Clinical Neurophysiology 127 (2016) 156–168

Contents lists available at ScienceDirect

Clinical Neurophysiology

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

Validation of an automated seizure detection algorithm for term neonates

Sean R. Mathieson [b], Nathan J. Stevenson [a], Evonne Low [a], William P. Marnane [a], Janet M. Rennie [b], Andrey Temko [a], Gordon Lightbody [a], Geraldine B. Boylan [a,*]

[a] Neonatal Brain Research Group, Irish Centre for Fetal and Neonatal Translational Research, Department of Paediatrics and Child Health, University College Cork, Cork, Ireland
[b] Academic Research Department of Neonatology, Institute for Women's Health, University College London, London, United Kingdom

ARTICLE INFO

Article history:
Accepted 30 April 2015
Available online 9 May 2015

Keywords:
Neonatal seizures
Automated seizure detection
Neonatal EEG
Hypoxic-ischaemic encephalopathy
Neonatal neurology

HIGHLIGHTS

- Seizure detection algorithm (SDA) validated on unseen, unedited EEG of 70 neonates.
- Results at SDA sensitivity settings of 0.5–0.3 acceptable for clinical use.
- Seizure detection rate of 52.6–75.0%, false detection rate 0.04–0.36 FD/h.

ABSTRACT

Objective: The objective of this study was to validate the performance of a seizure detection algorithm (SDA) developed by our group, on previously unseen, prolonged, unedited EEG recordings from 70 babies from 2 centres.
Methods: EEGs of 70 babies (35 seizure, 35 non-seizure) were annotated for seizures by experts as the gold standard. The SDA was tested on the EEGs at a range of sensitivity settings. Annotations from the expert and SDA were compared using event and epoch based metrics. The effect of seizure duration on SDA performance was also analysed.
Results: Between sensitivity settings of 0.5 and 0.3, the algorithm achieved seizure detection rates of 52.6–75.0%, with false detection (FD) rates of 0.04–0.36 FD/h for event based analysis, which was deemed to be acceptable in a clinical environment. Time based comparison of expert and SDA annotations using Cohen's Kappa Index revealed a best performing SDA threshold of 0.4 (Kappa 0.630). The SDA showed improved detection performance with longer seizures.
Conclusion: The SDA achieved promising performance and warrants further testing in a live clinical evaluation.
Significance: The SDA has the potential to improve seizure detection and provide a robust tool for comparing treatment regimens.

1. Introduction

The concept of "neuroprotective" intensive care has now reached neonatal units worldwide, in part driven by the results of randomized controlled trials showing that therapeutic hypothermia is beneficial for term babies with a recent hypoxic-ischaemic injury (Glass et al., 2011). The practice of neuroprotective care involves careful monitoring of carbon dioxide tension, blood pressure and other physiological variables and is ideally accompanied by continuous cotside EEG monitoring. Without EEG monitoring many seizures are missed. The inaccuracy of clinical recognition of seizures was demonstrated by Murray et al. (2008). In this study, comparing EEG evidence of seizures to the seizure detection acumen of NICU staff based on clinical evidence alone, of 526 EEG seizures, only 179 (34%) had any clinical accompaniment, overdiagnosis was common with only 48 of 177 (27%) clinically suspected events accompanied by EEG seizures such that only 48/526 (9%) of EEG seizures were correctly identified by clinical observation. Amplitude-integrated EEG (aEEG) is widely used in NICUs for seizure detection but has been shown to perform poorly (Rennie et al., 2004). In this study seizure detection by four non-experts using CFM traces at 3 paper speeds were compared against simultaneous EEG in 19 babies. Sensitivities of only 38%

* Corresponding author. Tel.: +353 21 4205040.
E-mail address: g.boylan@ucc.ie (G.B. Boylan).

http://dx.doi.org/10.1016/j.clinph.2015.04.075
1388-2457/

at 6 cm/h, 54% at 15 cm/h and 55% at 30 cm/h were achieved and agreement between observers was poor at all speeds (κ values from 0.01 to 0.39). Treating seizures to electrical quiescence has yet to be proven of any long-term benefit, but there is evidence from animal models (Wirrell et al., 2001), and clinical studies (Glass et al., 2009; Shah et al., 2014) which would support the principle that seizures do inflict further brain injury. Attempts to ameliorate such damage must be accompanied by prompt and reliable detection of seizures. In addition, good quality randomized controlled trials of new antiepileptic drugs are impossible without robust and reproducible EEG monitoring.

A significant barrier to the practice of neuroprotective critical care in the NICU is the lack of expertise in reporting neonatal EEG. Current cotside EEG monitors are sophisticated devices, offering the ability to record multiple channels of EEG continuously together with other physiological signals and video-recording of the baby's movements. They allow the continuous display of aEEG and other quantitative trends and are easy to set up and maintain. But few clinicians have the knowledge to interpret the plethora of information which is generated by such monitoring, and without this knowledge there is a danger that this equipment will be under utilised or (worse) the output will be misinterpreted at the cotside.

Our group has considerable experience with cotside EEG monitoring and has grown to appreciate the benefits that this provides. For many years now we have been working on a seizure detection algorithm (SDA), which would analyse one or more channels of "raw" EEG, continuously and in real-time, providing a visual and audible alert to the clinical team. The engineering challenges have proven formidable because EEG is a complex signal, and neonatal seizures have variable amplitude, frequency and morphology, and are rarely sustained for more than 5 min.

Other groups have developed SDAs for neonates, and have published their detection rates, using varying definitions of success (Liu et al., 1992; Gotman et al., 1997; Smit et al., 2004; Navakatikyan et al., 2006; Deburchgraeve et al., 2008; Mitra et al., 2009). Details of the performance of these and other SDAs are outlined in Table 1 and reviewed further in the discussion. Currently only two SDAs are commercially available. These are the Gotman algorithm incorporated into the Stellate EEG system (Natus Medical Inc, USA); and the 'Recognize' algorithm of Navakatikyan which is incorporated into the Brainz aEEG monitor (Natus Medical Inc., USA) which has only a 2 channel EEG capability. One problem which inhibits comparison of SDAs is the lack of an agreed definition of what constitutes best performance. Many SDAs are reported to have good detection rates, with a high number of seizures accurately detected when compared to expert neurophysiology as the "gold standard", and low numbers of missed seizures. However, the temporal aspect of seizure detection is rarely reported, for example one missed seizure of 8 min duration in an hour would be clinically important. Another important aspect of SDA performance assessment is the number of false detections. Many validation studies have used only short duration recordings, but any robust algorithm designed for current NICU use has to be able to perform reliably on very long recordings of 72 h or more. Respiration artefact is a particular problem often recorded in neonatal EEG and can mimic the stereotyped rhythmic seizure activity that is often seen in neonates.

We have previously reported the performance of our neonatal SDA on a set of 17 seizure babies recorded at Cork University Maternity Hospital (CUMH), Ireland (Temko et al., 2011a) using a 'leave one out' (LOO) cross validation method of analysis, whereby the data of one patient is used for testing and the others used for training the algorithm and the process is repeated for each patient and the mean result reported. A further LOO study was performed on 38 babies from CUMH (Temko et al., 2013) incorporating an adaptation to reduce the effects of prolonged artefact and showed improved performance. This study also incorporated analysis of an 'unseen' dataset of 51 babies from CUMH.

The aim of the present study was to validate the performance of our neonatal SDA on a larger database of unseen, unedited, continuous, multi-channel EEG data from 70 term newborns collected at 2 sites, CUMH and University College London Hospital (UCLH), and to provide comprehensive measures of SDA performance. While time based metrics assess the ability of the algorithm to detect the 'amount' of seizure activity (seizure burden) correctly and is, in a sense, the most precise engineering metric, event based metrics provide clinicians with valuable information as to the percentage of seizures that will be detected, with important implications for treatment and also how often the algorithm is likely to alarm falsely. We therefore report both time based and event based measures of performance.

2. Methods

2.1. Data acquisition and EEG annotation

Neonates were enrolled from the neonatal intensive care units of CUMH and UCLH from January 2009 to October 2011 as part of an on-going study of neonatal seizures. Neonates $\geqslant$37 weeks gestation were enrolled for EEG monitoring if they fulfilled two or more of the following criteria: Apgar score less than six at five minutes; a continued need for resuscitation after birth; any clinical evidence of encephalopathy, or seizures developed within 72 h of age.

This study was conducted with approval from the Clinical Research Ethics Committees of the Cork Teaching hospitals, Ireland and the National Health Service in the UK, via the Integrated Research Application Service. Written, informed consent was obtained from at least one parent of each neonate who participated in this study.

Table 1
Summary of SDAs proposed in the literature. (DB – database, h – hour, S – seizure, NS – non-seizure, Dur – duration, AUC – area under the receiver operator characteristic, Sens – sensitivity, spec, specificity, SDR – seizure detection rate, FA/h – false alarms per hour).

Algorithm	DB length h (*N*)	S:NS Dur	NS neonates	AUC	Sens (%)	Spec (%)	SDR (%)	FA/h (N/h)
Liu et al. (1992)	1.0 (14)	1:1	Yes		84	98		
Gotman et al. (1997)	237 (54)		Yes				66	2.3
Smit et al. (2004)	10.4 (19)		No		66	90		
Navakatikyan et al. (2006)	24 (55)	1:6.8	Yes		83	87	90	2
Lawrence et al. (2009)	2708 (40)		Yes				55	0.09
Deburchgraeve et al. (2008)	218 (26)		Yes				85	0.66
Cherian et al. (2011)	756 (24)	1:27.9	No		59		66	0.58
Mitra et al. (2009)	120 (76)	1:11.0	Yes				80	0.78
Temko et al. (2011a,b)	268 (17)	1:5.9	No	0.96	90	90	89	1
Temko et al. (2013)	2540 (51)		Yes	0.96			71	0.25