these simple features. This algorithm should be easy to implement for any EEG vendor. A click on the same IED peak will always produce identical markings and measures. Any rater differences stem only from each rater's subjective assessment in a narrow time window as to which sharp discharge is selected for analysis.

We found a poor interrater agreement of Halford score for single sharp discharges, in line with previous reports.[8] Morphology-based EEG assessment is one way of improving reproducibility. The agreement regarding whether an EEG contains any IEDs is higher than agreement regarding the evaluation of individual IEDs.[38–43] In visual EEG evaluation, a single sharp discharge is rarely sufficient to confidently conclude whether an IED is present. Our findings of imperfect AUCs for both the Halford score and the BEMS illustrate this. Definitions of IED do not include intra-EEG IED frequency or variation. However, in routine visual evaluation, consecutive sharp discharges might accumulate evidence to tip the scale in favor of IED. A measure for spike incidence may add predictive power concerning the clinical diagnosis of epilepsy, even though it is not an intrinsic morphologic feature.

This study represents a large cohort of unselected patients with epileptiform activity. A strength of this study is that we have examined one training dataset, one validation dataset, and a third independent and external dataset. This study has some limitations. Detailed information regarding how the clinical diagnosis of epilepsy was made is lacking. The EEG interpretation may have impacted the clinical conclusion. The predictive model's performance measures for the clinical epilepsy diagnosis are therefore likely to be optimistic. The study would have benefited from a broader panel of expert raters to substantiate measures of interrater reproducibility.

The designation of the first focal suspicious sharp discharge as epileptiform depends to a large degree on reproducible morphologic features that can be made into a clinical score. Best separating features were amplitude, slope, slow after-wave area, and difference from background. Duration and asymmetry did not contribute. The score was predictive of future epilepsy.

ACKNOWLEDGMENTS

The authors thank Harald Aurlien and Tom Eichele for thoughtful comments on the manuscript and Noeska Smit for contributing Fig. 5.

REFERENCES

1. Bouma HK, Labos C, Gore GC, Wolfson C, Keezer MR. The diagnostic accuracy of routine electroencephalography after a first unprovoked seizure. Eur J Neurol 2016;23:455–463.
2. Krumholz A, Shinnar S, French J, Gronseth G, Wiebe S. Evidence-based guideline: management of an unprovoked first seizure in adults: report of the guideline development subcommittee of the American Academy of Neurology and the American Epilepsy Society. Neurology 2015;85:1526–1527.
3. Koutroumanidis M, Arzimanoglou A, Caraballo R, et al. The role of EEG in the diagnosis and classification of the epilepsy syndromes: a tool for clinical practice by the ILAE Neurophysiology Task Force (Part 1). Epileptic Disord 2017;19:233–298.
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;49:231–237.
5. Stroink H, Schimsheimer RJ, de Weerd AW, et al. Interobserver reliability of visual interpretation of electroencephalograms in children with newly diagnosed seizures. Dev Med Child Neurol 2006;48:374–377.
6. Hostetler WE, Doller HJ, Homan RW. Assessment of a computer program to detect epileptiform spikes. Electroencephalogr Clin Neurophysiol 1992;83:1–11.
7. Bagheri E, Jin J, Dauwels J, Cash S, Westover MB. A fast machine learning approach to facilitate the detection of interictal epileptiform discharges in the scalp electroencephalogram. J Neurosci Methods 2019;326:108362.
8. Halford JJ, Arain A, Kalamangalam GP, et al. Characteristics of EEG interpreters associated with higher interrater agreement. J Clin Neurophysiol 2017;34:168–173.
9. Kane N, Acharya J, Benickzy S, et al. A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. clin Neurophysiol Pract 2017;2:170–185.
10. Kural MA, Duez L, Sejer Hansen V, et al. Criteria for defining interictal epileptiform discharges in EEG. A clinical validation study. Neurology 2020;94:e2139–e2147.
11. Xu Y, Nguyen D, Mohamed A, et al. Frequency of a false positive diagnosis of epilepsy: a systematic review of observational studies. Seizure 2016;41:167–174.
12. Arends JB, van der Linden I, Ebus SC, Debeij MH, Gunning BW, Zwarts MJ. Value of re-interpretation of controversial EEGs in a tertiary epilepsy clinic. Clin Neurophysiol 2017;128:661–666.
13. Benbadis SR. Errors in EEGs and the misdiagnosis of epilepsy: importance, causes, consequences, and proposed remedies. Epilepsy Behav 2007;11:257–262.

14. Benbadis SR, Kaplan PW. The dangers of over-reading an EEG. J Clin Neurophysiol 2019;36:249.
15. Gilbert DL, Sethuraman G, Kotagal U, Buncher CR. Meta-analysis of EEG test performance shows wide variation among studies. Neurology 2003;60:564–570.
16. Tjepkema-Cloostermans MC, de Carvalho RCV, van Putten M. Deep learning for detection of focal epileptiform discharges from scalp EEG recordings. Clin Neurophysiol 2018;129:2191–2196.
17. Scheuer ML, Bagic A, Wilson SB. Spike detection: inter-reader agreement and a statistical Turing test on a large data set. Clin Neurophysiol 2016;128:243–250.
18. 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.
19. Gotman J, Ives JR, Gloor P. Automatic recognition of inter-ictal epileptic activity in prolonged EEG recordings. Electroencephalogr Clin Neurophysiol 1979;46:510–520.
20. Jing J, Sun H, Kim JA, et al. Development of expert-level automated detection of epileptiform discharges during electroencephalogram interpretation. JAMA Neurol 2019;77:103–108.
21. Bagheri E, Dauwels J, Dean BC, Waters CG, Westover MB, Halford JJ. Interictal epileptiform discharge characteristics underlying expert interrater agreement. Clin Neurophysiol 2017;128:1994–2005.
22. Aanestad E, Gilhus NE, Brogger J. Interictal epileptiform discharges vary across age groups. Clin Neurophysiol 2020;131:25–33.
23. Aurlien H, Gjerde IO, Eide GE, Brogger JC, Gilhus NE. Characteristics of generalised epileptiform activity. Clin Neurophysiol 2009;120:3–10.
24. Terney D, Alving J, Skaarup CN, Wolf P, Beniczky S. The slow-wave component of the interictal epileptiform EEG discharges. Epilepsy Res 2010;90:228–233.
25. Seneviratne U, Cook M, D'Souza W. Consistent topography and amplitude symmetry are more typical than morphology of epileptiform discharges in genetic generalized epilepsy. Clin Neurophysiol 2016;127:1138–1146.
26. Holberg EEGAS. SCORE EEG™ [software]. Available at http://holbergeeg.com. Accessed October 10, 2020.
27. Tatum WO, Olga S, Ochoa JG, et al. American Clinical Neurophysiology Society guideline 7: guidelines for EEG reporting. J Clin Neurophysiol. 2016;33:328–332.
28. MATLAB. 9.4.0.949201 (R2018a) update 6. Natick, Massachusetts: The MathWorks Inc; 2018.
29. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134:9–21.
30. Frost JD Jr. Microprocessor-based EEG spike detection and quantification. Int J Biomed Comput 1979;10:357–373.
31. Henze DA, Wittner L, Buzsáki G. Single granule cells reliably discharge targets in the hippocampal CA3 network in vivo. Nat Neurosci 2002;5:790–795.
32. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 1979;74:829–836.
33. StataCorp L. Stata base reference manual release 16. Texas: Stata Press, 2019.
34. Berry SD, Ngo L, Samelson EJ, Kiel DP. Competing risk of death: an important consideration in studies of older adults. J Am Geriatr Soc 2010;58:783–787.
35. Cohen JJE. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20:37–46.
36. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996;49:1373–1379.
37. Kural MA, Tankisi H, Duez L, et al. Optimized set of criteria for defining interictal epileptiform EEG discharges. Clin Neurophysiol 2020;131:2250–2254.
38. 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.
39. Webber WR, Litt B, Lesser RP, Fisher RS, Bankman I. Automatic EEG spike detection: what should the computer imitate? Electroencephalogr Clin Neurophysiol 1993;87:364–373.
40. Gotman J, Wang LY. State dependent spike detection: validation. Electroencephalogr Clin Neurophysiol 1992;83:12–18.
41. Gotman J, Gloor P, Schaul N. Comparison of traditional reading of the EEG and automatic recognition of interictal epileptic activity. Electroencephalogr Clin Neurophysiol 1978;44:48–60.
42. Struve FA, Becka DR, Green MA, Howard A. Reliability of clinical interpretation of the electroencephalogram. Clin Electroencephalography 1975;6:54–60.
43. Houfek EE, Ellingson RJ. On the reliability of clinical EEG interpretation. J Nerv Ment Dis 1959;128:425–437.

frontiers | Frontiers in Neurology

TYPE Original Research
PUBLISHED 23 May 2023
DOI 10.3389/fneur.2023.1165592

OPEN ACCESS

EDITED BY
Giorgi Kuchukhidze,
University Hospital Salzburg, Austria

REVIEWED BY
Nishant Sinha,
University of Pennsylvania, United States
Dang Khoa Nguyen,
Montreal University, Canada

*CORRESPONDENCE
Eivind Aanestad
eivind.aanestad@helse-bergen.no

RECEIVED 14 February 2023
ACCEPTED 24 April 2023
PUBLISHED 23 May 2023

CITATION
Aanestad E, Gilhus NE, Olberg HK, Kural MA, Beniczky S and Brogger J (2023) Spike count and morphology in the classification of epileptiform discharges.
Front. Neurol. 14:1165592.
doi: 10.3389/fneur.2023.1165592

Spike count and morphology in the classification of epileptiform discharges

Eivind Aanestad[1,2]*, Nils Erik Gilhus[2,3], Henning Kristian Olberg[1,2], Mustafa Aykut Kural[4,5,6], Sándor Beniczky[4,5,6] and Jan Brogger[1]

[1]Department of Clinical Neurophysiology, Haukeland University Hospital, Bergen, Norway, [2]Department of Clinical Medicine, University of Bergen, Bergen, Norway, [3]Department of Neurology, Haukeland University Hospital, Bergen, Norway, [4]Department of Clinical Neurophysiology, Danish Epilepsy Centre Filadelfia, Dianalund, Denmark, [5]Department of Clinical Neurophysiology, Aarhus University Hospital, Aarhus, Denmark, [6]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

Purpose: The purpose of this study is to investigate the impact of Bergen Epileptiform Morphology Score (BEMS) and interictal epileptiform discharge (IED) candidate count in EEG classification.

Methods: We included 400 consecutive patients from a clinical SCORE EEG database during 2013–2017 who had focal sharp discharges in their EEG, but no previous diagnosis of epilepsy. Three blinded EEG readers marked all IED candidates. BEMS and IED candidate counts were combined to classify EEGs as epileptiform or non-epileptiform. Diagnostic performance was assessed and then validated in an external dataset.

Results: Interictal epileptiform discharge (IED) candidate count and BEMS were moderately correlated. The optimal criteria to classify an EEG as epileptiform were either one spike at BEMS >=58, two at >=47, or seven at >=36. These criteria had almost perfect inter-rater reliability (Gwet's AC1 0.96), reasonable sensitivity of 56–64%, and high specificity of 98–99%. The sensitivity was 27–37%, and the specificity was 93–97% for a follow-up diagnosis of epilepsy. In the external dataset, the sensitivity for an epileptiform EEG was 60–70%, and the specificity was 90–93%.

Conclusion: Quantified EEG spike morphology (BEMS) and IED candidate count can be combined to classify an EEG as epileptiform with high reliability but with lower sensitivity than regular visual EEG review.

KEYWORDS

epileptiform, morphology, count, quantitative, EEG, validation

Introduction

The definition of epileptiform activity includes qualitative criteria to guide EEG readers in detecting interictal epileptiform discharges (IEDs) (1). According to the criteria, typical morphological traits of an IED are a spiky peak, a wave duration that is different from the background waves, an asymmetric waveform, a slow after-wave, disrupted background activity, and a dipole suggesting that the source of the transient is in the brain. The diagnostic value of morphological IED features has been discussed extensively (2–5). Inter-rater agreement (IRA)

has been assessed for specific morphological features (2, 4), and optimal combinations of the criteria have been validated (3).

The IED criteria describe the evaluation of single transients without discussing the possible role of recurring sharp transients. It is unclear whether additional EEG phenomena such as IED counts are relevant for a diagnosis of epilepsy. An EEG conclusion should be based not only on a single graphoelement but also on all available data in an EEG recording. IRA seems to be better for overall EEG interpretation than individual IEDs (4, 6–9). The literature is sparse regarding spike count in routine scalp EEG. Kural et al. (10) demonstrated that a higher IED count is required to conclude that an EEG contains epileptiform activity when the IEDs have a less typical epileptiform morphology. Spike count has been assessed for specific epilepsy types, such as continuous spike–wave during sleep (11, 12), benign epilepsy with centro-temporal spikes (13), temporal lobe epilepsy (14), and juvenile myoclonic epilepsy (15). Latency to the first IED, a measure analogous to spike count has been examined in long-term EEG recordings (16).

We have previously described the Bergen Epileptiform Morphology Score (BEMS) (17), a score from 0 to 86 for sharp transients, where a higher value indicates a more typical epileptiform morphology. BEMS is calculated from carefully selected and visually relatable morphological IED features (spike slope, spike amplitude, spike similarity to background, and slow after-wave area) and patients' age, as IED morphology depends on age (18). BEMS classified the first sharp discharge in an EEG with a similar performance as an EEG rater both regarding the EEG conclusion (AUC = 0.86) and a future epilepsy diagnosis (AUC = 0.70). With BEMS as an established score for the single sharp discharge, it is possible to assess additional factors that may be relevant to spike identification. This study aimed to examine whether the combination of BEMS and the number of sharp discharges, referred to as IED candidate count in this study, can improve the diagnostic performance when classifying EEGs as epileptiform or non-epileptiform.

Materials and methods

Patients and EEGs

We selected a random subsample of EEGs from the total material described in our previous study (17). The original material included all consecutive patients who had standard EEGs or sleep-deprived EEGs recorded in our EEG laboratory at Haukeland University Hospital during the period of 4 March 2013–29 October 2017, which were reported in SCORE EEG (19). We included only those patients who had all their EEGs recorded at our laboratory during the inclusion period, no epileptiform activity in prior EEGs (interictal epileptiform activity or non-focal IEDs), and no prior clinical diagnosis of epilepsy (ICD-10 G40/G41 since 1999) in their hospital medical records. The first EEG for each patient that contained an epileptiform or non-epileptiform sharp discharge was analyzed. Hospital database records were examined for a clinical epilepsy diagnosis until 27 November 2019. The subsample of EEGs selected for this study was then randomized again and divided into two equally sized datasets (DS1 and DS2). DS1 was reserved as a training set to find optimal cut points for the predictor variables. An external dataset (DS3), described by the study mentioned in (10), was used for external validation. In total, 30 out of 60 patients in DS3 had epilepsy, confirmed by recording of their habitual paroxysmal events during long-term EEG monitoring (LTM). The patients included in DS3 were 1 year or older with a median age of 33 years. Their 20-min interictal EEG had to contain sharp transients. Patients were excluded if their LTM was inconclusive regarding epileptic or non-epileptic seizures.

EEG recordings

Electrodes were applied according to the 10–20 system, with a minimum of 21 and a maximum of 26 electrodes. The 26 electrode montages included three subtemporal electrodes on each side and Fpz. The recording length was 20 min for standard EEGs and 60 min for sleep-deprived EEGs. The sampling rate was 500 samples per second. NicoletOne™ EEG system was used to record and display EEGs for the clinical EEG classification which was used as an outcome, while EEGs were displayed in EEGLAB (20) for the marking of IED candidates.

IED candidates

We defined IED candidates as sharp transients that could be suspected to be IED, excluding physiological transients (21) and artifacts that mimic epileptiform discharges. Three clinical neurophysiologists, with at least 6 years of experience in EEG interpretation, marked all IED candidates chronologically in each EEG until a maximum count of 40, using a tool that has been described previously (17, 18). Only the channel in which the IED candidate had the most typical epileptiform features was marked for quantitative morphological analysis. We defined a maximum count to reduce the workload and with the assumption that higher counts would not have a significant impact on the performance of classification. In the event of spike trains or IEDs in close temporal proximity, only one distinct spiky component was analyzed per epoch of 1 s. If no IED candidate was identified in an EEG, a negative peak from the background activity on the last page of the EEG recording was marked instead. The IED candidates were marked independently by the three raters (rater 1–3) and blinded to patient data, any previous EEG markings, and the ordinary clinical EEG report. Raters 1 and 2 marked IED candidates in DS1, DS2, and DS3, while rater 3 marked in DS2 and DS3 to increase the number of raters for validation.

IED candidate-derived diagnostic markers

The following diagnostic markers were derived from the marked IED candidates in DS1:

- $BEMS_{max}$: The IED candidate with the highest BEMS in an EEG.
- $BEMS_{sum}$: The sum of BEMS for all IED candidates in one EEG.
- IED candidate count: The number of IED candidates in one EEG.
- Diagnostic classifier: We searched through combinations of three pairwise IED candidate counts and BEMS thresholds to find the combination with the highest average diagnostic accuracy and IRA when applied as three criteria-sets, where one criteria-set had to be fulfilled to classify an EEG as epileptiform. All

combinations were assessed for raters 1 and 2 in DS1 with pre-specified constraints for computational feasibility as follows: The range of BEMS was constrained between 40 and 70 points for the first criteria-set, 30 and 60 points for the second criteria-set, and 20 and 50 points for the third criteria-set. In addition, the BEMS threshold differences between criteria-sets could not be less than 10 points. A total of 5,456 combinations were assessed. The number of criteria-sets was chosen based on the combination of criteria sets given in the study by Kural et al. (10). Adding further criteria-sets was considered to be too computationally demanding, with diminishing returns regarding diagnostic performance.

In addition to the diagnostic markers, the mean BEMS for all IED candidates in one EEG, defined as $BEMS_{mean}$, was calculated.

Statistics

The diagnostic markers $BEMS_{max}$, $BEMS_{sum}$, IED candidate count, and the binary classification by the diagnostic classifier were grouped according to the EEG conclusion (focal IED or sharp transient). As a secondary outcome measure, these markers were grouped according to whether the patients were diagnosed with epilepsy or not during the follow-up. The diagnostic performance was assessed by measures of sensitivity, specificity, accuracy (the percentage that was correctly classified), and IRA. We calculated the intraclass correlation coefficient (ICC) as a measure of IRA between the raters. IRA between raters for binary classifications was calculated as Gwet's AC1 (22). Pearson's correlation coefficient was calculated for BEMS and IED candidate counts. Optimal cut points for the diagnostic markers were chosen in DS1 for raters 1 and 2 as the lowest possible value that corresponded to the highest accuracy, with specificity of >90% for the EEG conclusion. When the optimal cut points differed between the raters, the mean defined the common cut point, except for the diagnostic classifier, where a joint set of combinations was selected that maximized the sum of accuracy and Cohen's kappa for raters 1 and 2. The performance of the diagnostic classifier and the diagnostic markers was finally assessed in DS2 and DS3, regarding the EEG-conclusion of non-epileptiform transients and IED and the diagnosis of epilepsy during the follow-up. Since candidate counts were limited between 1 and 40, we calculated the estimated candidate count with a Poisson model with 1 as the lower censoring limit and 40 as the upper censoring limit.

Results

Patient and EEG characteristics

The original material from our previous study contained a total of 14,337 EEGs. A total of 4,473 EEGs were excluded due to an incomplete EEG history in the SCORE database, a previous diagnosis of epilepsy, or missing data. In total, 6,607 EEGs were excluded because no IEDs or sharp transients had been scored in the clinical report, or the EEG contained a seizure. A subsample of 400 EEGs from 400 different patients was randomly selected from the remaining 2,063 candidates and divided equally into DS1 and DS2. A total of 383 patients were analyzed after excluding 17 patients due to technical difficulties in the process of loading or reading the EEG. Patient age distributions were similar in DS1 and DS2, with a mean age of 39 years and standard deviation (SD) of 28 years (Table 1). In total, 42 out of 383 patients (11%) died during follow-up. The proportion that had EEGs containing clinically scored IEDs differed between DS1 (21%) and DS2 (13%), while those diagnosed with epilepsy were 30% in DS1 and 27% in DS2. The IED candidate count had a wide range with a maximum of 40/min. The estimated mean candidate peak rate was only 0.1–0.4/min for the three raters, which corresponds to 2–8 suspicious peaks in a 20-min EEG.

Relationship between IED candidate count and spike morphology

Spike morphology and IED candidate count had a positive association in each of the three datasets (Figure 1). $BEMS_{max}$ and IED candidate count had a correlation coefficient (CC) of 0.62 for rater 1, 0.67 for rater 2, and 0.66 for rater 3 in DS1 and DS2 combined (DS2 only for rater 3). The CC for raters 1, 2, and 3 in DS3 was 0.62, 0.60, and 0.61, respectively. $BEMS_{mean}$ and IED candidate count had a correlation of 0.41, 0.42, and 0.37 in the combined DS1 and DS2 for raters 1, 2 and 3, respectively. The significance level was $p < 0.001$ for all correlation coefficients.

Diagnostic performance in DS1

The accuracy and IRA data for $BEMS_{max}$, $BEMS_{sum}$, and IED candidate count in DS1 are shown in Figure 2. The cut points applied for all three raters were 50 for $BEMS_{max}$, 465 for $BEMS_{sum}$, and 18 for IED candidate count. The IRA was substantial for all diagnostic markers; ICC and Gwet's AC1 were 0.76 (95% CI = 0.69–0.81) and 0.88 (95% CI = 0.85–0.96) for $BEMS_{max}$, 0.68 (95% CI = 0.59–0.75) and 0.86 (95% CI = 0.80–0.93) for $BEMS_{sum}$, and 0.73 (95% CI = 0.65–0.79) and 0.90 (95% CI = 0.83–0.95) for IED candidate count, respectively. The diagnostic performance when applying the common cut points for the individual raters was as follows: For $BEMS_{max}$, the sensitivity was 64% (95% CI = 48–78), specificity was 89% (95% CI = 83–93), and accuracy was 84% (95% CI = 78–89) for rater 1 and sensitivity was 64% (95% CI = 48–78), specificity was 92% (95% CI = 86–96), and accuracy was 86% (95% CI = 80–90) for rater 2. For $BEMS_{sum}$, the sensitivity was 71% (95% CI = 55–84), specificity was 92% (95% CI = 87–96), and accuracy was 88% (95% CI = 82–92) for rater 1 and the sensitivity was 64% (95% CI = 48–78), specificity was 99% (95% CI = 95–100), and accuracy was 91% (95% CI = 87–95) for rater 2. For IED candidate count, the sensitivity was 64% (95% CI = 48–78), specificity was 96% (95% CI = 92–99), and accuracy was 89% (95% CI = 84–93) for rater 1 and the sensitivity was 60% (95% CI = 43–74), specificity was 99% (95% CI = 95–100), and accuracy was 90% (95% CI = 85–94) for rater 2. The diagnostic classifier with the highest combined accuracy and IRA for both raters was as follows: One IED candidate with BEMS >= 58, two IED candidates with BEMS >= 47, or seven IED candidates with BEMS >= 36. The sensitivity, specificity, and accuracy were 67% (95% CI = 51–80), 92% (95% CI = 86–95), and 86% (95% CI = 81–91) for rater 1 and 62% (95% CI = 46–76), 96% (95% CI = 92–99), and 89%

10.3389/fneur.2023.1165592

TABLE 1 Patient and EEG characteristics.

	DS1			DS2			Total		
Sample size, *n*=	196			187			383		
Age in years, mean (SD) (min-max)	37.3	(28.3)	(0–100)	40.5	(26.7)	(0–94)	38.9	(27.5)	(0–100)
Death rate during follow-up, %	12.8			9.1			11.0		
EEG-conclusion IED, %	21.4			13.4			17.5		
Epilepsy at follow-up, %	30.1			27.3			28.7		
IED candidates per minute for rater 1 (Censored), mean (min-max)	1.4	(0.0–25.6)		1.1	(0.0–23.2)		1.2	(0.0–25.6)	
IED candidates per minute for rater 1 (Uncensored), mean	0.36			0.34			0.35		
IED candidates per minute for rater 2 (Censored), mean (min-max)	1.3	(0.0–39.6)		0.6	(0.0–16.8)		1.0	(0.0–39.6)	
IED candidates per minute for rater 2 (Uncensored), mean	0.12			0.09			0.10		
IED candidates per minute for rater 3 (Censored), mean (min-max)	*			1.0	(0.0–22.9)		1.0	(0.0–22.9)	
IED candidates per minute for rater 3 (Uncensored), mean	*			0.14			0.14		

*: Rater 3 did not mark IED candidates in DS1.

(95% CI = 84–93) for rater 2, respectively. IRA was almost perfect with Gwet's AC1 = 0.89 (95% CI = 0.83–0.95).

Diagnostic performance in DS2 and DS3

Diagnostic performance for the various markers in DS1 and DS2 is shown in Table 2. Gwet's AC1 was >0.89 for all diagnostic markers in DS2 and varied from 0.57 to 0.73 in DS3. The diagnostic classifier had the highest IRA in both datasets with Gwet's AC1 of 0.96 in DS2 and 0.73 in DS3. The sensitivity for the markers was 60–67% in DS2 and 33–70% in DS3. The specificity was 96–99% in DS2 and 86–91% in DS3. Supplemental Digital Content 1 shows rater-specific performance measures with 95% confidence intervals. The diagnostic performance of our current standard of care, the clinical EEG conclusion, which also served as a reference standard for the diagnostic markers, had a sensitivity of 41% (95% CI = 28–56) and a specificity of 97% (95% CI = 93–99) for the follow-up diagnosis of epilepsy in DS2.

Discussion

We have shown in a large study of routine scalp EEGs that there is a positive correlation between characteristic epileptiform IED morphology and IED candidate count. EEGs with distinct epileptiform discharges had a high IED candidate count, while EEGs with less characteristic epileptiform activity as defined by BEMS had a lower IED candidate count. Quantified EEG spike morphology (BEMS) and IED candidate count can be combined to classify an EEG as epileptiform with high reliability but with somewhat lower sensitivity than regular visual EEG review.

Interictal epileptiform discharge (IED) candidate count was an important predictor of epileptiform activity in our study. We suggest that the IED candidate count should be added in a future update of the criteria for epileptiform discharges that was proposed by the International Federation of Clinical Neurophysiology (IFCN). IED candidate count is not an inherent property of epileptiform morphology but rather provides a context for its interpretation. We suspect that a higher IED candidate count strengthens the EEG

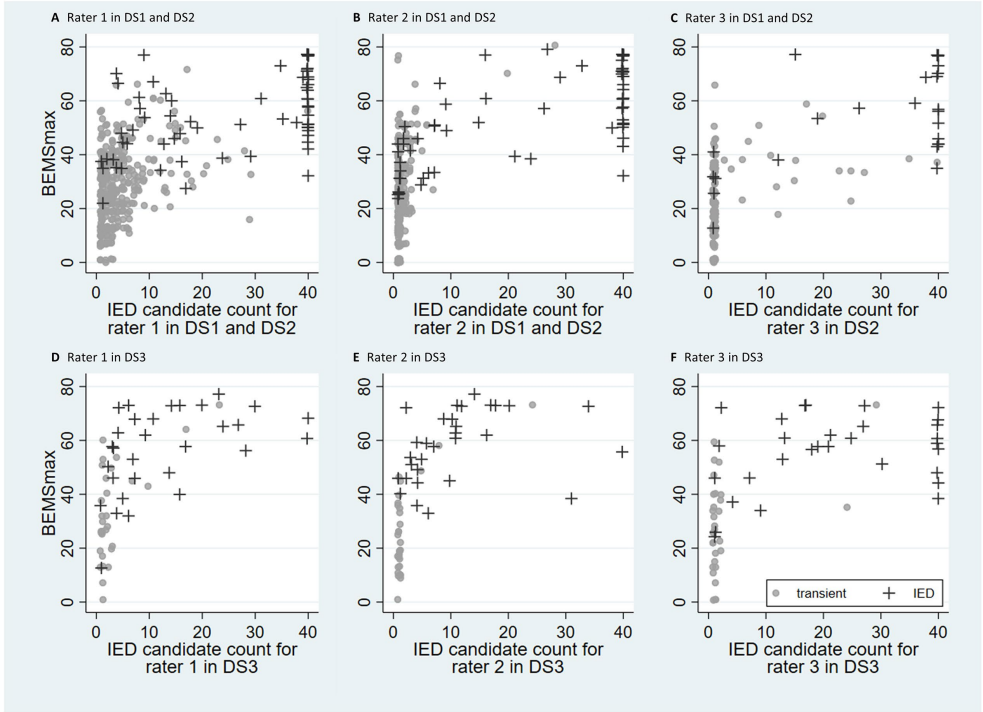

FIGURE 1
Scatter plots with IED candidate count and $BEMS_{max}$ for the three raters. The datasets DS1 and DS2 (n=383) are shown in **(A–C)** (rater 3 examined DS2 only). The dataset DS3 (n=60) is shown in **(D–F)**. Observations are labeled according to epileptiform (+) or non-epileptiform (●) EEG in DS1 and DS2, and epilepsy (+) or not epilepsy (●) in DS3. Jitter (0.4) has been added to increase visibility of overlapping symbols. Most EEGs in DS1 and DS2 **(A–C)** fell into two clusters, similar for all three raters. The cluster in the lower left corner contains EEGs with infrequent IED candidates and a low $BEMS_{max}$, while the cluster in the upper right corner contains EEGs with frequent IED candidates with a high $BEMS_{max}$. A minority of the EEGs were scattered between the two clusters. This two-cluster pattern was less evident for DS3 **(D–F)**.

reader's confidence in spike detection and reduces the likelihood of false positives, a well-known challenge in visual spike detection (23). Background noise can imitate epileptiform activity once or twice but not repetitively. Some physiological sharp transients can occur repeatedly but have recognizable morphology (21). When spotting a definite epileptiform discharge by visual interpretation, less prominent discharges in the same region are more easily included as IEDs. It depends on the signal-to-noise ratio whether one or more discharges are needed to distinguish a spike from background activity. Signal averaging is a well-known method in signal analysis and applies the same principle. Each addition of raw signal to the running average flattens background noise while the signal of interest remains unchanged. While one epileptiform discharge might fulfill only a few IFCN criteria, the average of many discharges meets more criteria, and morphological uncertainties are eliminated.

We have described the IED candidate count in a large dataset where all EEGs had at least one epileptiform or non-epileptiform sharp discharge scored at the time of the clinical EEG report. The average IED candidate count was estimated to be between two and eight per 20-min EEG. IED candidate count is entirely based on visual interpretation of scalp EEG and does not reveal the intracranial or "true" spike count. Some intracranial IEDs are not detected in scalp EEG (24). Their visibility depends on variables, such as source depth, cortical area, and geometry (25).

Our study had a high IRA. To be able to test the reproducibility of our classification model, we divided our internal EEGs into two independent data sets (DS1 and DS2) and used the second data set for validation, also including a third EEG rater. The inter-rater agreement was substantial to almost perfect between the three raters for the diagnostic classifier, demonstrating robustness regarding variability in the selection of IED candidates between the raters. Our assessment of IED candidate morphology is entirely objective by applying the algorithm for the BEMS score, analogous to a subjective visual assessment of epileptiform criteria. The BEMS algorithm is publicly available for use by equipment manufacturers.

The diagnostic classifier was built using a traditional and explainable analytic approach that classified EEGs as epileptiform or non-epileptiform with high reproducibility and specificity but with lower sensitivity than a routine clinical EEG examination. Possible explanations for the limited sensitivity could be that the BEMS score did not capture enough information per IED candidate or that the clinical information available to the clinical interpreter of the EEG

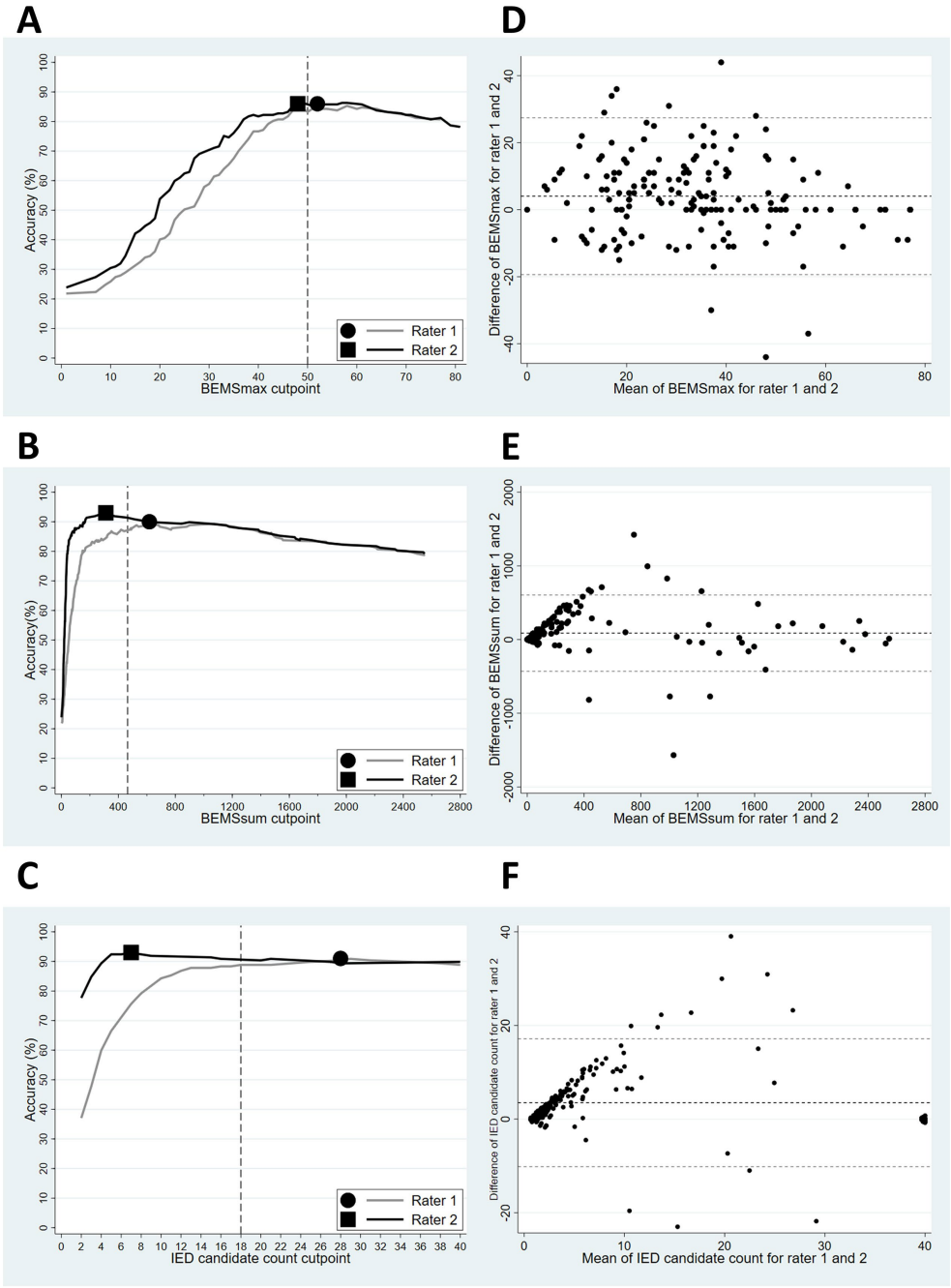

FIGURE 2
(A–C): Accuracy for $BEMS_{max}$ **(A)**, $BEMS_{sum}$ **(B)** and IED candidate count **(C)**, with the EEG conclusion as outcome for rater 1 and 2 in DS1. The optimal rater specific cut points are indicated by a dot (rater 1) and a square (rater 2). The vertical dashed lines indicate common cut points. **(D–F)**: Bland–Altman plots for $BEMS_{max}$ **(D)**, $BEMS_{sum}$ **(E)**, and IED candidate count **(F)** for rater 1 and 2 in DS1. Mean values are plotted along the *x*-axis, mean differences between rater 1 and 2 along the *y*-axis. $BEMS_{max}$ had the most evenly distributed means between rater 1 and 2, while $BEMS_{sum}$ and IED candidate count had increasing differences between the two raters for higher means.

contained decisive information to tip the scales. We found a gray area of ambiguous EEGs that had neither infrequent and unconvincing IED candidates nor numerous and highly epileptiform IED candidates. Experienced clinical EEG interpreters can add diagnostic value in such cases.

Future studies of EEG interpretation should focus on difficult borderline EEGs.

There are several limitations to this study. We did not have access or the capacity to analyze clinical and paraclinical patient data that can be thought to influence or explain BEMS and IED candidate count, e.g., what evidence was available to the clinicians that diagnosed the patients with epilepsy, type of epilepsy syndrome, seizure burden, use of anti-seizure medication, imaging data, neurological comorbidities, IED candidate localization, and topography. Individual rater threshold differences were relatively large for IED candidate count and $BEMS_{sum}$ (Figure 2), affecting the diagnostic performance negatively since a common threshold will differ from each of the rater's optimal cut point. The IED candidate count threshold differences imply that the difficulties when deciding whether an EEG waveform is an IED candidate are comparable to that of IED classification. The low inter-rater threshold difference and healthy Bland–Altman plot for $BEMS_{max}$ (Figure 2) suggest that the IED candidate selection by a human rater combined with automated quantitative BEMS scores reliably identifies the IED candidate with the most typical epileptiform characteristics in an EEG.

We validated the diagnostic performance and IRA of the diagnostic markers by examining an external EEG dataset (DS3) that had different patient characteristics, prevalence of positive outcomes, and reference standards. DS3 consisted of patients who had required long-term video-EEG monitoring (LTM) in their work-up, as opposed to our internal datasets which included routine EEGs from a wide variety of referrers and reasons for requesting an EEG. The pretest IED probability was lower in DS1 and DS2 compared with DS3. The low prevalence of focal IEDs in our internal datasets approximates the actual prevalence in the patient population that is referred to our EEG laboratory for a routine EEG, which can be estimated at 8% from our previous studies (18). Validating the diagnostic markers on DS3 was a "trial by fire" due to the different dataset characteristics outlined above. The optimal cut point for any quantitative diagnostic marker depends on pretest probability, and its application will be more suitable for similar datasets. The gold standard in DS3, the classification of habitual seizures as epileptic or non-epileptic by LTM, is an outcome measure of a higher standard than those in DS1 and DS2, which were focal IEDs as classified by the attending physician and the presence of a follow-up diagnosis of epilepsy in the hospital database records.

Conclusion

Interictal epileptiform discharge (IED) candidate count is a relevant predictor variable in the classification of EEGs as epileptiform or non-epileptiform. IED candidate count correlated positively with IED candidate morphology. Based on our data, we suggest the following criteria for definite interictal epileptiform activity: either at least one very typical epileptiform discharge (BEMS> = 58), or at least two moderately typical epileptiform discharges (BEMS> = 47), or at least seven less distinct epileptiform discharges (BEMS> = 36).

Data availability statement

The datasets presented in this article are not readily available because of ethical and privacy restrictions. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by REK vest, Universitetet i Bergen, Det medisinske fakultet, Postboks 7,804, 5,020 Bergen. The reference code for our research project is "2017/1512/REK vest." Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

TABLE 2 Mean performance of diagnostic markers between 3 raters when applying the common cut points that were developed in DS1 for rater 1 and 2.

	DS2 (*N*=187)					DS3 (*N*=60)		
	Interrater agreement	EEG conclusion		Epilepsy		Interrater agreement	Epilepsy	
	Gwet's AC1	Sensitivity %	Specificity %	Sensitivity %	Specificity %	Gwet's AC1	Sensitivity %	Specificity %
BEMSmax*	0.90	60	96	29	95	0.57	70	86
BEMSsum**	0.94	67	98	32	98	0.69	56	94
IED candidate count***	0.93	60	98	29	97	0.70	33	96
Diagnostic classifier****	0.96	60	99	26	97	0.73	63	91
Clinical EEG conclusion				51	96			

Asterisks indicate appliance of common cutpoints. *: BEMSmax > = 50. **: BEMSsum > = 465. ***: IED candidate count > = 18. ****: either 1 IED candidate with BEMS > = 58, 2 with BEMS > = 47 or 7 with BEMS > = 36.

Author contributions

EA drafted the manuscript. EA, NG, and JB were responsible for the study design, methodology, and revising the manuscript. EA, JB, and HO collected data for the internal dataset. SB and MK provided and organized the external dataset. EA and JB performed the statistical analysis. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Helse Vest, project number F-10226.

Conflict of interest

JB, EA and HO are minority shareholders in Holberg EEG AS, the providers of the SCORE EEG software used in this study.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2023.1165592/full#supplementary-material

References

1. 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) 2:170–5. doi: 10.1016/j.cnp.2017.07.002

2. 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:e2139–e2147. doi: 10.1212/WNL.0000000000009439

3. Kural MA, Tankisi H, Duez L, Sejer Hansen V, Udupi A, Wennberg R, et al. Optimized set of criteria for defining Interictal Epileptiform Eeg discharges. *Clin Neurophysiol*. (2020) 131:2250–4. doi: 10.1016/j.clinph.2020.06.026

4. Jing J, Herlopian A, Karakis I, Ng M, Halford JJ, Lam A, et al. Interrater reliability of experts in identifying Interictal Epileptiform discharges in electroencephalograms. *JAMA Neurol*. (2020) 77:49–57. doi: 10.1001/jamaneurol.2019.3531

5. Tatum WO, Shellhaas RA. Epileptiform discharges. *Neurology*. (2020) 94:862–3. doi: 10.1212/WNL.0000000000009432

6. Gotman J, Gloor P, Schaul N. Comparison of traditional Reading of the Eeg and automatic recognition of Interictal epileptic activity. *Electroencephalogr Clin Neurophysiol*. (1978) 44:48–60. doi: 10.1016/0013-4694(78)90104-9

7. Gotman J, Wang L-Y. State dependent spike detection: validation. *Electroencephalogr Clin Neurophysiol*. (1992) 83:12–8. doi: 10.1016/0013-4694(92)90127-4

8. Houfek EE, Ellingson RJ. On the reliability of clinical Eeg interpretation. *J Nerv Ment Dis*. (1959) 128:425–7. doi: 10.1097/00005053-195905000-00006

9. Webber WR, Litt B, Lesser RP, Fisher RS, Bankman I. Automatic Eeg spike detection: what should the computer imitate? *Electroencephalogr Clin Neurophysiol*. (1993) 87:364–3. doi: 10.1016/0013-4694(93)90149-P

10. Aykut Kural M, Qerama E, Johnsen B, Fuchs S, Beniczky S. The influence of the abundance and morphology of Epileptiform discharges on diagnostic accuracy: how many spikes you need to spot in an Eeg. *Clin Neurophysiol*. (2021) 132:1543–9. doi: 10.1016/j.clinph.2021.03.045

11. Larsson PG, Wilson J, Eeg-Olofsson O. A new method for quantification and assessment of Epileptiform activity in Eeg with special reference to focal nocturnal Epileptiform activity. *Brain Topogr*. (2009) 22:52–9. doi: 10.1007/s10548-008-0072-3

12. Reus EEM, Visser GH, Cox FME. Determining the spike–wave index using automated detection software. *Publish Ahead of Print*. (2020) 38:198–201. doi: 10.1097/wnp.0000000000000672

13. Chavakula V, Sánchez Fernández I, Peters JM, Popli G, Bosl W, Rakhade S, et al. Automated quantification of spikes. *Epilepsy Behav*. (2013) 26:143–2. doi: 10.1016/j.yebeh.2012.11.048

14. Clemens Z, Janszky J, Clemens B, Szucs A, Halasz P. Factors affecting spiking related to sleep and wake states in temporal lobe epilepsy (Tle). *Seizure*. (2005) 14:52–7. doi: 10.1016/j.seizure.2004.09.003

15. Turco F, Bonanni E, Milano C, Pizzanelli C, Steinwurzel C, Morganti R, et al. Prolonged epileptic discharges predict seizure recurrence in Jme: insights from prolonged ambulatory Eeg. *Epilepsia*. (2021) 62:1184–92. doi: 10.1111/epi.16875

16. Werhahn KJ, Hartl E, Hamann K, Breimhorst M, Noachtar S. Latency of Interictal Epileptiform discharges in long-term Eeg recordings in epilepsy patients. *Seizure*. (2015) 29:20–5. doi: 10.1016/j.seizure.2015.03.012

17. Aanestad E, Gilhus NE, Brogger J. A new score for sharp discharges in the Eeg predicts epilepsy. *J Clin Neurophysiol*. (2023) 40:9–16. doi: 10.1097/WNP.0000000000000849

18. Aanestad E, Gilhus NE, Brogger J. Interictal Epileptiform discharges vary across age groups. *Clin Neurophysiol*. (2020) 131:25–33. doi: 10.1016/j.clinph.2019.09.017

19. Holberg EEG AS. Score Eeg™ [software]. Available at: http://Holbergeeg.Com (Accessed May 8, 2023).

20. Delorme A, Makeig S. Eeglab: an open source toolbox for analysis of single-trial Eeg dynamics including independent component analysis. *J Neurosci Methods*. (2004) 134:9–21. doi: 10.1016/j.jneumeth.2003.10.009

21. Wüstenhagen S, Terney D, Gardella E, Meritam Larsen P, Rømer C, Aurlien H, et al. Eeg Normal variants: a prospective study using the score system. *Clin Neurophysiol Pract*. (2022) 7:183–0. doi: 10.1016/j.cnp.2022.06.001

22. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's kappa and Gwet's Ac1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol*. (2013) 13:61. doi: 10.1186/1471-2288-13-61

23. Benbadis SR, Kaplan PW. The dangers of over-Reading an Eeg. *J Clin Neurophysiol*. (2019) 36:249. doi: 10.1097/wnp.0000000000000598

24. Pyrzowski J, Le Douget J-E, Fouad A, Siemiński M, Jędrzejczak J, Le Van QM. Zero-crossing patterns reveal subtle Epileptiform discharges in the scalp Eeg. *Sci Rep*. (2021) 11:4128. doi: 10.1038/s41598-021-83337-3

25. Ramantani G, Maillard L, Koessler L. Correlation of invasive Eeg and scalp Eeg. *Seizure*. (2016) 41:196–0. doi: 10.1016/j.seizure.2016.05.018

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230860113 (print)
9788230845202 (PDF)