Formally evaluating the variability and generalizability of visual HFO evaluations: a new semi-supervised framework

Aaron M Spring[1,2,3], Daniel J Pittman[2,3], Yahya Aghakhani[1], Jeffrey Jirsch[4], Neelan Pillay[1], Luis E Bello-Espinosa[1,5], Colin Josephson[1], Paolo Federico[1,2,3,6]

[1]Department of Clinical Neurosciences, University of Calgary, Calgary, Canada
[2]Hotchkiss Brain Institute, University of Calgary, Calgary, Canada
[3]Seaman Family MR Research Centre, Foothills Medical Centre, Calgary, Canada
[4]Department of Medicine, University of Alberta, Edmonton, Canada
[5]Department of Paediatrics, University of Calgary, Calgary, Canada
[6]Department of Radiology, University of Calgary, Calgary, Canada

Rationale

HFO identification algorithms are tested against the gold standard of the agreement between two visual reviewers. However, the variability of these visual markings has yet to be formally evaluated. We propose a framework to formally evaluate the variability of visual markings, and facilitate physician-supervised algorithmic HFO identification.

Methods

Data from 10 iEEG datasets were filtered for ripples (80-250Hz) and root-mean-square (RMS) normalized. The following events were identified: 64 candidate HFOs (high threshold, 3.1 SD), 64 low threshold (2.3 SD), and 64 distractor events. Six epileptologists reviewed each event and determined whether an HFO was present. A nested Generalizability Theory model was used to estimate the generalizability and variability of the visual HFO evaluations.

Results

Table 1 lists the variances of the model facets. Model generalizability was 0.946 or 0.974 for evaluator or event, respectively.

Discussion

The highest source of variance (58%) was the event type, indicating the algorithms successfully identified candidate HFOs. The high variance between evaluators (10.6%) and high interaction between the evaluator and any given event (8.97%) highlights the large degree of subjectivity in the visual evaluations; however, the negligible interaction between the evaluator and event type (1.42%) suggests that each evaluator remains internally consistent between event types.

The high generalizability indicates that both universal HFO evaluations and internal evaluator thresholds can be inferred from the model. Furthermore, this model could be extended to assess any HFO detection algorithm against “universal” HFO evaluations, or to establish the required number of events and evaluators for semi-supervised HFO identification.

Table 1 – Generalizability Theory Variance Estimates

Facet	Absolute Variance	Relative Variance
p	3.202	10.60%
t	17.602	58.29%
d	0.894	2.96%
d×t	0	0.00%
p×t	0.428	1.42%
p×d	0.314	1.04%
p×d×t	0.259	0.86%
e:(d×t)	2.082	6.89%
p×e:(d×t)	2.709	8.97%
error	2.706	8.96%
Total	**30.196**	**100.00%**

Absolute and relative variances corresponding to the facets of the nested generalizability theory model p×e:(d×t), where p = evaluator, e = event, t = event type (candidate HFO, low-threshold HFO, or distractor), d = dataset, × denotes a crossed relationship between facets, and : denotes a nested relationship.