

# Interobserver variation when interpreting prognostic SSEPs: Do we agree when a response is absent?

W. McDevitt, P. Bill, P. Davies, B. Scholefield

## Abstract

**Purpose:** To determine the rate of agreement when interpreting prognostic Somatosensory Evoked Potentials (SSEPs) as absent, present or indeterminate.

**Methods:** SSEPs recorded in comatose paediatric cardiac arrest patients were randomised into a web-based questionnaire & sent to Consultant Neurophysiologists (Cons) and Healthcare Scientists (HS) within the West Midlands region for interpretation.

**Results:** Overall, 40% of SSEPs were interpreted with unanimous agreement ( $n=12$ ). 65% amongst Cons ( $n=5$ ) and 43% amongst HS groups ( $n=7$ ).  $0.62$  was the overall rate of agreement between all participants. This rose when interpreting present cortical responses ( $K=0.92$ ) and fell with absent ( $K=0.52$ ) and indeterminate ( $K=0.36$ ) responses. Cons had a higher rate of agreement ( $K=0.68$ ) than HS ( $K=0.62$ ) & duplicate SSEPs were interpreted consistently in 91% of cases.

**Conclusions:** Variation exists when interpreting prognostic SSEPs. The rate of agreement decreases when interpreting cortical SSEPs as absent or indeterminate which is alarming considering the implication of an absent cortical SSEP.

## Introduction

There is no national consensus in the UK regarding the use of prognostic SSEPs in the intensive care unit (ICU)<sup>1</sup>, despite its widespread use in medical management. This may be due to inherent bias in studies advocating its efficacy and the rate of interobserver variation attached to the test is not negligible<sup>1-3</sup>.

In this context, interobserver variation measures the amount observers disagree with one another when interpreting the same SSEP. Several studies have concluded that a degree of interobserver variation exists amongst experts when interpreting prognostic SSEPs<sup>1-3</sup>. There is a lack of evidence regarding *intraobserver* variability<sup>3</sup> (i.e. Does the expert interpret the same SSEP consistently) and to the knowledge of the authors, neither has been assessed in the UK.

Before creating a regional consensus regarding the use of prognostic SSEPs it is important to address these issues. Therefore, this study was conducted to assess the inter- and intraobserver variation between and within experts in the West Midlands when interpreting prognostic SSEPs as either absent, present or indeterminate.

## Methods

All SSEPs were recorded in children (Mean age 63.9 months  $\pm 73.8$ SD) who had suffered cardiac arrest and remained comatose after Cardio Pulmonary Resuscitation in the ICU of Birmingham Children's Hospital. SSEPs were recorded using Synergy software (Medelec Synergy®, Oxford Instruments) via electrical stimulation of the left and right median nerve (either at the wrist or elbow) using bipolar skin (Ag/Ag-Cl) electrodes. Evoked potentials (EP) were recorded via skin surface electrodes placed at Erb's point, at the cervical spine (CV2), and C3' & C4'. Two sets of at least 150 averaged EPs were recorded from each arm. The low and high frequency filters were set at 3Hz and 3kHz, respectively. Stimulus duration was 0.2ms and stimulation intensity started at 8mA and increased in multiples of 8mA until EPs were recorded. Time base was 50ms and the artefact reject was set at 50uV.

98 SSEPs (36 of which were repeats to assess intraobserver variability) were randomised into an online questionnaire & sent via email to 8 West Midlands Hospitals. Consultants (Cons) in Clinical Neurophysiology and the Lead Healthcare Scientist (HS) had to interpret whether the cortical component of the SSEP was either absent, present or indeterminate.

## Data analysis

The interobserver rate of agreement was calculated using Fleiss' kappa coefficient ( $K$ ).  $K$  values have been arbitrarily defined as poor, slight, fair, moderate, substantial and almost perfect agreement<sup>4</sup> (Table one). These were determined for the HS group, Cons group and overall. Percentages of unanimous agreement and consistent interpretations amongst repeat SSEPs were also calculated. This meant both inter- and intraobserver variability were assessed. 95% & 99% Confidence intervals (CI) are reported in brackets.

Fleiss' Kappa Coefficient	Level of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81+	Almost perfect

Table one: Agreement levels for Fleiss' Kappa Coefficients<sup>4</sup>

- Alternate hypothesis: Variation will exist between assessors.
- Null hypothesis: No variation will exist between assessors.

## Results

Overall, 40% (95% CI: 28-54) of SSEPs were interpreted with unanimous agreement ( $n=12$ ). 65% (95% CI: 51-76) amongst Cons group ( $n=5$ ) and 44% (95% CI: 30-57) amongst HS group ( $n=7$ ). \*Please refer to table two\*. Figure two highlights the challenges of prognostic SSEP interpretation when used to predict outcome.

$0.62$  (SE: 0.01) was the overall rate of agreement between all participants. Agreement rose when interpreting cortical responses as present ( $K=0.92$ , SE: 0.02) and fell with absent ( $K=0.52$ , SE: 0.02) and indeterminate ( $K=0.36$ , SE: 0.02) interpretations. Consultants had a higher rate of agreement ( $K=0.68$ , SE: 0.03) than Healthcare Scientists ( $K=0.62$ , SE: 0.02) and the  $p$  value for all  $K$  statistics = <0.0001 (Table three).

Duplicate SSEPs were interpreted consistently in 91% of cases (99% CI: 0.86 - 0.95). The difference between HS & Cons groups was marginal (0.5%) \*Please refer to figure one\*.

Assessor group	Number of unanimously agreed SSEPs	Percentage/ %	95% Confidence interval
Healthcare Scientist	27	43.55	30.99 - 56.74
Consultant	40	64.52	51.34 - 76.26
Total	25	40.32	28.05 - 53.55

Table two: The number and percentages of unanimous agreement of groups when interpreting SSEPs as absent, present and indeterminate

Cortical SSEP	Healthcare Scientists $n=7/K$	Consultants, $n=5/K$	Overall, $n=12/K$
Absent	0.46	0.54	0.52
Present	0.91	0.93	0.92
Indeterminate	0.43	0.29	0.36
Overall	0.62	0.68	0.62

Table three: The interobserver variation between groups, and as a whole, when interpreting SSEPs as absent, present and indeterminate

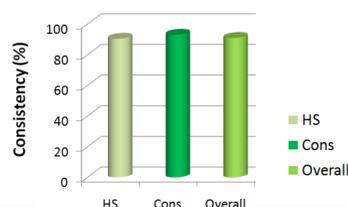


Figure one: A bar chart to show the percentages of consistent interpretation when assessors were asked to interpret duplicate SSEPs - a measure of intraobserver variation

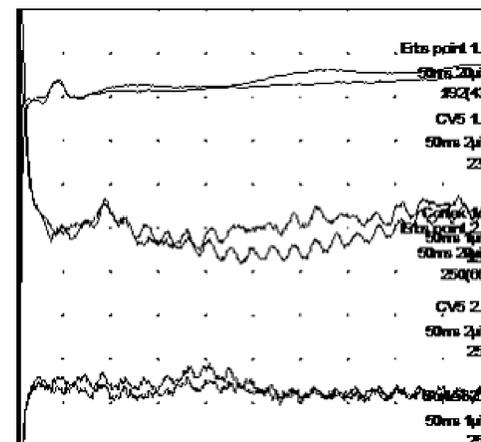


Figure two: A prognostic SSEP recorded in a 1 year old male with CHARGE syndrome; treated at BCH after suffering a 50 minute cardiac arrest. In our study, 55% of assessors deemed the trace as indeterminable & 45% as present. Sadly, the child was pronounced dead 4 days after recording this SSEP.

## Conclusions

- Interobserver variation exists when interpreting prognostic SSEPs.
- Higher rates of disagreement when interpreting absent and indeterminate SSEPs.
- SSEPs were interpreted consistently in 91% of duplicate SSEPs.

## Application

- Regional guidelines were created after the study and circulated within the West Midlands with the aim of reducing the rate of interobserver variability.

## Future work

- Re-audit in one years time to assess whether proposed guidelines have decreased the rate of interobserver variability.

## References

1. Friberg H, Cronberg T, Dunser MW, Duranteau J, Horn J, Oddo M. Survey on Current Practices for Neurological Prognostication after Cardiac Arrest. *Resuscitation* (2015), <http://dx.doi.org/10.1016/j.resusc.2015.08.011>: 158-162.
2. Sandroni C, Cariou A, Cavallaro F, Cronberg T, Friberg H, Hoedemaekers C, Horn J, Nolan JP, Rossetti AO, Soar J. Prognostication in comatose survivors of cardiac arrest: An advisory statement from the European Resuscitation Council and the European Society of Intensive Care Medicine. *Intensive Care Medicine* (2014), **40**: 1816 - 1831.
3. Hakimi K, Kinney G, Kraft G, Micklesen P, Robinson L. Reliability in Interpretation of Median Somatosensory Evoked Potentials in the setting of Coma: Factors and Implications. *Neurocritical care* (2009), **11**: 353-361
4. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* (1977), **33**: 159-174

