Clinical Neurophysiology 126 (2015) 1124–1131

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units

F. Fürbass [a,*], P. Ossenblok [c], M. Hartmann [a], H. Perko [a], A.M. Skupch [a], G. Lindinger [d], L. Elezi [b], E. Pataraia [d], A.J. Colon [e], C. Baumgartner [b], T. Kluge [a]

[a] Department Safety & Security, AIT Austrian Institute of Technology GmbH, Vienna, Austria
[b] General Hospital Hietzing with Neurological Center Rosenhuegel, 2nd Neurological Department, Vienna, Austria
[c] Clinical Physics, Academic Center of Epileptology, Heeze & Maastricht UMC+, The Netherlands
[d] Department of Clinical Neurology, Medical University of Vienna, Vienna, Austria
[e] Department of Clinical Neurophysiology, Academic Center of Epileptology, Heeze & Maastricht UMC+, PO Box 61, 5590 AB Heeze, The Netherlands

ARTICLE INFO

Article history:
Accepted 24 September 2014
Available online 2 October 2014

Keywords:
Epileptic seizure detection
Automatic
Online
Parameter free
Alarm device
Prospective multi-center study

HIGHLIGHTS

- Large prospective multi-center study of an automatic seizure detection system including 205 patients.
- Comparison between two automatic seizure detection systems using the same prospectively recorded dataset.
- Performance numbers on the publicly available CHB–MIT dataset and on 310 retrospective patients datasets.

ABSTRACT

Objective: A method for automatic detection of epileptic seizures in long-term scalp-EEG recordings called EpiScan will be presented. EpiScan is used as alarm device to notify medical staff of epilepsy monitoring units (EMUs) in case of a seizure.
Methods: A prospective multi-center study was performed in three EMUs including 205 patients. A comparison between EpiScan and the Persyst seizure detector on the prospective data will be presented. In addition, the detection results of EpiScan on retrospective EEG data of 310 patients and the public available CHB–MIT dataset will be shown.
Results: A detection sensitivity of 81% was reached for unequivocal electrographic seizures with false alarm rate of only 7 per day. No statistical significant differences in the detection sensitivities could be found between the centers. The comparison to the Persyst seizure detector showed a lower false alarm rate of EpiScan but the difference was not of statistical significance.
Conclusions: The automatic seizure detection method EpiScan showed high sensitivity and low false alarm rate in a prospective multi-center study on a large number of patients.
Significance: The application as seizure alarm device in EMUs becomes feasible and will raise the efficiency of video-EEG monitoring and the safety levels of patients.

1. Introduction

Long-term video EEG-monitoring in epilepsy monitoring units (EMUs) plays a central role in pre-surgical evaluation of patients with epilepsy (Smith, 2005). This time-consuming procedure lasting for several days up to weeks requires high effort from staff to ensure patient safety and to evaluate the high amount of data. Safety in EMUs is an on-going discussion. It is generally accepted that precautions have to be in place to promptly detect seizures (Carlson, 2009) and to avoid additional harm to the patients. A study by Atkinson et al. (2012) with $N = 20$ patients showed that only 40% of seizures showed staff response. Changing the safety

* Corresponding author at: AIT Austrian Institute of Technology, Donau-City-Straße 1, 1220 Vienna, Austria. Tel.: +43 50550 4230; fax: +43 50550 4125.
E-mail address: franz.fuerbass@ait.ac.at (F. Fürbass).

http://dx.doi.org/10.1016/j.clinph.2014.09.023
1388-2457/

protocol for EMUs can thus lead to a decrease in patient accidents and an increase in detected seizures (Spanaki et al., 2012).

Automatic epileptic seizure detection (ESD) is one method to improve patient safety and efficiency in the EMU. Although these systems have a long history of numerous methodical approaches that proved to be effective in some trials (Gotman and Gloor, 1976; Gotman, 1982, 1990) wide spread clinical application were not accomplished until now. Today, the small number of epilepsy monitoring units using such systems stays in contrast with the increasing awareness of patient security issues during long-term recording and the high costs of this examination method. A low false alarm rate is of major importance for alarm systems to avoid ignorance by staff as found by Lee and Shah (2013). Many epilepsy centers do not use automatic seizure detection systems because of a very high number of false detections.

Several publications proposed patient specific seizure detectors or detectors for certain seizure patterns (Beniczky et al., 2013). These approaches will be of limited value in clinical practice because details of the type of epilepsy or the localization of the seizure onset zone (SOZ) are mostly unknown. Attempts to use the first seizure of a patient for patient specific detectors are limited because of the long time delay to the first seizure. Several studies reported a delay between 2 and 3.7 days in EMUs for pre-surgical evaluation, depending on the type of epilepsy (Todorov et al., 1994). In addition, the average number of seizures that can be recorded in one week of video EEG is rather small (median of 3 in one week in our data). Furthermore, it is important to detect whether or not a patient has one or multiple types of seizures. This implies that detection systems cannot be efficiently trained or configured for patients in the EMU and that only parameter-free detection systems without restriction to seizure types are feasible.

Automatic analysis of the EEG can be done either ad-hoc during the recording of the patient or post hoc after the patient recoding has finished. These situations are also referred to as "online" or "offline" detection, respectively. This article will solely present results of the online seizure detector EpiScan but the major differences to offline detectors are depicted shortly to allow objective comparison to other publications. First of all, because online detection systems may be used as alarm devices whereas offline systems support the EEG evaluation after recording. Furthermore, online detection systems must have a very short time delay to trigger alarms. An artificially delayed alarm allows the collection of information about the trend of the supposed seizure and can avoid false alarms. A system reacting in the range of a few seconds is more close to an alarm device, whereas a system with a detection delay of several minutes or hours behaves like a typical post hoc system. When comparing the performance of ad-hoc to post hoc systems or ad-hoc systems with different delays care has to be taken.

The amount and kind of data to evaluate an automatic seizure detection system is an important and frequently discussed issue. A sufficient number of long-term patient recordings are needed in order to draw reliable conclusions about sensitivity, specificity or the differentiation between two competing systems or datasets. One critical point in assessment of seizure detectors is the estimation of the sensitivity. Seizures are rare events with high inter- and intra-patient variability. The detection sensitivity of an automatic system represents a random variable with high variance and unknown distribution. In statistics the central limit theorem states that a sampling distribution approaches the normal distribution if the sample size is sufficient, no matter how the population distribution was shaped. A sample size of $N = 30$ is considered as appropriate for moderately skewed population distributions and will give a rough estimate of the performance. Population distributions far from normal need a sample size of $N = 500$ or more. For the sensitivity and false alarm rate of a seizure detection system we cannot assume a distribution close to normal and thus have to carefully determine the amount of data necessary to get significant results.

However, sensitivity based on a high number of patients alone does not validate a clinical application if only parts of the recordings are used. Only complete and uncut datasets reflect the real clinical situation and can prove sensitivity and specificity at the same time. A detection system may easily be able to detect 100% of the seizures in a dataset when only ictal EEG fragments are used but will show an excessive false alarm rate when evaluated on full long-term recordings. In addition, changes of the EEG during the day/night cycle need to be included in the evaluation leading to a necessary continuous recording length of more than 24 h.

The Computational Encephalography research group (www.eeg-vienna.com) of the Austrian Institute of Technology (AIT) has developed an automatic seizure detection system for long-term scalp EEG recordings called EpiScan. The detection algorithm of EpiScan works as an alarm device which allows notification of medical staff in case of a seizure. The system does not require parameters or patients specific settings. In this article the results of a prospective multi-center study will be presented. The results of EpiScan will be compared to the results of the Persyst seizure detector using the same prospective dataset. A comparison to the EpiScan performance on the development dataset and the MIT–CHB dataset will be carried out.

2. Methods

2.1. Data analysis

EpiScan is based on a computational method, which automatically detects epileptic seizures in digitized EEG. This method was developed over several years by a team of physicians, mathematicians and medical experts (Schachinger et al., 2006; Perko et al., 2007; Kluge et al., 2009; Hartmann et al., 2011; Fürbass et al., 2012). It is intended to analyze the EEG ad-hoc and to act as an online detection system. The EpiScan method analyses the digital EEG during recording in intervals of a quarter-second. Frequencies below 0.7 Hz and above 99 Hz are removed by finite impulse response filters. Line noise is removed with notch filters at 50 and 60 Hz. EEG segments with artifacts like i.e. excessive amplitudes or artifacts from loose electrodes are removed automatically (Skupch et al., 2013) and are not used for detection. This will avoid false alarms based on measurement problems. The EEG is then scanned for rhythmic patterns in the time and frequency domain by algorithms called Epileptiform Wave Sequence Analysis (EWS) and Periodic Waveform Analysis (PWA), respectively (Hartmann et al., 2011; Fürbass et al., 2012). An energy detector scans for tonic or tonic–clonic seizures with strong muscle artifacts. All extracted features are normalized by a spatio-spectral model of the brain activity that is continuously updated by past information from the EEG. A set of classifiers is used to remove events with physiological origin. The use of these adaption and classification algorithms avoids repeated detections of physiological or pathological patterns that are no seizures and is therefore another important mechanism to avoid excessive false alarms. The parameters of the classifiers were optimized using an automatic parameter optimization method (Dollfuss et al., 2013).

2.2. Quantity and quality of data needed for evaluation

The amount of data in a study is a critical parameter for the reliability of the results. Standard measures in statistics like i.e. the mean or confidence intervals of a result assume a sufficient high number of replicates in order to be valid estimates. An objective estimate of the number of participants for a seizure detector study

is hardly feasible but it is easy to show that $N < 30$ is too low. Viewing the problem from the neurophysiological perspective it has to be considered that epilepsy is a symptomatic disease with numerous etiologies. Thus 30 patients will not adequately represent the full range of possible manifestations (see experiment using virtual trials in supplementary data).

2.3. Definition of seizures

An epileptic seizure is defined as a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain (Fisher et al., 2005). Although the EEG is an important tool for epilepsy diagnosis this definition does not state how well the seizure activity can be identified in the EEG. Some seizures are hardly recognized without using additional information from video or other clinical information because of artifact overlap or subtle EEG patterns. Furthermore, clinical practice often includes subtle seizure-like events in the list of seizures to support the neurologist.

In order to remove the bias of clinical procedures and EEG measurement issues, it is common to restrict the evaluation of EEG-based epileptic seizure detection systems on clearly visible electrographic seizures. In such an approach one or several experienced EEG reviewers select seizures according to a visual perception value and define seizure onset and duration. Such a pre-sorting of seizures is preferable during development of the detection system but not appropriate in the clinical practice of EMUs.

To take this into account we used a two-step procedure to evaluate our seizure detection system. In a first step the detection performance of EpiScan was assessed using seizures defined by clinical and electrographic observations without restriction to EEG correlates. This includes all seizures that were marked during recording of the EEG and seizures that were found retrospectively by the standard EMU review procedure using video, EEG, and observation reports from nursery staff. Results using seizures from this first step are referred to as $C + E$ evaluation group. In a second step the detection performance for seizures with different levels of EEG perception value (P) will be given. The perception values of seizures were assigned from experienced EEG technicians in several video-blinded reviewing sessions. The reviewers were asked to decide whether the EEG at a defined time point shows a seizure. They were allowed to switch montages and to review the EEG before and after the given time point. The possible answers included six levels of increasing perceptions values (see Table 1). Based on these six perception values data were divided to form four groups called $C + E$, E75, E50, and E25 for the evaluation of the seizure detection systems. The evaluation groups E75, E50 and E25 include all seizures that had at least at perception value of >75%, >50% or >25%, respectively (see supplementary data for examples). Reviewers will often use the middle of a scale or 50% if they are uncertain about the decision. This case was avoided by forcing a decision between "rather a seizure" (rather yes) or "rather not a seizure" (rather no).

2.4. Dataset

2.4.1. Prospective multi-center study

A prospective multi-center study was performed to evaluate the seizure alarm system EpiScan. During the study, long-term EEG recordings from 205 consecutive patients were evaluated. Data were recorded at three epilepsy-monitoring units, the 2nd Neurological Department of the General Hospital Hietzing with Neurological Center Rosenhuegel in Vienna (NCR), the Department of Clinical Neurology of the Medical University of Vienna (MUV) and the Epilepsy Center Kempenhaeghe in Heeze, the Netherlands (KEMP). Data were recorded between January 2012 and March 2013. All centers used the international 10–20 electrode placement system for data recording. The data was recorded using a sampling rate of 256 Hz in center MUV and NCR and a sampling rate of 200 Hz in center KEMP. The inclusion criteria were a signed patient agreement form and an age above 18.

An ITmed EEG recording machine was used in center MUV. Patients had to stay in bed to allow video-EEG in this facility. The center NCR uses a Micromed recording system including a headbox with internal memory that allowed unplugging of several minutes without loss of EEG. Due to technical reasons it was not possible to use the EEG in the unplugged time periods for the study (about 3% of the recorded data at NCR). The center KEMP uses a Stellate recording device with long patch cables. Here all patients stayed in a living-room like environment that allowed free movement. They were able to use fitness devices or the bathroom without disruption of the EEG recording. This environment induced lots of additional movement artifacts namely from cycling, chewing, and tooth brushing making this dataset especially challenging to analyze for an automatic detection system. In all three centers AEDs were withdrawn preceding or during the five day period of video-EEG monitoring depending on the patient. The amount of data that was collected at each center as well as number of patients is summarized in Table 2.

2.4.2. Retrospective data

EpiScan was developed using a dataset of 310 patients. This dataset will also be evaluated in this article to further increase

Table 2
Overview of the EEG data included in the prospective multicenter study. In total 205 patients participated, including 94 patient with seizures. The number of seizures and hours of recorded EEG is given.

Epilepsy center	N	N with sz.	Number of sz.	Hours of EEG
EEG data of the prospective study				
NCR	83	27	142	6513
KEMP	60	47	211	5127
MUV	62	20	173	4044

Table 1
The EEG perception value (P) of a seizure. Seizure markers received from EMUs have no perception value ($C + E$). A seizure perception value is assigned through EEG reviewing, the higher the more clearly a pattern was perceived as seizure in the EEG. An evaluation group includes all seizures that have at least a given minimum perception value.

Q: Is this EEG pattern a seizure?		Evaluation groups with included seizure			
Possible answer	P (%)	E75	E50	E25	$C + E$
Seizure perception value					
Surely yes	100–90	x	x	x	x
Probably yes	90–75	x	x	x	x
Rather yes	75–50		x	x	x
Rather no	50–25			x	x
Probably no	25–10				x
Surely no	10–0				x

the statistical relevance of the results and to show differences and similarities between a retrospective and prospective dataset. The development dataset was recorded at several different EMUs using the international 10–20 electrode placement system at a sampling rate of 256 Hz. The dataset included 693 markers that were a mixture of relevant information for the diagnosis and real seizures. The EMU review procedure on this development data had not been standardized for the use in a clinical study. Therefore a seizure evaluation group $C+E$ is undefined and thus no evaluation on $C+E$ seizures could be done. The dataset was evaluated retrospectively and the same protocol as for the prospective dataset to define seizures for the E25, E50, and E75 evaluation groups was applied.

Although EpiScan was developed and tested on data from patients with age above 18 an evaluation on a small pediatric dataset gives a first insight if a clinical application will be feasible. The CHB–MIT scalp EEG database was used. It was created by a team of investigators from Children's Hospital Boston (CHB) and the Massachusetts Institute of Technology (MIT) and is publicly available from the Physionet website (http://www.physionet.org/physiobank/database/chbmit/). The database includes data from 24 patients from 1.5 years to 22 years of age and a mean of 10 years (Goldberger et al., 2000; Shoeb, 2009; Hunyadi et al., 2012). Seizure markers were uses as given in the dataset without assigning perception values. Unfortunately, the CHB–MIT data is given in bipolar longitudinal montage only. A full montage set with other reference electrodes cannot be restored from this information. This will have a negative influence on the detection performance.

To mimic the behavior of prospective data, all retrospective patient recordings were used in their full length without restriction. No file selection and no segmentation of patient data were applied. The number of patients with seizures was defined based on the lowest available seizure perception value for each dataset. For the retrospective dataset E25 was used, for the prospective study data $C+E$ was used to define the number of patients with seizures. Table 3 gives an overview of the complete dataset used for this publication.

2.5. Definition of detection performance

EpiScan alarms were compared to manually defined seizure markers. Analysis was done separately for the four evaluation groups $C+E$, E25, E50, and E75. A seizure epoch was defined as a three minutes time range starting from the beginning of the seizure marker. An EpiScan alarm occurs on a specific time point without having time duration. An EpiScan alarm was defined as true positive (TP) when it appeared within a seizure epoch. Several EpiScan alarms in one seizure epoch were defined as one TP. Alarms outside of a seizure epoch were defined as false positives (FP). False positives occurring within a time span of less than 30 s were counted as a single false alarm. A seizure epoch without a matching EpiScan alarm was defined as false negative (FN). Fig. 1 summarizes these definitions.

The sensitivity of the automatic detection was calculated for each patient. It was defined as the ratio between the numbers of true positives (TP) to the number of all seizures (TP + FN). The false alarm rate of the automatic seizure detection was defined by the number of false alarms in 24 h (FA/24 h).

$$\text{Sensitivity} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FN}}$$

$$\text{FA/24h} = \frac{\#\text{FP}}{\text{duration of recording days}}$$

2.6. Comparison to Persyst seizure detection

Currently, Persyst is considered to be the most prevalent seizure detection system. We compared the results from the prospective study data with the results obtained from the Persyst seizure detection in Version 12 (Version 12, Rev. B, 2012.11.27, http://www.persyst.com/). All EEG datasets were converted to EDF format (http://www.edfplus.info/) with a maximum of 99,000 cycles per file. The correctness of the files was validated with the Polyman EDF checker (http://www.edfplus.info/downloads/). Each EDF file was processed separately with the Persyst seizure detection engine. The results of the "SzDetect" table were manually copied to Microsoft Excel tables. These tables were automatically analyzed by reading them into Matlab. The values for sensitivity and false alarm rate were calculated with the same procedures that were used for the EpiScan results.

2.7. Statistical methods

All confidence interval (CI) values are calculated using a parameter-free bootstrapping method for confidence intervals with 1000 bootstrap samples as described in DiCiccio and Efron (1996). The two-sample t-test was used to validate whether two samples came from a distribution with the same mean. The default alpha value was 0.05.

3. Results

3.1. Detection results of the prospective multi-center study

An overview of the dataset collected in the prospective multicenter study is given in Table 2. In total, 15,684 h of EEG including 205 patients with 526 seizures were evaluated. The results for the different perception values are shown in Fig. 2. For those seizures where all reviewers agreed on "probably yes" or higher (E75) EpiScan showed a mean sensitivity of 81% (95% confidence interval = 74–86%) combined with a false alarm rate of 7.1 false alarms per day. As expected including more ambiguous EEG pattern and thus lower perception value for the seizures led to a decrease in sensitivity. With a perception value of E50 ("rather yes" or higher) a mean sensitivity of 78% (95% confidence interval = 70–84%) with a false alarm rate of 7.08 per day was reached.

When calculating the average sensitivity for all 94 patients with seizures regardless of whether the seizure is visible in the EEG or not ($C+E$) we achieved a mean sensitivity of 72% (95% confidence

Table 3
Data used to evaluate EpiScan: Data of the prospective multicenter study (Study), the retrospective data of the development dataset (Devel), the public available pediatric dataset from MIT (CHB–MIT), and the cumulative dataset including all (ALL). The number of patients and the subgroup of the number of patients with seizures (*N* with sz.) are given. The total number of seizures and hours of recorded EEG are shown in the last two columns.

Dataset name	Prosp./Retrosp.	*N*	*N* with sz.	Number of sz.	Hours of EEG
EEG data for EpiScan evaluation					
Study	P	205	94	526	15,684
Devel	R	310	124	1113	25,567
CHB–MIT	R	24	24	197	1355
All	P/R	539	242	1836	42,594

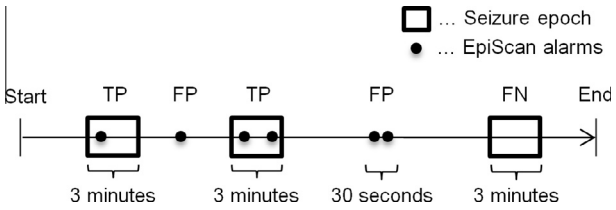

Fig. 1. Definition of true positives (TP), false positives (FP) and false negative (FN) detections by comparing seizure epochs defined by epileptologists to EpiScan alarms.

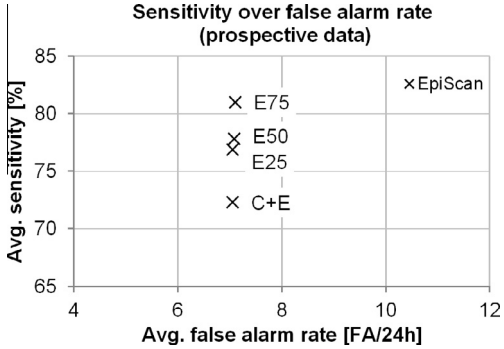

Fig. 2. Results of the prospective multi-center study of EpiScan for different levels of seizure perception values ($C+E$, E25, E50, E75). A steady increase in detection sensitivity can be observed for seizures groups with higher reviewer perception values.

interval = 65–79%). The false alarm rate reduced slightly to 7.05 false alarms per day.

Some seizures were not detected by the clinical protocol of the EMUs at the different centers. Reasons were subtle clinical signs, strong artifact superposition, unobtrusive visual EEG patterns or they were simply overlooked by the reader. During the prospective study, EpiScan detected 16 (3% of all seizures) previously undetected seizures.

We found no statistically significant difference in detection sensitivity between the three participating centers ($p > 0.06$) which shows the robustness of EpiScan against influences of different recording setups.

We further investigated how sensitivity and false alarm rate of individual patients are distributed in the dataset. This gives more insight into the performance of the detection system. We divided the patients in five groups: group one contained patients where EpiScan detected less than 25% of the seizures. Group two contained those patients where 25–50% of the seizures where detected. Group three, four and five contained patients where more than 50%, more than 75% and 100% of the seizures were detected, respectively. A histogram of the $C+E$ detection sensitivities is shown Fig. 3. The results were plotted separately for each center and for the complete dataset. The histograms were normalized to the number of patients of the given center to allow comparison of the results between the different recording sites. Fig. 3 reveals that in more than half of the patients 100% of the seizures were detected. The distribution of the detection sensitivity had a very similar pattern in all centers proving a very stable detection quality for different patient cohorts and recording conditions.

A similar analysis was performed for the false alarm rate. Patients were divided into groups with a false alarm rate per day of less than one, between one and five, five and ten, and between ten and 24 false alarms per day. Fig. 4 shows a histogram for all four centers as well as the combined data. The difference of the false alarm rate from center MUV was statistically significant ($p < 0.05$) compared to the other two centers.

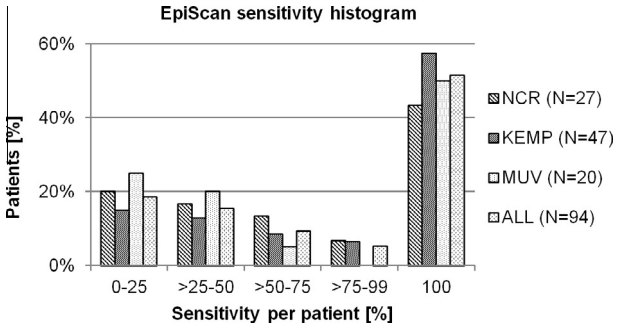

Fig. 3. Histogram of EpiScan detection sensitivities in the prospective study using all seizures ($C+E$). The normalized histograms of the three individual centers (NCR, KEMP, MUV) shows that more than half of the patients were detected with 100% sensitivities.

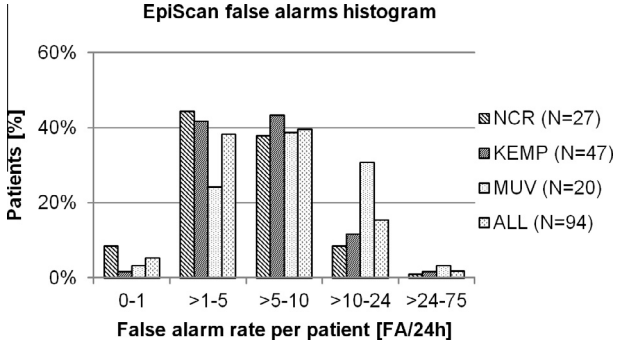

Fig. 4. Histogram of EpiScan false alarms of the three individual centers (NCR, KEMP, and MUV) and combined (ALL) normalized to percent. The mean values are: NCR = 5.1, KEMP = 6.9, MUV = 9.8, ALL = 7.05 FA/24 h.

3.2. Comparison with Persyst seizure detection

In addition to the analysis of the seizure detection system EpiScan we also performed an analysis of the Persyst seizure detector. Performance of EpiScan and Persyst 12 are presented for the data of the prospective study using all patients with seizures ($N = 94$). A comparison of the detection performance for different seizure perception values is given in Fig. 5. The results show an increase of the performance with increasing visibility of the seizure in the EEG for both detectors. For the clearly visible electrographic seizures in the E75 seizure group we found a sensitivity of 81% for EpiScan compared to 75% for the Persyst seizure detection. Similar increases in sensitivity for EpiScan were found for all other groups (+3.7% sensitivity for $C+E$, +6.2% sensitivity for E25, +5.5% sensitivity for E50, $p < 0.76$ for all groups). In addition to the higher sensitivity we found lower false alarm rate for the EpiScan seizure detector (−27% or −2.68 FA/24 h) compared to the Persyst seizure detector for all detection groups. The differences in sensitivity and false alarm rate are not statistically significant using a paired sample t-test.

3.3. Results on retrospective datasets

3.3.1. Results of the development dataset

The results on the development dataset, which contains an extraordinary high number of patients (Table 3), will be shown to further raise the reliability of the EpiScan detection performance. Here, correlation on patient diagnosis will be presented that is not yet available for the prospective study data.

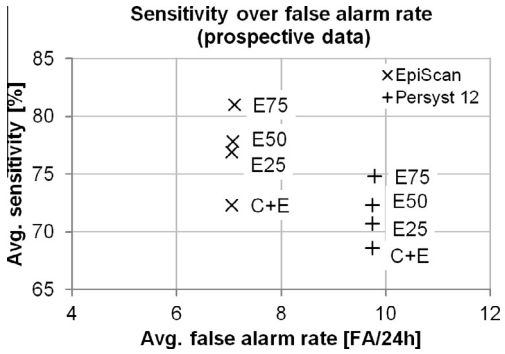

Fig. 5. Comparison of EpiScan and Persyst 12 detection performance using the prospective dataset showing the superior performance of EpiScan compared to Persyst 12. The performance increases with increasing level of perception value which is true for both seizure detectors. An average sensitivity of 81% is reached by EpiScan for unequivocal electrographic seizures (E75), compared to 75% reached by Persyst.

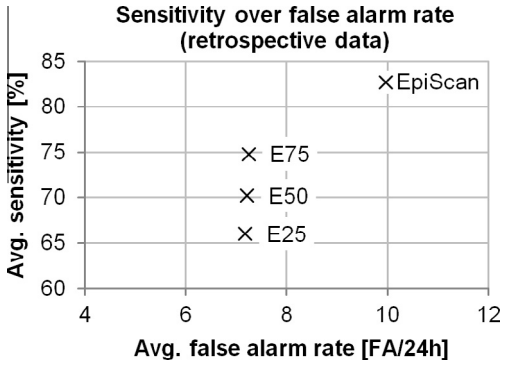

Fig. 6. EpiScan detection performance on the development dataset. The variation of the seizure perception value shows that detection performance increases if subtle electrographic seizures are removed from evaluation.

(Fig. 6) depicts the detection performance of EpiScan for this dataset. Results for the three perception value groups E25, E50 and E75 are shown. As for the prospective dataset, an increased seizure perception value (defined in Table 1) results in better detection performance. For perception value E75 the detection performance is 75% sensitivity with a false alarm rate of 7.2 FA/24 h.

We looked a possible correlation between the type of epilepsy and the detection performance of an automatic system. Table 4 compares the results for patients suffering from mesial temporal lobe epilepsy (mTLE), temporal lobe epilepsy (TLE), extra temporal lobe epilepsy (XTLE) and frontal lobe epilepsy (FLE). The best results of an average sensitivity of 87% were found for the subgroup with mTLE because of many regular rhythmic patterns during seizures. The TLE subgroup achieved an average sensitivity of 83% which is a very good result. The missed seizures are due to a few patients with neocortical seizure onset zone (SOZ) which often exhibited unique seizure patterns for each patient that were not always detected. A similar explanation for the lower detection sensitivity applies to the XTLE and FLE group, which also include some patients with neocortical SOZ. An additional problem of the FLE group was that a typical seizure shows high amplitude muscle artifacts but some seizures from certain patients lack this property and show only average amplitude artifact and no other obvious seizure activity in the EEG.

3.3.2. Results on pediatric EEG data

In addition to the large dataset recorded from adults a small pediatric dataset was analyzed. The detection performance of EpiScan on the MIT dataset had an average sensitivity of 67% (95% confidence interval from 53% to 79%) with 7.7 FA/24 h on average. No conclusions about the expected performance of EpiScan on pediatric data can be drawn as the amount of patients with seizures is too small.

Table 4
Average detection performance of EpiScan for patient groups with different diagnoses. The development dataset with unequivocal electrographic seizures (E75) were used. Mesial temporal lobe epilepsy (mTLE) showed the best results.

Diagnosis	Sensitivity [%]	False alarm rate [FA/24 h]
Average detection performance of EpiScan compared between patient diagnoses		
mTLE ($N = 11$)	87	6.6
TLE ($N = 52$)	83	6.7
XTLE ($N = 50$)	64	7.3
FLE ($N = 11$)	54	7.2
No epilepsy	–	7.2

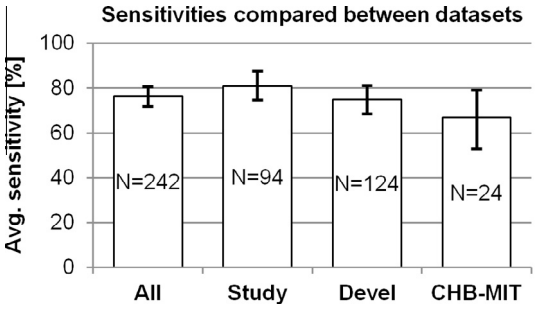

Fig. 7. Average detection sensitivities of EpiScan compared between different patients groups with different sample sizes. All sensitivities use the E75 seizure definition. At $N = 242$ the 95% confidence interval reduces to 8%, whereas a sample size of $N = 24$ will result in a range from 53% to 79%.

3.4. Cumulative meta-analysis of all available data

The prospective and all available retrospective data was used to assess statistical variables with high confidence level. The recording length, number of patients and number of seizures of all patients are listed in Table 3. Fig. 7 shows the nearly invariant detection performance of the different datasets. When combining all data into a large dataset the confidence interval reduces to 8.6% for all 242 patients with seizures. The difference in sensitivity between retrospective dataset (Devel) and prospective dataset (study) is small and statistically insignificant ($p = 0.91$). The average false alarm rate of all 539 patients was 7.2 false alarms in 24 h with a 95% confidence interval of 6.7–7.8 FA/24 h. No statistical significant difference between the different datasets could be found ($p > 0.38$ for all combinations).

4. Discussion

Seizure detection is an eagerly awaited feature in clinical practice of EMUs. We presented a multi-center study for the EpiScan online seizure detection system. EpiScan was tested as an online device in three different EMUs and its sensitivity and false alarm rate was calculated.

We were able to show that a sensitivity of 81% can be reached for seizures that are clearly visible in the EEG. We compared these results to the Persyst software, the most widely used seizure detection system. For the Persyst system we found a performance of about 75% showing that the new detection system EpiScan performs at least as good as the Persyst system.

The average sensitivity of EpiScan of about 81% was achieved for adult patients of age 18 and above. A small pediatric dataset

showed that EpiScan achieved comparable results also for EEG of children. No significant differences compared to the results of the development or prospective dataset were found. We concluded that an application of EpiScan on pediatric patients is also possible. However, the size of the pediatric dataset of $N = 24$ was too small to achieve a statistically valid comparison. More data are necessary to draw a reliable conclusion about the EpiScan performance for pediatric datasets.

A low false alarm rate is an essential feature of online seizure detection system. An alarm rate of several alarms per hour would render a seizure detection system useless regardless of the achieved sensitivity. We found an average false alarm rate of the EpiScan system of 7.1 false alarms per day. The comparison with the Persyst software showed that the Persyst system has a false alarm rate which is about 27% higher than that of EpiScan. As pointed out in the Methods section, false alarms within 30 s were counted as single false positive. An alarm occurs on a specific time point without having time duration. The concatenation of several false alarms within 30 s to a single false positive therefore corresponds to a false positive with maximal length of 30 s. The rational for this definition of a maximum length for a false alarm was that a reviewer should be able to determine if an EpiScan marker is a false alarm by looking at one single page of EEG. By restricting the maximum length of an artifact to 30 s we ensured that it would not be necessary to scroll through the EEG when classifying a marker as a false alarm. Variation of this maximal length parameter had only little impact on the false alarm rate. Increasing the time range for false alarms to 3 min will reduce the false alarm rate by 15%. We found a small increase in false alarm rate of less than 1% when the minimal perception value of the seizures was increased. This effect relates to the fact that EpiScan detected numerous seizures with a low perception value. When increasing the minimal perception value these seizure epochs do not longer count as true positive but add to the false alarms. Thus the marker converts to a false alarm according to the definitions in Fig. 1.

We found non-significant differences in sensitivity between the different centers that participated in the study. Only the center MUV showed a significantly higher false alarm rate of more than 10 false alarms per day due to artifacts appearing as epileptiform discharges during many of the recordings. The artifact contamination in the data from center KEMP was also high but did not raise the false alarm rate as in center MUV. A detailed analysis showed that most of the KEMP artifacts consisted of movement related electrode artifacts which did not trigger one of the seizure detection methods in EpiScan. Although some other artifacts like tooth brushing and movements during cycling did raise false alarms the incidence rate was too low to have a significate effect.

It is a question of debate if only a prospective clinical study of a seizure detection system can give a reliable proof of the performance. The results of our clinical study showed no statistically significant differences compared to the performance that was determined in offline experiments on a large development dataset. We did find a slightly increased performance in the study dataset that was due to the lower number of patients in the clinical study as compared to the offline-dataset. While usability issues can only be addressed in a clinical setting we argue that the use of large offline-datasets will give information about the performance of a seizure detection system that is as good as a prospective clinical study. The amount of data in an extensive offline analysis can be much larger than in any reasonable clinical trial of an automatic seizure detection system. It is of course necessary to ensure that the dataset used for offline-evaluation reflects the data found in an EMU. Taking for instance data from TLE-patients only, from seizure patients only, taking patients that only show a low level of artifacts or taking subsets of the datasets based on different inclusion criteria might strongly bias the results of an offline analysis. In addition it is important that large periods without seizures are included in every dataset in order to get a reliable estimation of both sensitivity and false alarm rate. Studies on many patients but very few hours of EEG only show the sensitivity but not the specificity of the method. In addition care must be taken not to over-fit the detection system to a given dataset by using the same small dataset for development and subsequent testing. Ideally the dataset for development and testing will be different. In practice a sufficient large dataset of several 10,000 h of data will also reduce this problem of over-fitting. In addition it is important to include complete 24 h recordings of the EEG in order to analyze the complete day/night cycle of a patient during the validation of a detection system.

We believe that patient safety in EMUs can be increased even by an imperfect seizure detection system. It is often assumed that medical staff will reach near 100% surveillance. This however is not always the case. A study by Atkinson et al. (2012) claimed that only 40% of the seizure showed a staff response. Although this number seems low, the general statement that a certain number of seizures in EMUs do not get immediate staff response is in line with our findings. Even at the large centers that participated in this study the automatic seizure detection system did find additional seizures that were not detected by the manual review procedures. The amount of missed seizures during recording depends heavily on the available staff in EMUs, their training and the time when a seizure occurs. Centers that do have highly trained staff available in their EMUs 24 h a day and that show a staff-patient-ratio of up to 1 technician for three patients during the day will be less likely to miss seizures. However, smaller hospitals often cannot afford this number of staff at their EMUs leading to an increased number of missed or un-responded seizures, especially during the night when even less staff is available. Here, an automated seizure detection system would be of great value by providing additional safety to the patient, since it will not depend on human factors, staff availability or time of day but on the quality of the visible EEG pattern only.

It is evident from all performed studies so far that an automatic seizure detection system will never replace the human EEG analysis. It will always be an additional source of information. Especially on highly specialized EMUs for pre-surgical evaluation such a system would be used in addition to today's procedures in order to find possible additional seizures that were missed by the visual analysis. However, even the larger centers are under constant pressure to cut costs. The available staff for EEG analysis generally decreases. Here, an automatic seizure detection system provides reliable monitoring of the EEG that will help to ensure the level of patient safety.

5. Conclusion

An automatic seizure detection and alerting system was validated in a prospective multi-center study and on retrospective data. In total 42,000 h of uncut long term EEG recordings were used to assess detection performance by means of sensitivity and false alarm rate on a high statistical confidence level. The results showed 81% sensitivity for seizures with high perception value at 7.1 false alarms per day in the prospective study. The analysis of 539 patient recordings showed no significant difference between prospective and retrospective detection results.

The multi-center study of EpiScan proved the ability of the system to work as seizure alarm device in the clinical long-term video EEG monitoring. The evaluation of the seizure detection performance on patient data with various diagnoses and ages showed the universal applicability of EpiScan. The comparison to the currently most prevalent seizure detection system from Persyst

showed that EpiScan reaches a lower false alarm rate on the collected prospective dataset but the difference was not of statistical significance. We conclude that the application of the EpiScan seizure detection system in EMUs could increase the efficiency and the safety level for patients.

Disclosure

None of the authors has any conflict of interest to disclose. We confirm that we have read the journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Acknowledgements

Many thanks for the support of our work to all people involved and especially to Roy Krijn and to all staff members of the IT departments at AKH Vienna and the General Hospital Hietzing. Special thanks for the comprehensive and reliable EEG analysis to the technicians involved in the study, namely Harrie Geeris, Michaela Demel, and Julia Tarra.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.clinph.2014.09.023.

References

Atkinson M, Hari K, Schaefer K, Shah A. Improving safety outcomes in the epilepsy monitoring unit. Seizure J Br Epilepsy Assoc 2012;21:124–7.

Beniczky S, Polster T, Kjaer TW, Hjalgrim H. Detection of generalized tonic–clonic seizures by a wireless wrist accelerometer: a prospective, multicenter study. Epilepsia 2013;54:e58–61.

Carlson C. First do no harm: safety in the epilepsy monitoring unit. Epilepsy Curr 2009;9:162–3.

DiCiccio TJ, Efron B. Bootstrap confidence intervals. Stat Sci 1996;11:189–228.

Dollfuss P, Hartmann MM, Skupch A, Fürbass F, Kluge T. Automatic optimization of parameters for seizure detection systems. In: 35th Annu Int Conf IEEE Eng Med Biol Soc EMBC. 2013. p. 1976–9.

Fisher RS, van Emde Boas W, Blume W, Elger C, Genton P, Lee P, et al. Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Epilepsia 2005;46:470–2.

Fürbass F, Hartmann M, Perko H, Skupch A, Dollfuss P, Gritsch G, et al. Combining time series and frequency domain analysis for a automatic seizure detection. In: 2012 Annu Int Conf IEEE Eng Med Biol Soc EMBC. 2012. p. 1020–3.

Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. Circulation 2000;101:e215–20.

Gotman J. Automatic recognition of epileptic seizures in the EEG. Electroencephalogr Clin Neurophysiol 1982;54:530–40.

Gotman J. Automatic seizure detection: improvements and evaluation. Electroencephalogr Clin Neurophysiol 1990;76:317–24.

Gotman J, Gloor P. Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. Electroencephalogr Clin Neurophysiol 1976;41:513–29.

Hartmann MM, Fürbass F, Perko H, Skupch A, Lackmayer K, Baumgartner C, et al. EpiScan: online seizure detection for epilepsy monitoring units. Eng Med Biol Soc 2011 Annu Int Conf IEEE. 2011. p. 6096–9.

Hunyadi B, Signoretto M, Van Paesschen W, Suykens JAK, Van Huffel S, De Vos M. Incorporating structural information from the multichannel EEG improves patient-specific seizure detection. Clin Neurophysiol 2012;123:2352–61.

Kluge T, Hartmann M, Baumgartner C, Perko H. Automatic detection of epileptic seizures in scalp EEG-recordings based on subspace projections. Epilepsia [Internet] 2009 [cited 2011 Mar 15]; 26–27. Available from: <http://onlinelibrary.wiley.com/doi/10.1111/j.1528-1167.2009.02377_1.x/abstract>.

Lee JW, Shah A. Safety in the EMU: reaching consensus. Epilepsy Curr 2013;13:107–9.

Perko H, Baumgartner C, Kluge T. Online algorithm for epileptic seizure detection. Lecture: annual meeting of swiss society of biomedical engineering. Neuchâtel, Schweiz; 2007.

Schachinger D, Baumgartner Ch, Kluge T. Bivariate phase space divergence: a measure for the automatic detection and anticipation of epileptic seizures in ECoG data. Lecture: international special topic conference on information technology in biomedicine. Ioannina, Greece; 2006.

Shoeb AH. Application of machine learning to epileptic seizure onset detection and treatment [Internet, thesis]. Massachusetts Institute of Technology; 2009 [cited 2013, July 9]. Available from: <http://dspace.mit.edu/handle/1721.1/54669>.

Skupch AM, Dollfuss P, Fürbass F, Gritsch G, Hartmann MM, Perko H, et al. Spatial correlation based artifact detection for automatic seizure detection in EEG. In: 2013 35th Annu Int Conf IEEE Eng Med Biol Soc EMBC. 2013. p. 1972–5.

Smith S. EEG in the diagnosis, classification, and management of patients with epilepsy. J Neurol Neurosurg Psychiatry 2005;76(Suppl. 2):ii2–7.

Spanaki MV, McCloskey C, Remedio V, Budzyn D, Guanio J, Monroe T, et al. Developing a culture of safety in the epilepsy monitoring unit: a retrospective study of safety outcomes. Epilepsy Behav 2012;25:185–8.

Todorov AB, Lesser RP, Uematsu SS, Yankov YA, Todorov AA. Distribution in time of seizures during presurgical EEG monitoring. Neurology 1994;44:1060–4.