Clinical Neurophysiology xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

Validation of a new automated neonatal seizure detection system: A clinician's perspective

P.J. Cherian [a,*], W. Deburchgraeve [b], R.M. Swarte [c], M. De Vos [b,d], P. Govaert [c], S. Van Huffel [b], G.H. Visser [a]

[a] Section of Clinical Neurophysiology, Department of Neurology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[b] Department of Electrical Engineering (ESAT), Katholieke Universiteit, Leuven, Belgium
[c] Section of Neonatology, Department of Pediatrics, Erasmus MC-Sophia, University Medical Center, Rotterdam, The Netherlands
[d] Neuropsychology Lab, Department of Psychology, University of Oldenburg, Oldenburg, Germany

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Neonatal seizures
Automated seizure detection
Perinatal asphyxia
HIE
NICU

ABSTRACT

Objective: To validate an improved automated electroencephalography (EEG)-based neonatal seizure detection algorithm (NeoGuard) in an independent data set.

Methods: EEG background was classified into eight grades based on the evolution of discontinuity and presence of sleep–wake cycles. Patients were further sub-classified into two groups; gpI: mild to moderate (grades 1–5) and gpII: severe (grades 6–8) EEG background abnormalities. Seizures were categorised as definite and dubious. Seizure characteristics were compared between gpI and gpII. The algorithm was tested on 756 h of EEG data from 24 consecutive neonates (median 25 h per patient) with encephalopathy and recorded seizures during continuous monitoring (cEEG). No selection was made regarding the quality of EEG or presence of artefacts.

Results: Seizure amplitudes significantly decreased with worsening EEG background. Seizures were detected with a total sensitivity of 61.9% (1285/2077). The detected seizure burden was 66,244/97,574 s (67.9%). Sensitivity per patient was 65.9%, with a mean positive predictive value (PPV) of 73.7%. After excluding four patients with severely abnormal EEG background, and predominantly having dubious seizures, the algorithm showed a median sensitivity per patient of 86.9%, PPV of 89.5% and false positive rate of 0.28 h^{-1}. Sensitivity tended to be better for patients in gpI.

Conclusions: The algorithm detects neonatal seizures well, has a good PPV and is suited for cEEG monitoring. Changes in electrographic characteristics such as amplitude, duration and rhythmicity in relation to deteriorating EEG background tend to worsen the performance of automated seizure detection.

Significance: cEEG monitoring is important for detecting seizures in the neonatal intensive care unit (NICU). Our automated algorithm reliably detects neonatal seizures that are likely to be clinically most relevant, as reflected by the associated EEG background abnormality.

© 2011 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Neonatal seizures occur commonly in the neonatal intensive care unit (NICU) and usually suggest serious neurological dysfunction (Lombroso, 1996). The most common cause of neonatal seizures in the NICU is hypoxic ischaemic encephalopathy (HIE). The majority of seizures occurring in encephalopathic neonates are subclinical, being detected only by continuous electroencephalography monitoring (cEEG) (Murray et al., 2008). However, detection and treatment of neonatal seizures are hampered by limited access to cEEG. It is an expensive and labour-intensive technique and its interpretation requires specialised training. Even in the few NICUs where it is available, expert evaluation of the EEG by a neurologist/clinical neurophysiologist is not available around the clock, thus limiting its usefulness. In the past years, many NICUs have adopted amplitude integrated EEG (aEEG), which displays time-compressed and asymmetric band-pass-filtered data from a single EEG channel. However, the limited spatial and temporal sampling results in loss of EEG information and also the results of visual interpretation are variable (Bourez-Swart et al., 2009; Rennie et al., 2004; Shellhaas et al., 2007). Improvement of cEEG can be achieved by the development of a reliable automated seizure detection method with a fair degree of sensitivity and low false positive detection rates (Aarabi et al., 2006; Faul et al., 2005; Gotman et al., 1997; Greene et al., 2007; Navakatikyan

* Corresponding author. Address: Section of Clinical Neurophysiology, Ba 451a, Department of Neurology, Erasmus MC, University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands. Tel.: +31 10 70 35 662; fax: +31 10 70 34 621.
E-mail address: j.perumpillichira@erasmusmc.nl (P.J. Cherian).

1388-2457/$36.00 © 2011 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.clinph.2011.01.043

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

et al., 2006). Presently, there is no commercially available multichannel EEG-based neonatal seizure detection algorithm that is widely accepted for clinical use.

Another reason that cEEG coupled with automated neonatal seizure detection is not being widely used may be the lack of convincing evidence that treatment of subclinical (electrographic) neonatal seizures with antiepileptic drugs (AED) improves clinical outcome. In conditions like HIE, it is difficult to separate the effects of seizures from the effects of primary hypoxic brain injury. It is known that electrographic neonatal seizures occur mostly in association with an abnormal EEG background (Laroia et al., 1998; McBride et al., 2000), which in turn reflects the severity of the underlying brain dysfunction (Sarnat and Sarnat, 1976). EEG background is also a strong predictor of outcome (Bye et al., 1997; Holmes et al., 1982; Watanabe et al., 1980b), especially after perinatal asphyxia. To be clinically relevant, we feel that the detection and interpretation of neonatal seizures should be done in the context of EEG background activity. Recently, we have developed an automated neonatal seizure detection system that mimics a human interpreter reading the EEG, with promising results (Deburchgraeve et al., 2008). We have further improved it, especially in rejecting artefacts (Deburchgraeve, 2010). The aim of the present study was to test this improved algorithm, called NeoGuard, in a new and much larger EEG data set. The secondary aim was to assess the performance of the algorithm in detecting electrographic seizures occurring upon varying degrees of EEG background abnormality, which has not been done previously.

2. Patients and methods

EEG data from 119 consecutive newborns with presumed perinatal asphyxia, who underwent video-EEG monitoring for $\geqslant$24 h, between March 2003 and August 2007 as part of an ongoing study formed the database. The inclusion criteria for cEEG were clinical features of encephalopathy, and having at least one of the following features of birth asphyxia: (a) arterial pH of umbilical cord blood $\leqslant$7.1, (b) Apgar score $\leqslant$5 at 5 min and (c) high clinical suspicion (like foetal distress, umbilical cord prolapse, difficult labour or a history of convulsions). Babies with congenital cardiac abnormalities and multiple congenital anomalies were excluded; 45/119 neonates had recorded seizures. Twenty-one of them, described in our previous article describing the seizure detection algorithm were excluded, leading to a large new set of 'unseen' monitoring data from 24 newborns. Three of these patients (Nos. 1, 4 and 5) were subsequently found to have arterial ischaemic stroke on magnetic resonance imaging (MRI). All the patients except two (Nos. 17 and 19, with gestational ages of 30 and 35 weeks) were born at term. MRI of the brain was available in 21, done according to an asphyxia protocol and scored by a neuroradiologist (Swarte et al., 2009). The study had the approval of the Erasmus MC Medical Ethical Review Board.

2.1. EEG registrations

Digital video-EEG with polygraphy, started mostly $\leqslant$24 h post partum (PP), was registered continuously for 1–3 days using a Nervus™ monitor (Taugagreining hf, Reykjavik, Iceland). Seventeen scalp electrodes were applied according to the full 10–20 International System (Cherian et al., 2009) in all except two patients. A restricted 10–20 system with 13 electrodes (Fp1-2, C3-4, Cz, F7-8, T3-4, T5-6, O1-2) was used in one, whilst in the other, nine electrodes (Fp1-2, C3-4, Cz, T3-4, O1-2) were used. Polygraphy included ECG, respiration, electro-oculogram (EOG), chin EMG and limb movements. EEG sampling frequency was 256 Hz. Band-pass filter was 0.3–70 Hz.

EEGs were interpreted by a clinical neurophysiologist as part of patient care and this information was used to make treatment decisions. All patients in this study received loading doses of phenobarbitone (20 mg kg^{-1}). In patients with persistent electrographic seizures, intravenous midazolam was given as a loading dose of 0.1 mg kg^{-1}, followed by maintenance doses of 0.1–0.5 mg kg^{-1} h^{-1}. Babies with seizures refractory to midazolam received intravenous lidocaine, 2 mg kg^{-1}, followed by a maintenance dose of 6 mg kg^{-1} h^{-1}, which was tapered off over 30 h. The AED treatment was according to a departmental protocol with every next step taken when no effect on electrographic seizures was seen within 1–2 h, and is comparable to the ones widely used in European NICUs (Vento et al., 2010).

2.2. Scoring EEG background

We used an eight grade system (Table 1), derived from the literature (Holmes et al., 1982; Murray et al., 2009; Pezzani et al., 1986; Watanabe et al., 1980a) and modified it based on our clinical experience. This scoring system emphasises discontinuity (voltage and length of the discontinuous periods) and its evolution over 24 h, presence of variability, reactivity and SWC. Severity of discontinuity is easy to assess and is a robust predictor of outcome (Biagioni et al., 1999). Discontinuity was defined as voltage attenuation (<25 μV for mild and <10 μV for severe) in at least 50% of the EEG channels, lasting for $\geqslant$5 s. To score discontinuity, the most frequently occurring measure of the amplitude and duration of the discontinuous periods were used. EEGs were blindly scored by a clinical neurophysiologist (PJC) together with a neonatologist (RMS). Consensus was reached in all cases. We omitted parameters such as frequency content and duration of the EEG bursts, frequency of sporadic or rhythmic sharp waves within and outside the bursts, etc. because we wanted the classification to be readily applicable at the bedside. We also avoided the term burst-suppression, as this has been defined variably in neonates (Biagioni et al., 1999).

2.3. Electrographic seizures

We defined as 'definite seizures', ictal-appearing electrographic discharges that showed a clear variation from background activity,

Table 1
EEG background scoring based on evolution of discontinuity and SWC over 24 h.

Grade	DurIBI	AmpIBI	SWC	Improve	Reac	Variab	Other findings
0	NA	$\geqslant$25–50	++	NA	+	+	Normal for CA
1	$\leqslant$5	10–25	±	↑SWC	+	+	↑ShW, <6 s runs of theta/delta
2	<10	<10	–	↓IBI	+	+	*
3	10–20	<10	–	To grade 2	+	+	–
4	>20	<10	-	To grade 2	±	±	–
5	10–20	<10	–	No	±	±	–
6	>20	<10	–	To grade 5	–	–	
7	10–20	<10	–	Worsen	–	–	↑IBI > 20
8†	>20	<10	–	No/ worsen	–	–	–

SWC, sleep-wake cycles; h, hours; CA, conceptional age; NA, not applicable; IBI, inter-burst interval; DurIBI, length of IBI in seconds; AmpIBI, mean amplitude of IBI in μV; Improve, improvement in 24 h; React, reactivity; Variab, variability; ShW, sharp waves.
*Excessive inter-hemispheric asynchrony and amplitude asymmetry were also classified as at least grade 2. †Includes also persistent low voltage (<10 μV) and isoelectric EEG background. ↑Increased. ↓Decreased.

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

displaying a repetitive pattern of sinusoidal oscillations or sharp waves, or a mixture of both (Fig. 1A and B), lasting ⩾10 s, with evolution in amplitude and frequency over time (Bye and Flanagan, 1995; Deburchgraeve et al., 2008; Lombroso, 1993), whether they had clinical correlates or not. We also classified discharges as 'dubious seizures' (a) runs of sharp waves/oscillations or a mixture of both occurring arrhythmically (with marked variability in the interval and morphology between individual complexes in a seizure discharge for the major part of its duration) or (b) rhythmic discharges of shorter (<10 s) duration, or periodically occurring complexes consisting predominantly of sharp waves. It was difficult to identify the onset and offset of such discharges (Figs. 2B and 3). We chose to group them under 'seizures' as they were seen to recur paroxysmally during the monitoring. Inter-rater agreement amongst two experienced clinical neurophysiologists (PJC and GHV) for scoring of seizure occurrence in 1-h segments of EEGs from 10 patients from another data set was 73% (Deburchgraeve et al., 2008), with a Kappa value (κ) of 0.4. Subsequently, these data were reviewed jointly by both the neurophysiologists and the scored seizures were agreed upon by consensus. For this study, the EEGs were reviewed in their entirety by a neurologist/clinical neurophysiologist (PJC), blinded to the clinical and MRI results, and the onset, duration, location, spread, frequency and rhythmicity of the electrographic seizures were documented.

2.4. Seizure detection algorithm

The technical details of this algorithm have been discussed in detail in a previous article (Deburchgraeve et al., 2008). Briefly, this consists of two detectors running in parallel: (a) a spike-train detector that detects high-energy segments of the EEG and analyses the correlation between them and (b) oscillatory seizure detector that detects increase in low-frequency activity (1–8 Hz) with high autocorrelation. This algorithm has been further improved, especially in artefact reduction. We identified the three most important types of artefacts hindering seizure detection, namely ECG spike artefacts, blood vessel pulsation and respiration artefacts. An algorithm has been developed (Deburchgraeve, 2010) to remove these three artefacts from the EEG using the input from the simultaneously recorded polygraphy channels (ECG and respiration movement). The algorithm searches for activity in the EEG channels that have a high correlation with activity in the polygraphy channels. Only after automatically removing such activity from the EEG channels is the seizure detection algorithm applied. The algorithm runs on the Matlab™ (The MathWorks, Natick, MA, USA) platform. It is capable of running online during bedside registration of EEG. For this study, all the cEEG data were converted to European Data Format and the algorithm was run offline, independently by two of the authors (P.J.C. and W.D.). No selection of EEG data based on the quality of registrations or presence of artefacts was made. The detections were reviewed independently and were compared to the visual scoring done by the clinical neurophysiologist. For each patient, a list of true positives, missed detections (false negatives) and false positives were made. As the algorithm provides an output of event-based detection (not epoch-based, as reported by studies evaluating shorter segments of EEG), we do not have a value for true negatives. Missed seizures as well as false positives were jointly reviewed. Consensus was reached in all cases. We calculated the sensitivity per patient (SensPP), the total sensitivity (SensT_PP) and positive predictive value (PPV) as described previously (Deburchgraeve et al., 2008) as well as the detected seizure burden (cumulative duration) in seconds (Vanhatalo, 2011). We also looked at the number of false positive detections per hour (Fp h^{-1}) as well as their cumulative duration (s).

2.5. Statistical analysis

Since there were not enough number of patients for a three-way classification based on EEG background (mild, moderate and severe abnormalities), dichotomous classification of patients was done (gpI: mild to moderately severe (grades 1–5) and gpII: severely abnormal (grades 6–8)) for comparing the electrographic characteristics of seizures as well as the performance of the seizure detection algorithm. Comparison of the two groups was done using Mann–Whitney U-test or the Fisher's exact test. A p value of <0.05 was considered significant.

3. Results

A total of 756 h of EEG data were studied. Median duration of EEG recording was 25 h (range 17–78). The offline running of the algorithm on EEG data of 24 h (about 1 GB of data) from a patient on a laptop computer with a 2.26 GHz Core™ 2 Duo processor (Intel Corporation, Santa Clara, CA, USA), was done in about 12 h. The algorithm gives an output of the number of events detected, the position in time where the event was detected and the duration of the detected events. NeoGuard also allows visual review of the detections (Fig. 1B). A total of 2077 seizures were scored visually (median 67 per patient, range 7–236). The total seizure burden was 97,574 s.

Ten patients had mild to moderate abnormalities of EEG background (gpI, grades 1–5), whilst 14 had severe abnormalities (gpII, grades 6–8). MRI scans were not done in three neonates in gpII. In the 21 newborns who had an MRI, lesions of basal ganglia and thalamus were significantly increased in patients in EEG gpII, when compared to gpI (9/11 vs. 2/10, $p = 0.009$, Fisher's exact test). Follow-up data for at least 2 years were available in all patients except one in gpI, who had a normal MRI. Severely abnormal outcomes (death or severe handicap) were significantly increased in patients in gpII, with 10 patients dying and four developing severe handicap (gpII vs. gpI, 14/14 vs. 2/9, $p < 0.001$, Fisher's exact test).

3.1. Electrographic characteristics that may influence automated seizure detection

For the whole group, the median values (range) of various seizure parameters were frequency 2.3 Hz (1–8), amplitude 50 μV (10–150), duration 36 s (18–174) and percentage of arrhythmic seizures 11% (0–100). EEG background abnormality (grades 1–8) as well as the seizure characteristics in each patient is shown in Tables 2A and B. There was a significant negative correlation between the seizure amplitude and the EEG background grade (Spearman's rho = −0.475, $p = 0.019$). There were trends for increasing expression of arrhythmic seizures, increase in seizure burden, seizure firing frequency and decrease in individual seizure duration with increasing severity of EEG background abnormality, though the results did not reach statistical significance (results not shown). Increasing expression of arrhythmic seizures (expressed as a percentage of total number of seizures) tended to be associated with lower seizure amplitude (Fig. 4). In patients in EEG gpII, the total number of seizures expressed was significantly increased whilst the amplitude of the seizures was significantly decreased (Table 3, Fig. 5). There was a tendency for patients in this group to express more seizures with a higher firing frequency.

3.2. Dubious seizures

Four patients in gpII had predominantly (>90%) 'dubious' seizures, with low amplitude arrhythmic discharges expressing oscillatory morphology or mixed patterns (Fig. 3). One patient in

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

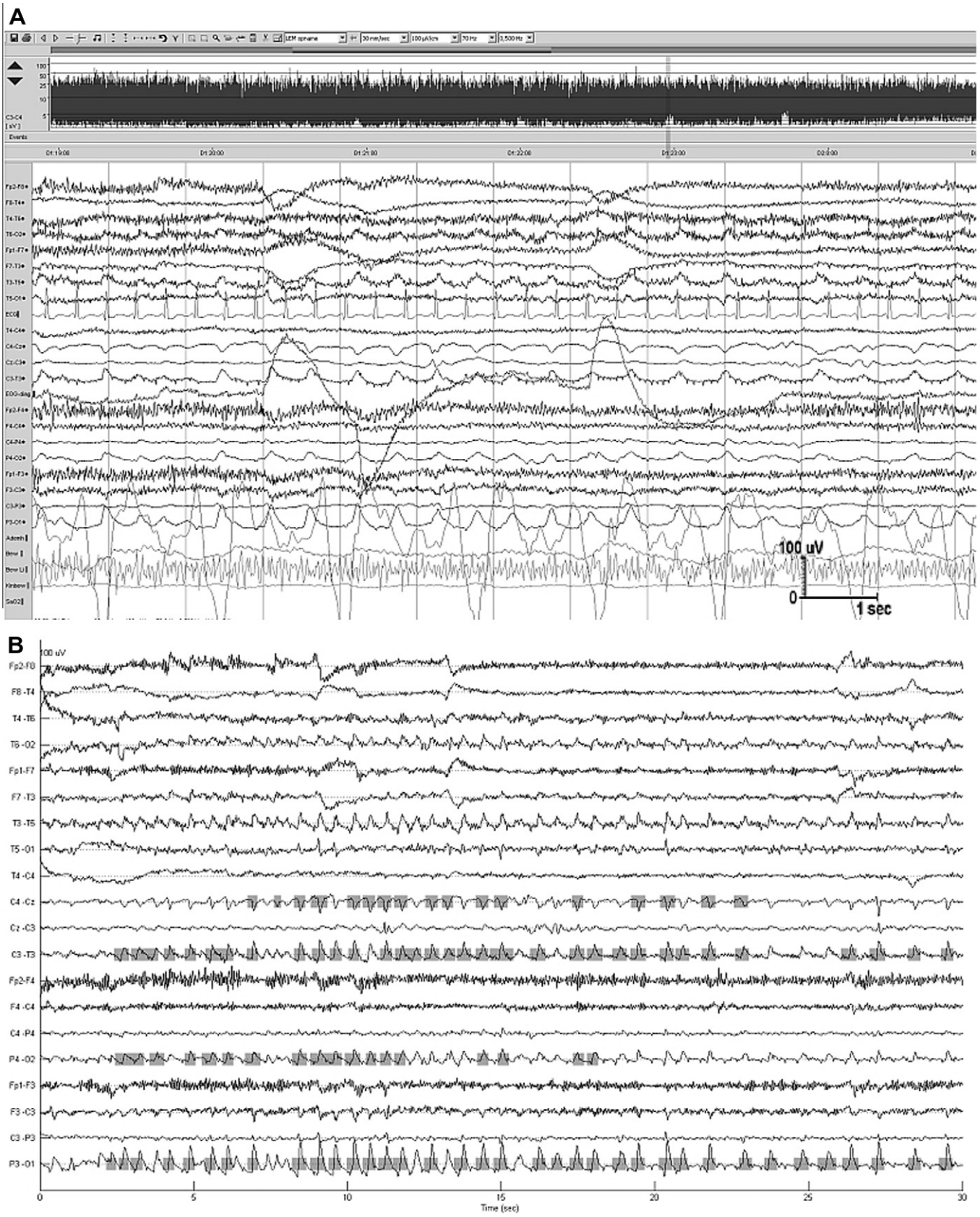


Fig. 1. (A) A left central-temporal-parietal seizure with spread to the right side, expressing a mixed (sharp waves and oscillations) pattern in patient No. 8. It is difficult to identify the seizure (grey vertical bar) in the aEEG trend shown at the top, probably due to contamination of EEG by muscle artefacts. (B) Detection of the seizure (grey blocks) by the algorithm after automated removal of artefacts (time base 30 s/page, the first 12 s correspond to the EEG segment shown in (A)).

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

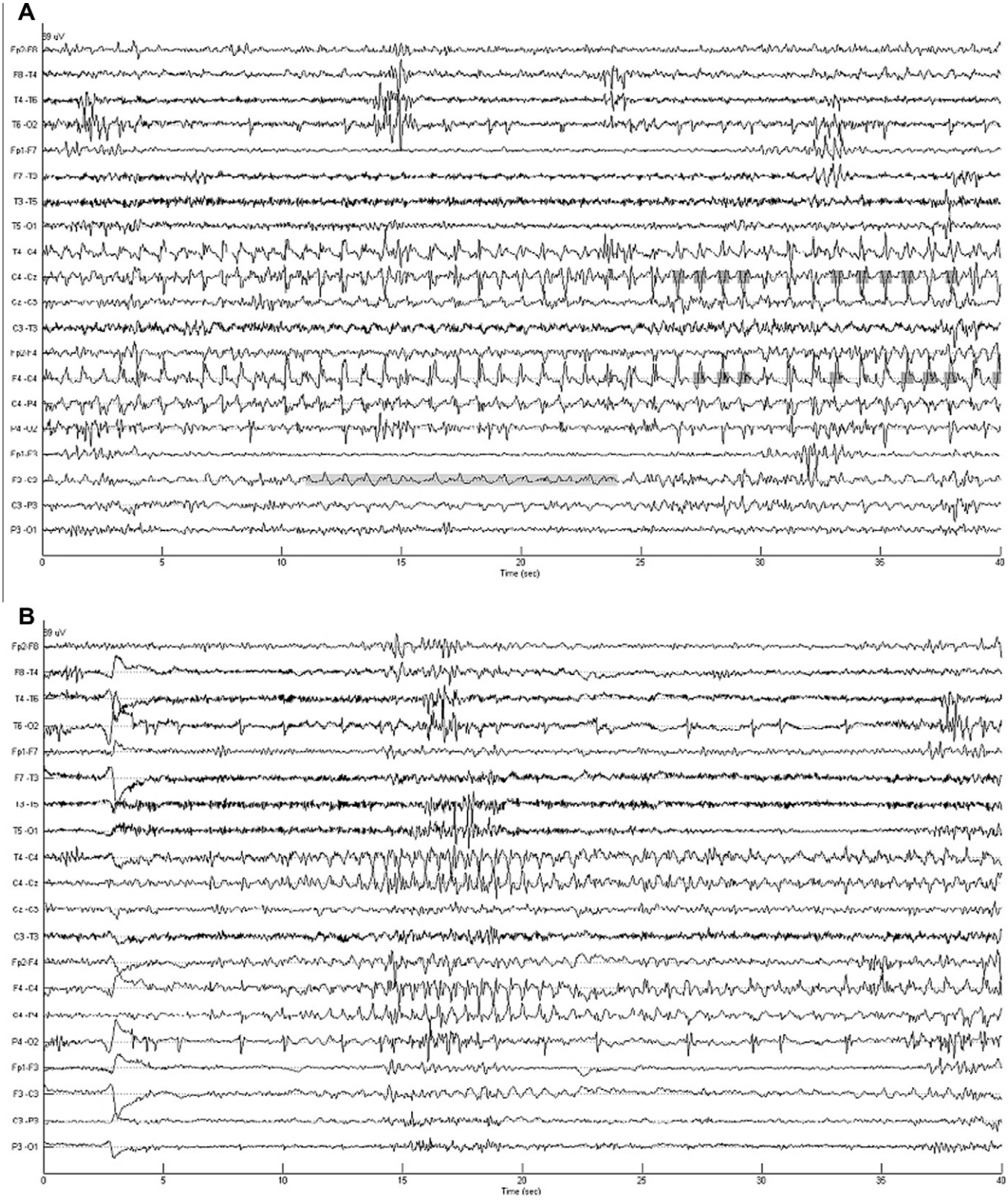


Fig. 2. (A) Seizure in patient No. 5 with right middle cerebral artery territory stroke, showing rhythmic sharp waves and mixed patterns over the right central-parietal regions with spread to the left side, detected by the algorithm. (B) Dubious seizures in the same patient, characterized by brief rhythmic discharges and periodic sharp waves, occurring over the same regions. These shorter discharges were variably detected by the algorithm.

gpI, with a middle cerebral artery territory stroke (patient No. 5) had in addition to definite seizures, 'dubious' seizures consisting of short runs of rhythmic sharp waves (Fig. 2B) (Oliveira et al., 2000), as well as periodic sharp waves (lateralised periodic discharges or LPDs, also known as periodic lateralised epileptiform discharges or PLEDs), involving the same regions.

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

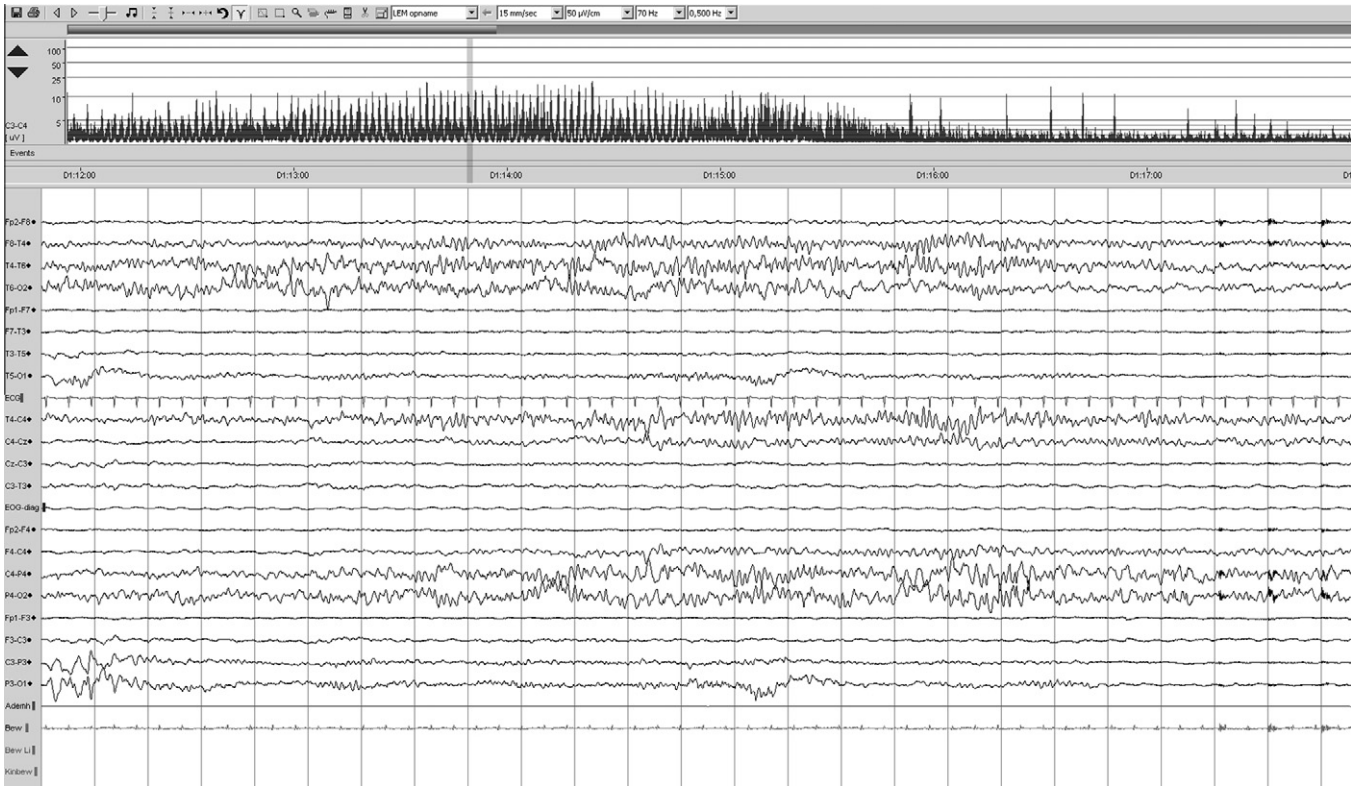


Fig. 3. Dubious seizure in patient No. 12. Arrhythmic low amplitude discharges occurring over the right temporal-parietal regions (a paroxysmal pattern is well seen in the aEEG trend at the top), not detected by the algorithm, associated with very severe abnormality (grade 8, isoelectric or persistent low voltage) of EEG background.

Table 2A
Results of visual scoring of EEG and automated seizure detection in patients with mild to moderate abnormalities (grades 1–5) of EEG background.

No.	EEG gr	Morphol	Freq	Amp	Dur	Arrhy%	Sz detection		Sens	Fp	PPV	Fp/h	durFp
							No.	Burden					
1	1	Mixed, osci	2	60	25	0	52/53	1009/1343	98	21	84	0.88	234
2	2	Shw	1	50	21	0	10/18	93/376	56	1	92	0.04	17
3	5	Shw	1	35	86	80	28/48	2076/4136	58	19	60	0.42	512
4	5	Shw	1	50	49	8	30/34	1432/1656	88	34	47	2	1947
5	1	Mixed	1	150	66	11	56/63	2455/4126	89	0	100	0	0
6	2	Osc, shw	2	100	174	15	12/13	1908/2128	92	4	75	0.17	99
7	2	Shw	5	50	40	6	104/109	4078/4356	95	0	100	0	0
8	2	Mixed	2	60	108	0	8/8	745/860	100	0	100	0	0
9	2	Mixed, osc, shw	7	100	63	4	93/98	5757/6195	95	0	100	0	0
10	4	Mixed, shw	1	65	18	100	6/7	109/127	86	6	50	0.26	113

3.3. Performance of the algorithm

The results are given in Tables 2A–3. When both definite and dubious seizures were considered together, 1285/2077 seizures were detected, giving a total sensitivity (SensT, the detected percentage of all seizures present in the data set) of 61.9%. The detected seizure burden was 66,244/97,574 s (67.9%). The average of all sensitivities of the 24 patients (SensT_PP) was 65.9% (median 85.1%, range 0–100), with a mean PPV of 73.7% (median 86.6%, range 10–100). In four patients with severely abnormal EEG background activity (gpII) and predominantly (>90%) dubious seizures, the algorithm performed very poorly (SensPP 0–7%). As it was doubtful whether this recurring paroxysmal activity (Fig. 3) constituted genuine seizures, we excluded these patients. In the remaining 20 patients (totalling 643 h of EEG data), the algorithm showed a SensT_PP of 86.9%, PPV of 89.5% and Fp h^{-1} of 0.28 h^{-1} [1263/1512 seizures detected (SensT of 83.5%), that is, 76.8% or 65,958/85,904 s of the seizure burden].

3.4. Comparison of seizure detection in patients in EEG gpI and gpII

The algorithm showed a trend towards increased sensitivity in patients with mild to moderate abnormalities of EEG background (gpI) (after excluding patients with 'dubious' seizures) but the difference was not statistically significant (Table 3).

3.5. Electrographic characteristics of missed seizures in 20 patients

The algorithm showed a low sensitivity (32–55%) in four patients in gpII. In patient Nos. 15 and 19, low amplitude arrhythmic electrographic discharges constituted about 60% of the seizures. In patient Nos. 16 and 20, the missed seizures were of lower amplitude (20–35 μV) and shorter (15–20 s), but similar in morphology and location to the detected seizures.

In two patients in gpI, a relatively lower sensitivity (56 and 58%) was seen: in patient No. 2, low amplitude seizures (20–30 μV) were missed whilst in patient No. 3 with a moderately severe

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

Table 2B
Results of visual scoring of EEG and automated seizure detection in patients with severe abnormalities (grades 6–8) of EEG background.

No.	EEG gr	Morphol	Freq	Amp	Dur	Arrhy%	Sz detection		Sens	Fp	PPV	Fp/h	durFp
							No.	Burden					
11	8	Mixed	1	150	37	0	110/112	3931/4114	98	6	95	0.29	98
12	8	Mixed, osc	8	10	20	100	0/210	0/4200	0	45	0	0.09	45
13	8	Osc	4	30	20	100	1/70	13/1400	1	8	11	0.33	180
14	6	Shw	2	25	106	0	47/50	4059/5282	94	3	94	0.13	99
15	6	mixed	3	35	35	60	30/72	727/1853	42	7	81	0.17	129
16	8	Osc	1	150	32	0	18/33	783/1047	55	15	55	0.65	231
17	8	mixed, osc	8	50	89	44	95/113	6008/10,104	84	1	99	0.04	42
18	8	Shw	5	20	33	0	169/174	5584/5755	97	10	94	0.42	596
19	6	Mixed, osc	8	25	20	65	14/44	240/880	32	12	54	0.41	347
20*	8	Mixed, osc	3	25	30	0	10/27	360/1481	37	69	13	3.45	3588
21	8	Mixed, shw, osc	6	25	25	100	12/156	165/3490	7	0	100	0	0
22	8	Mixed, shw	3	30	75	10	170/200	10,069/14,994	85	10	94	0.42	200
23*	7	Mixed, shw	7	15	20	90	9/129	108/2580	7	80	10	3.2	4603
24	8	Mixed, shw, osc	1	50	64	30	201/236	14,535/15,091	85	31	87	0.47	1026

EEG gr, EEG background grade; Morphol, seizure morphology (osc, oscillations; shw, sharp waves; mixed, mixed patterns); Freq, seizure frequency (Hz); Amp, amplitude (μV); Dur, duration (s); Arrhy%, percentage of arrhythmic seizures; Sz detection, No. (number) and burden (cumulative duration in s) of seizures detected by the algorithm; Sens, sensitivity per patient; Fp, No. of false positive detections. Values reported are means. PPV, positive predictive value; Fp/h, false positives per hour; durFp, total duration of false positives (s). Patient nos. 12, 13, 21 and 23, highlighted in grey, with persistent, severely abnormal EEG background and predominantly low-amplitude arrhythmic discharges (dubious seizures) were excluded from the final analysis.
* Patient nos. 20 and 23 had a high number of false positive detections due to electrode and respiratory artefacts.

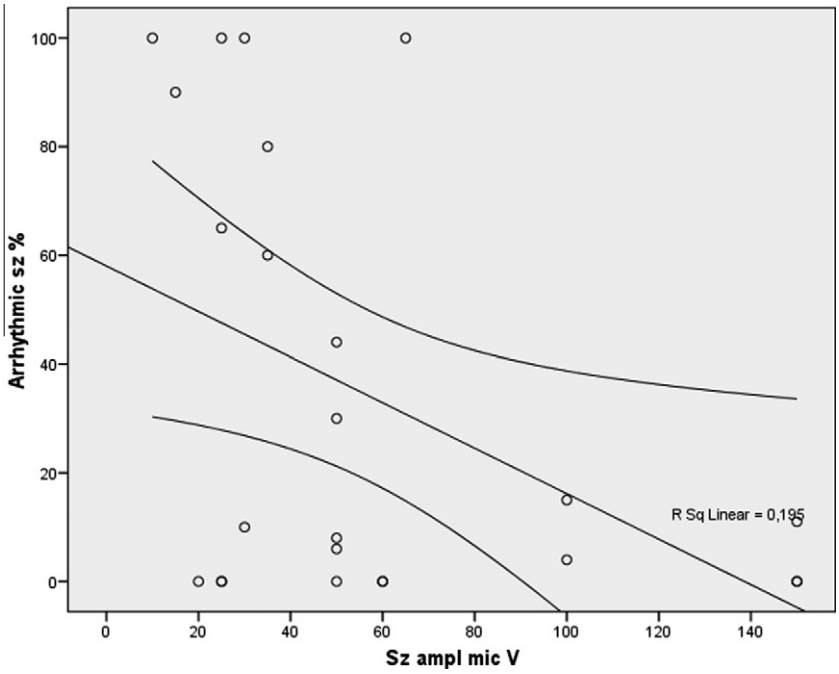


Fig. 4. Scatter plot showing a trend for increasing percentage of arrhythmic seizures to be associated with lower seizure amplitudes (Spearman's rho = −0.34, $p = 0.11$). Line of fit with 95% CI is shown.

EEG background abnormality (grade 5), low amplitude arrhythmic seizures constituted about 80% of the seizures.

3.6. Artefacts

The number of false positive detections per patient is shown in Tables 2A and 2B. Both biological (pulsation, respiratory movements, eye movements including nystagmus and tremor) and non-biological (electrode, ventilator and bed warming) artefacts caused false positive detections. Two patients showed a higher number of artefacts. In one (patient No. 23), these were related to head movements associated with respiration, involving O1 or O2 electrodes. In the other (patient No. 20), majority of the false positives were related to artefacts involving the C3 electrode.

4. Discussion

4.1. Performance of the algorithm

In this new and extensive data set, the neonatal seizure detection algorithm, NeoGuard, showed a good sensitivity and PPV, suggesting its suitability for cEEG monitoring in a NICU. No patients

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

Table 3
Comparison of seizure characteristics and the performance of the algorithm in patients with mild to moderate (gpI, grades 1–5,) and severe (gpII, grades 6–8) abnormalities of EEG background activity.

Characteristic	gpI (n = 10)	gpII (n = 14)	p value[†]
Duration EEG (h)	25 (17–78)	24 (20–66)	0.75
Sz number	41 (7–109)	112 (27–236)	0.008
Sz duration (s)	56 (18–174)	33 (20–106)	0.25
Sz frequency (Hz)	1.25 (1–7)	3 (1–8)	0.052
Sz amplitude (μV)	60 (35–150)	30 (15–150)	0.007
Sz burden (s)	1892 (127–6195)	3802 (880–15,091)	0.16
Arrhyth sz %	7 (0–100)	30 (0–100)	0.44
SensT_PP ($n = 24$)	90.6 (56–100)	55 (1–98)	0.015
($n = 20$)*	90.6 (56–100)	84.6 (32–98)	0.104
PPV (n = 24)	88 (47–100)	86.6 (10–100)	0.38
(n = 20)*	88 (47–100)	90.3 (13–99)	0.595
Det. sz burden (s)			
($n = 24$)	19,662/25,303 (77.7%)	46,582/72,271 (64.5%)	0.10
($n = 20$)[*]	19,662/25,303 (77.7%)	46,296/60,601 (76.4%)	0.55
Fp/h ($n = 24$)	0.11 (0–2)	0.41 (0–3.5)	0.158
($n = 20$)[*]	0.11 (0–2)	0.42 (0.04–3.5)	0.128

H, hours; Sz, seizure; Arrhyth sz %, percentage of seizures that were arrhythmic; Sz burden, cumulative duration of seizures. Performance of the algorithm: SensT_PP, average of sensitivities in each patient; PPV, positive predictive value; Det. sz burden, cumulative duration of seizures detected; Fp/h, false positives per hour.
[*] For the final analysis, 4 patients in gpII with persistent severe abnormality of EEG background and having predominantly dubious seizures were excluded. Values reported are medians (range).
[†] Mann–Whitney test.

with seizures were missed. Though we did not compare the results of the algorithm with visual detection using multichannel aEEG, a potential advantage of the automated detection is its good sensitivity (86.9%) and high PPV (89.5%). Use of aEEG alone (one or two channels) has been reported to show a low sensitivity (27–56%) and low inter-observer agreement (0.3) even amongst expert users (Shah et al., 2008).

It is difficult to make an actual comparison of this performance with that of other algorithms. Every research group may have a (often unconscious) bias to include patients and seizure types that best suits their detection method. Majority of the authors give no details of the characteristics of their patients. This is an important omission, as we show in our data that seizure characteristics like amplitude, duration and rhythmicity are influenced by changes in the EEG background. Ideal way forward would be to do a multi-centre study, where a broad and well-defined patient group is selected, an inter-observer agreement for visual scoring of seizures (in combination with varying severity of EEG background) is tested, and subsequently, a head-to-head comparison of seizure detection methods is carried out.

4.2. Inter-observer agreement for visual scoring of seizures

It is not surprising that inter-observer agreement was only fair (Kappa value $\kappa = 0.4$) between two expert raters for visual scoring of seizure occurrence. Correlation of human experts for scoring seizure data from epilepsy patients has identified that both very short and long seizures, high seizure rate per hour and ambiguous offsets can result in poor correlation (Wilson et al., 2003). As opposed to patients with chronic epilepsy, such situations are encountered more frequently in neonatal seizures occurring in encephalopathic infants in the NICU, making a compelling argument for the use of automated seizure detection.

4.3. Missed seizures

Seizures with very low amplitude and short duration were missed by the algorithm and this problem has also been reported by other authors (Mitra et al., 2009). We also found that the automatic detection of arrhythmic seizures (predominantly oscillations or mixed patterns) of low amplitude to be poor, whilst arrhythmic seizures with sharp wave morphology were well detected. In our cohort, arrhythmic seizures of low amplitude occurred only in patients with severely abnormal EEG background. Further adjustments of seizure detection parameters of the algorithm to detect similar seizures may not add clinical value, and might come at the cost of more false positive detections.

4.4. The problem of dubious seizures

We encountered two types of dubious seizures. The algorithm performs differently with each type. The first type is the low-amplitude arrhythmic discharges with poor repetitiveness seen in four of our patients with very severely abnormal EEG background. These discharges, though occurring paroxysmally, do not fit well into our operational definition of seizures, as their beginning and end are difficult to differentiate from the EEG background and they do not show a clear-cut evolution in amplitude and frequency. As expected, the algorithm fails to detect majority of these discharges. The patients showing these patterns had also severely abnormal outcomes (death or severe disability) despite receiving optimal medical care, including aggressive treatment of subclinical seizure discharges under EEG guidance, comparable to that received by our other patients. The atypical morphology and evolution also suggest a different underlying mechanism for these discharges. They are probably signatures of severe underlying cortical injury, and treating them aggressively with AED is unlikely to offer a clinical benefit. Due to these reasons, we feel that leaving four patients out in our final analysis of the performance of the algorithm is justified (thereby including only patients who are 'clinically relevant'). The second type of dubious seizure is more difficult to categorise and we encountered this in a single patient with an ischaemic stroke involving the right middle cerebral artery territory. They constituted brief rhythmic discharges (BRDs) lasting 5–7 s or LPDs (also known as PLEDs) (Hirsch et al., 2005). Although they did not fit well into our definition of seizures (shorter duration or no clear-cut evolution) and were not scored as seizures by the clinical neurophysiologist, many of these discharges were detected by the algorithm. They were localised to the brain regions that had also expressed definite seizures and hence we did not classify these detections as false positives. Some authors feel that BRDs have the same clinical significance as longer electrographic seizures (Oliveira et al., 2000). LPDs in neonates probably have the same clinical significance as in older children and adults (Scher and Beggarly, 1989). It is however controversial whether they constitute ictal or interictal phenomena.

4.5. Continuing challenge of artefacts

The artefact detection capabilities of the algorithm have been improved, especially by reducing ECG artefacts. As expected when cEEG monitoring is done for longer periods in a NICU, more artefacts (cumulative duration Fp) were incorrectly detected. However, the Fp h^{-1} were improved from 0.66 to 0.28 h^{-1} (Deburchgraeve et al., 2008). Nonetheless, this effect of improvement is largely explained by a few patients who showed many artefacts. In the majority of patients, the performance of the algorithm was already very good and not hindered by false positive detections.

Both biological and non-biological artefacts offer challenges to automated seizure detection. Technical innovation can help in

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

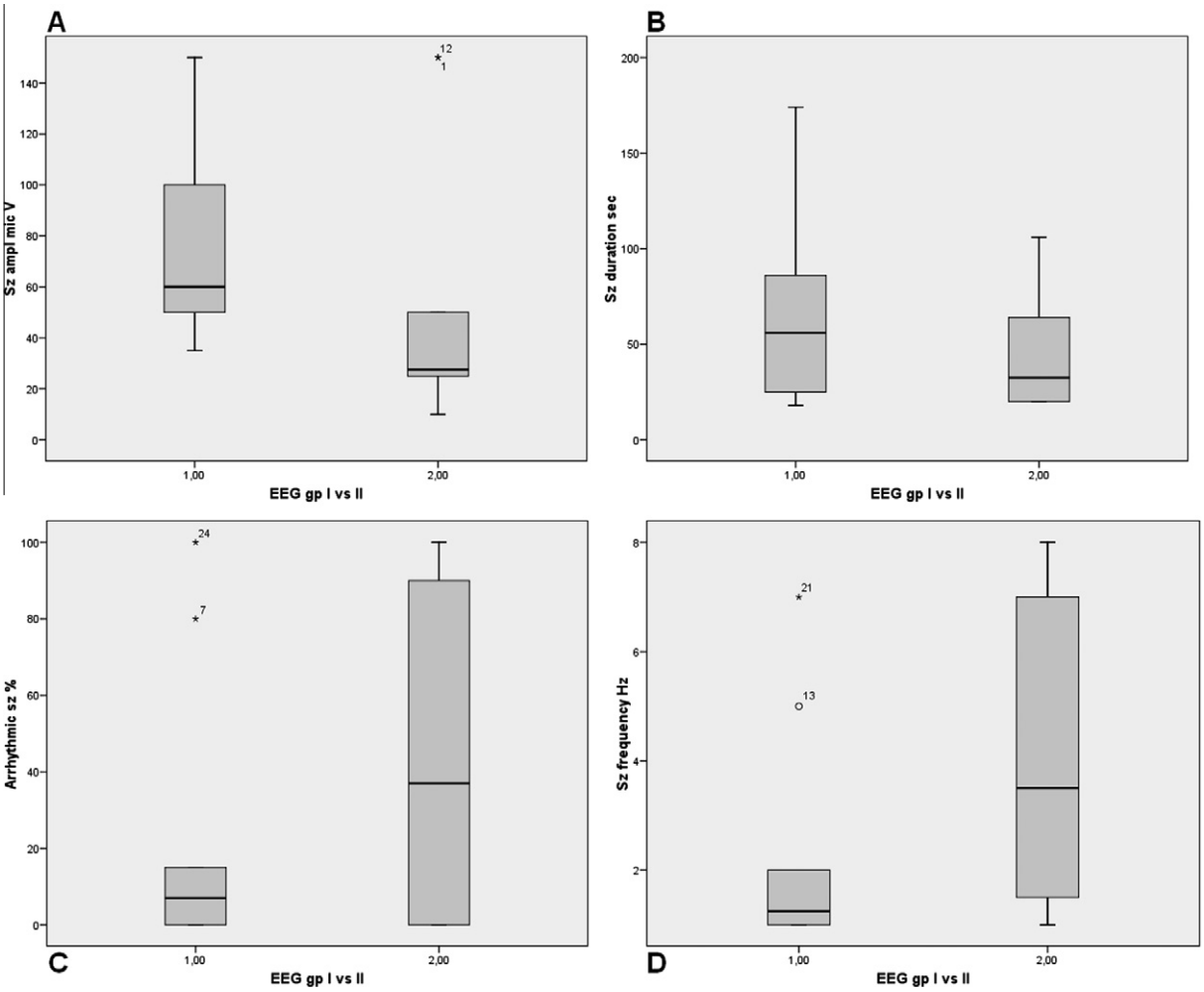


Fig. 5. Box plots comparing various ictal parameters ((A) amplitude in μV, (**B**) duration in seconds, (C) percentage of arrhythmic seizures and (D) seizure frequency in Hz) of individual seizures in patients with mild to moderate (gp**I** = grades 1–5, $n = 10$) and severe (gpII = grades 6–8, $n = 14$) abnormalities of EEG background. Only the amplitude of seizures was significantly different between the groups (Mann–Whitney $U = 24.5$, $p = 0.007$).

reducing the false positives due to biological signals like ECG, tremor, respiration, etc. The technical details of the improvements made to our algorithm are described in detail elsewhere (Deburchgraeve, 2010). External, non-biological artefacts are more challenging and the best solution is for the EEG technician and the clinical neurophysiologist to remain alert to them during the monitoring. Electrode artefacts, pulsation artefacts, etc. can be reduced by slightly repositioning the patient's head or the electrodes. In other words, to be successful, cEEG monitoring needs to have ongoing interaction between the caregivers (neurotechnologist, neonatologist, nursing staff and the clinical neurophysiologist). The automatic detections need to be visually verified first and subsequently, the algorithm can help to monitor the response to treatment. Automated seizure detection is best viewed as a tool to reduce the workload, and should not be seen as a black box into which EEG data can be fed with a resulting output that gives all the answers for therapeutic decision making.

4.6. Clinical significance of seizure detection

Automated seizure detection can guide treatment decisions, help prognostication and also help to study the pathophysiology of seizures in various encephalopathies. From a practical viewpoint, one can argue that this technique is clinically relevant only when the detection of subclinical electrographic seizures results in treatment decisions that ultimately are beneficial to the patient. It would be meaningful only when the brain injury or dysfunction is in a potentially reversible stage. This is best done by combining seizure detection with the studying of dynamic changes in EEG background activity (Scheuer and Wilson, 2004). Abnormality of EEG background activity and its evolution over time are ideally suited to assess the severity and course of the underlying encephalopathy (Biagioni et al., 1999; Holmes and Lombroso, 1993; Watanabe et al., 1980b). However, whether the seizure patterns themselves are affected by the severity of hypoxic brain injury has not been previously studied. With these considerations in mind, we assessed the morphological characteristics of electrographic seizures as well as the performance of the automated seizure detection algorithm in the context of EEG background activity. Even with the relatively small number of patients, we were able to show that seizure characteristics such as total seizure number, burden, amplitude, frequency and rhythmicity are influenced by the severity of the underlying encephalopathy and in turn, might affect seizure detection. This type of relevant information about seizure characteristics is lost when NICUs rely only on simple monitoring tools like aEEG. We are presently using these ictal characteristics in our detection algorithm and these can easily be incorporated into the output/visual display. These seizure characteristics need to be studied in larger cohorts. We feel that the presently achieved sensitivity and false positive rate of NeoGuard denotes a good balance. It

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043

remains a research tool (not commercially available), albeit a user-friendly one. We will make it available for wider use by researchers after completing a multi-centre evaluation.

More research into the electrographic characteristics of neonatal seizures (Shewmon, 1990) is needed, which will help us to better understand the phenomena that we are trying to detect. Development of a large data set of seizures (categorised according to severity of EEG background abnormality), with assessment of inter-rater agreement between centres is urgently needed. This will also help in comparing the performance of different seizure detection algorithms. Whilst developers are busy with the methodological and technical issues of automated detection, important research needs to be done in filling the gaps in present knowledge about the pathophysiology of neonatal seizures in various encephalopathies as well as the effect of their treatment.

Financial Support

W.D. was supported by a PhD grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen); the Belgian Federal Science Policy Office IUAP P6/O4 (DYSCO, 'Dynamical systems, control and optimization', 2007–2011); GOA Manet; GOA AMBioRICS and FWO project G.034107 N on data fusion.

M.V. was supported by a postdoctoral fellowship from the Katholieke Universiteit Leuven, Belgium.

Acknowledgement

We thank our neurotechnologists, especially Ms. Els Bröker and Ms. Jolanda Geerlings, for doing the EEG registrations and our patients and their parents for their cooperation.

Part of this work was presented as a poster at the 64th Annual Meeting of the American Epilepsy Society in December 2010, at San Antonio, Texas.

References

Aarabi A, Wallois F, Grebe R. Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis. Clin Neurophysiol 2006;117:328–40.

Biagioni E, Bartalena L, Boldrini A, Pieri R, Cioni G. Constantly discontinuous EEG patterns in full-term neonates with hypoxic–ischaemic encephalopathy. Clin Neurophysiol 1999;110:1510–5.

Bourez-Swart MD, van Rooij L, Rizzo C, de Vries LS, Toet MC, Gebbink TA, et al. Detection of subclinical electroencephalographic seizure patterns with multichannel amplitude-integrated EEG in full-term neonates. Clin Neurophysiol 2009;120:1916–22.

Bye AM, Cunningham CA, Chee KY, Flanagan D. Outcome of neonates with electrographically identified seizures, or at risk of seizures. Pediatr Neurol 1997;16:225–31.

Bye AM, Flanagan D. Spatial and temporal characteristics of neonatal seizures. Epilepsia 1995;36:1009–16.

Cherian PJ, Swarte RM, Visser GH. Technical standards for recording and interpretation of neonatal electroencephalogram in clinical practice. Ann Indian Acad Neurol 2009;12:58–70.

Deburchgraeve W. Development of an automated neonatal EEG seizure monitor. Department of Electrical Engineering. PhD Thesis. Leuven: Katholieke Universiteit; 2010.

Deburchgraeve W, Cherian PJ, De Vos M, Swarte RM, Blok JH, Visser GH, et al. Automated neonatal seizure detection mimicking a human observer reading EEG. Clin Neurophysiol 2008;119:2447–54.

Faul S, Boylan G, Connolly S, Marnane L, Lightbody G. An evaluation of automated neonatal seizure detection methods. Clin Neurophysiol 2005;116:1533–41.

Gotman J, Flanagan D, Zhang J, Rosenblatt B. Automatic seizure detection in the newborn: methods and initial evaluation. Electroencephalogr Clin Neurophysiol 1997;103:356–62.

Greene BR, Boylan GB, Reilly RB, de Chazal P, Connolly S. Combination of EEG and ECG for improved automatic neonatal seizure detection. Clin Neurophysiol 2007;118:1348–59.

Hirsch LJ, Brenner RP, Drislane FW, So E, Kaplan PW, Jordan KG, et al. The ACNS subcommittee on research terminology for continuous EEG monitoring: proposed standardized terminology for rhythmic and periodic EEG patterns encountered in critically ill patients. J Clin Neurophysiol 2005;22:128–35.

Holmes G, Rowe J, Hafford J, Schmidt R, Testa M, Zimmerman A. Prognostic value of the electroencephalogram in neonatal asphyxia. Electroencephalogr Clin Neurophysiol 1982;53:60–72.

Holmes GL, Lombroso CT. Prognostic value of background patterns in the neonatal EEG. J Clin Neurophysiol 1993;10:323–52.

Laroia N, Guillet R, Burchfiel J, McBride MC. EEG background as predictor of electrographic seizures in high-risk neonates. Epilepsia 1998;39:545–51.

Lombroso CT. Neonatal EEG polygraphy in normal and abnormal newborns. In: Niedermeyer E, Lopes da Silva F, editors. Electroencephalography, basic principles, clinical applications, and related fields. Baltimore: Williams & Wilkins; 1993. p. 803–75.

Lombroso CT. Neonatal seizures: A clinician's overview. Brain Dev 1996;18:1–28.

McBride MC, Laroia N, Guillet R. Electrographic seizures in neonates correlate with poor neurodevelopmental outcome. Neurology 2000;55:506–13.

Mitra J, Glover JR, Ktonas PY, Thitai Kumar A, Mukherjee A, Karayiannis NB, et al. A multistage system for the automated detection of epileptic seizures in neonatal electroencephalography. J Clin Neurophysiol 2009;26:218–26.

Murray DM, Boylan GB, Ali I, Ryan CA, Murphy BP, Connolly S. Defining the gap between electrographic seizure burden, clinical expression and staff recognition of neonatal seizures. Arch Dis Child Fetal Neonatal Ed 2008;93:F187–91.

Murray DM, Boylan GB, Ryan CA, Connolly S. Early EEG findings in hypoxic–ischemic encephalopathy predict outcomes at 2 years. Pediatrics 2009;124:e459–67.

Navakatikyan MA, Colditz PB, Burke CJ, Inder TE, Richmond J, Williams CE. Seizure detection algorithm for neonates based on wave-sequence analysis. Clin Neurophysiol 2006;117:1190–203.

Oliveira AJ, Nunes ML, Haertel LM, Reis FM, da Costa JC. Duration of rhythmic EEG patterns in neonates: new evidence for clinical and prognostic significance of brief rhythmic discharges. Clin Neurophysiol 2000;111:1646–53.

Pezzani C, Radvanyi-Bouvet MF, Relier JP, Monod N. Neonatal electroencephalography during the first twenty-four hours of life in full-term newborn infants. Neuropediatrics 1986;17:11–8.

Rennie JM, Chorley G, Boylan GB, Pressler R, Nguyen Y, Hooper R. Non-expert use of the cerebral function monitor for neonatal seizure detection. Arch Dis Child Fetal Neonatal Ed 2004;89:F37–40.

Sarnat HB, Sarnat MS. Neonatal encephalopathy following fetal distress. A clinical and electroencephalographic study. Arch Neurol 1976;33:696–705.

Scher MS, Beggarly M. Clinical significance of focal periodic discharges in neonates. J Child Neurol 1989;4:175–85.

Scheuer ML, Wilson SB. Data analysis for continuous EEG monitoring in the ICU: seeing the forest and the trees. J Clin Neurophysiol 2004;21:353–78.

Shah DK, Mackay MT, Lavery S, Watson S, Harvey AS, Zempel J, et al. Accuracy of bedside electroencephalographic monitoring in comparison with simultaneous continuous conventional electroencephalography for seizure detection in term infants. Pediatrics 2008;121:1146–54.

Shellhaas RA, Soaita AI, Clancy RR. Sensitivity of amplitude-integrated electroencephalography for neonatal seizure detection. Pediatrics 2007;120:770–7.

Shewmon DA. What is a neonatal seizure? Problems in definition and quantification for investigative and clinical purposes. J Clin Neurophysiol 1990;7:315–68.

Swarte R, Lequin M, Cherian P, Zecic A, van Goudoever J, Govaert P. Imaging patterns of brain injury in term-birth asphyxia. Acta Paediatr 2009;98:586–92.

Vanhatalo S. Development of neonatal seizure detectors: An elusive target and stretching measuring tapes. Clin Neurophysiol 2011;122:435–7.

Vento M, de Vries LS, Alberola A, Blennow M, Steggerda S, Greisen G, et al. Approach to seizures in the neonatal period: a European perspective. Acta Paediatr 2010;99:497–501.

Watanabe K, Miyazaki S, Hara K, Hakamada S. Behavioral state cycles, background EEGs and prognosis of newborns with perinatal hypoxia. Electroencephalogr Clin Neurophysiol 1980a;49:618–25.

Watanabe K, Miyazaki S, Hara K, Hakamada S. Behavioral state cycles, background EEGs, and prognosis of newborns with perinatal hypoxia. Electroencephalogr Clin Neurophysiol 1980b;49:618–25.

Wilson SB, Scheuer ML, Plummer C, Young B, Pacia S. Seizure detection: correlation of human experts. Clin Neurophysiol 2003;114:2156–64.

Please cite this article in press as: Cherian PJ et al. Validation of a new automated neonatal seizure detection system: A clinician's perspective. Clin Neurophysiol (2011), doi:10.1016/j.clinph.2011.01.043