

Interrater reliability in interpretation of electrocorticographic seizure detections of the responsive neurostimulator

*Mark Quigg, §Felice Sun, *Nathan B. Fountain, †Barbara C. Jobst, ‡Victoria S. S. Wong, §Emily Mirro, ‡Sarah Brown, and ‡David C. Spencer

Epilepsia, 56(6):968–971, 2015

doi: 10.1111/epi.12998

SUMMARY

Objective: Electrocorticographic (ECoG) recordings from patients with medically intractable partial-onset seizures treated with a responsive neurostimulator system (the RNS System) that detects and stores physician-specified ECoG events provide a new data resource. Interpretation of these recordings has not yet been validated. The purpose was to evaluate the interrater interpretation of chronic ambulatory ECoG recordings obtained by the RNS System.

Methods: Five pairs of five experts independently classified 7,221 ECoG recordings obtained from 128 patients with medically intractable partial seizures who participated in a randomized controlled trial of the safety and efficacy of the RNS System. ECoG detections—“long episodes” or “saturations”—were classified as “seizures” or “not seizures” based on a reference definition. Interrater agreement rates and kappa score reliabilities were calculated between rater pairs from the ECoG sample as a whole and within individual patients who had more than the median number of individual ECoG recordings.

Results: The overall interrater agreement was 79%, with a reliability $\kappa = 0.57$ (moderate agreement). Agreement between pairs of reviewers ranged from 0.69 to 0.85. Agreement rates were 94% or better for 50% of patients. Only 25% of patients had ECoG recordings agreement rates worse than 75%. ECoGs with mixed interpretations (one reviewer “seizure”/the other—“not seizure”) consisted of periods of low amplitude activity that evolved in amplitude or periodic discharges near 2 Hz.

Significance: Although reliability as a whole was moderate, for the majority of patients, detections yielded highly reliably interpreted events of either electrographic seizures or nonictal epileptiform activity.

KEY WORDS: Electrocorticography, Focal epilepsy, EEG, Brain electrical stimulation, Seizure detection.



Mark Quigg is a professor in the Department of Neurology and director of EEG/EP at the University of Virginia in Charlottesville, Virginia.

Accepted March 15, 2015; Early View publication April 20, 2015.

*Department of Neurology, University of Virginia, Charlottesville, Virginia, U.S.A.; †Department of Neurology, Dartmouth University, Lebanon, New Hampshire, U.S.A.; ‡Department of Neurology, Oregon Health and Science University, Portland, Oregon, U.S.A.; and §NeuroPace Inc, Mountain View, California, U.S.A.

Address correspondence to Mark Quigg, Department of Neurology, University of Virginia, Box 800394, Charlottesville, VA 22902, U.S.A. E-mail: quigg@virginia.edu

© 2015 The Authors. *Epilepsia* published by Wiley Periodicals on behalf of International League Against Epilepsy.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

KEY POINTS

- The interrater agreement rate in interpretation of seizures in ECoG recordings of the RNS System was 79% with a reliability of 0.57 (moderate).
- More than 50% of patients had ECoG recordings with interrater agreement rates in interpretation of seizures $\geq 94\%$.
- Twenty-five percent of patients accounted for most of the disputed interpretations.
- ECoG recordings with low interrater agreement consisted of either low amplitude rhythmic fast activity or quasi-periodic discharges.

The responsive neurostimulator system (RNS System; NeuroPace Inc., Mountain View, CA, U.S.A.) provides an opportunity not only to treat medically intractable partial seizures but also to evaluate the characteristics of long-term ambulatory electrocorticography (ECoG). The system includes a cranially implanted neurostimulator connected to one or two depth or cortical strip leads that are implanted near one or two seizure foci. The RNS System detects specific patterns of ECoG activity, provides closed-loop electrical stimulation at a seizure focus, and stores ECoG activity, all based on physician-programmed parameters.

To assess long-term ambulatory ECoG data, a first step is to determine the interrater reliability in identifying ECoG seizures. Despite the long history of ECoG,¹ few have evaluated the reliability in identifying potential electrographic seizures recorded from intracranial electrodes. None have evaluated ECoG recordings obtained from chronic ambulatory devices. Osorio et al.² evaluated 34 seizures from 10 patients recorded from traditional short-term intracranial electrodes and reported an interrater agreement (defined as the agreement of two of three reviewers) at a rate of 93% and a kappa score of 0.68.

The purpose of this study was to quantify the interrater reliability in the classification of events recorded during ambulatory ECoG.

METHODS

This retrospective evaluation obtained during the randomized controlled trial of the RNS System comprised 191 adults with medically intractable focal epilepsy^{3,4}.

The neurostimulator stores samples of ECoG recordings based on physician-defined criteria. A subset has characteristics suggestive of electrographic seizures: (1) a *long-episode*, an event for which detection of epileptiform activity continued beyond a preset physician-defined duration; and (2) a *saturation*, an event with amplitudes exceeding amplifier sensitivity. Approximately four 90 s ECoG samples

were stored at any one time, after which new ECoG recordings overwrote the oldest. To free neurostimulator memory, patients were instructed to transfer data daily and after a clinical seizure from the neurostimulator to an external remote monitor. Thus, the numbers of ECoG recordings from each patient varied with detection parameters, the nature of the EEG pattern, and the frequency with which patients transferred the data.

Subjects included had a continuous period ≥ 84 days beginning at least 18 months after implant in which settings were optimized and held constant. The ECoG recordings reviewed were the first 100 most recent long episodes and saturations (for a maximum total of 200 events) stored during the 84-day sampling period. ECoG recordings consisted of four bipolar channels recorded at 250 Hz, a band pass of 4–90 Hz, and a standard review sensitivity of 0.8 mV peak–peak.

Qualified reviewers interpreted ECoG studies while blinded to patient, neurostimulator settings and leads, and each other's interpretations. We defined an "electrographic seizure" as a sustained rhythmic discharge, including repetitive spiking or spike-and-wave pattern faster than or equal to 2 Hz, with definite evolution in frequency, location, or morphology, and clearly distinguishable from background, lasting at least 10 s in duration.⁵ "Nonseizure" samples consisted of epileptiform activity or nondescript patterns.

An agreement rate and Fleiss' kappa statistic⁶ were calculated for the all and for individual pairs of reviewers. Because particular patients with "difficult" ECoG patterns could account for the bulk of disagreements, intrapatient agreement rates were calculated for patients with greater than the median number of ECoG recordings reviewed (representing those with sufficient samples from which to calculate intrapatient agreement).

ECoG recordings were then grouped by interrater results into (1) those that both reviewers identified as electrographic seizures; (2) those that both identified as nonseizures; and (3) those that one reviewer ranked as electrographic seizure and the other as nonseizure. To determine which patterns fell into low and high reliability, 10 ECoG studies from each group were randomly selected for further description.

RESULTS

From the 191 subjects enrolled in the trial, 128 met inclusion criteria. Five reviewers forming five pairs reviewed a total of 7,221 ECoG recordings (mean ECoG recordings per patient 56.4, median 52.5, and range 1–160). The overall agreement rate (both reviewers of each pair agreed "seizure" or "not seizure") was 79%, for a kappa statistic of 0.57 (moderate agreement).

Agreement rates between pairs of reviewers ranged from 0.69 to 0.85 (Table 1). The range of kappa statistics between pairs ranged from 0.38 (fair agreement) to 0.70

Table 1. Interrater agreement rates and reliability among pairs of reviewers in interpretation of ECoG detections as “seizures” or “not seizures” recorded in the RNS System.

Reviewer pair	ECoG studies reviewed	Interrater agreement (%)	κ
D.S./V.W.	1,277	0.85	0.70
N.F./M.Q.	1,600	0.82	0.64
D.S./B.J.	1,403	0.81	0.60
V.W./M.Q.	1,222	0.78	0.55
N.F./B.J.	1,719	0.69	0.38

(substantial agreement). Reviewers within institutions (D.S./V.W. and M.Q./N.F.) had better agreement rates and reliability than those pairs across institutions (V.W./M.Q., N.F./B.J., and D.S./B.J.).

A minority of patients accounted for those ECoG studies with low agreement rates. Figure 1 shows the distribution of agreement rates of ECoG recordings within individual patients. The median inpatient agreement rate was 94%, meaning that 50% of patients had interrater agreement rates of 94% or better. Therefore, most patients generally had unambiguous detections that could reliably be categorized as electrographic seizure or as nonictal epileptiform activity. Furthermore, the lowest quartile of patients accounted for those ECoG studies with agreement rates <75%, indicating that a minority of patients accounted for the most ambiguities in interpretation.

Based on review of 10 ECoG samples from each group, a representative unambiguous ECoG is shown in Figure 2A and consists of runs of rhythmic sharp activity with changing frequency and amplitude as befits the characteristic of spatial or temporal evolution.

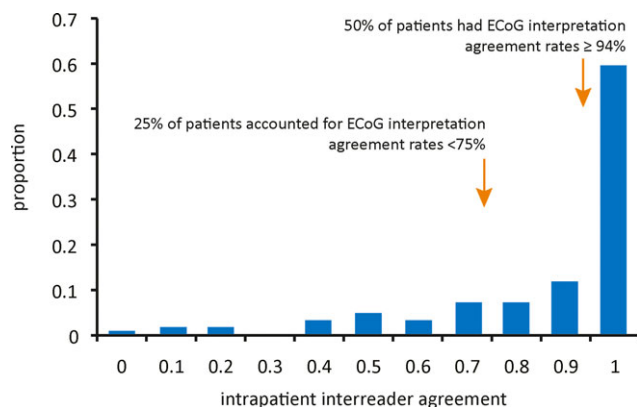


Figure 1.

Distribution of interrater agreement rates within patients from interpretations of long episodes and saturations as detected by the RNS System. Patients represented those whose inpatient ECoG samples were above the median reviewed.

Epilepsia © ILAE

Although “nonseizures” did not qualify as electrographic seizures per the definition, all contained abnormal epileptiform activity (Fig. 2B).

Those with the high rates of disagreement (Fig. 2C) fell into three categories: (1) low amplitude rhythmic activity with limited spatial or temporal evolution until the end of the discharge; (2) discharges with clear evolving activity that were <10 s because ictal onsets—and thus, the overall duration—were ambiguous (some overlap with the first category was present); and (3) quasi-periodic discharges during which runs became more organized with frequencies near 2 Hz.

DISCUSSION

Experts classified ECoG seizure detections (long episodes or saturations) stored by the neurostimulator with moderate interrater reliability. The majority of detections were electrographic seizures, but those detections classified as “nonseizures” usually contained abnormal, epileptiform activities. The bottom quartile of patients accounted for interrater agreement rates worse than 75%, meaning that detections from the majority of patients yielded highly reliably interpreted events.

This is the first study of the interrater reliability of seizure detections obtained from long-term ambulatory ECoG, and one of the few studies of interrater reliability of ECoG seizure interpretation. Disagreements may arise from the definition of seizures and follow those ambiguities well established in interpretation of ictal versus nonictal epileptiform activity.^{7,8} One of the strengths of this study—the total number of events exceeded 7,000—may also be a weakness, as the on-line system facilitated rapid review of many recordings. A more traditional print-based review of far fewer recordings may have yielded higher agreement rates. The limited spatial acquisition (four channels maximum) may also affect reliability when compared to the ≥ 128 channels of intracranial EEG monitoring.

ECoG studies with poor interrater agreement—mainly periodic discharges or low amplitude fast activities—are those that are difficult to classify in other areas of EEG interpretation. For example, periodic discharges fall into the hole of clinical correlation noted in algorithms of possible nonconvulsive status epilepticus.⁷ Although periodic discharges as seizures remain controversial, their appearance in ambulatory ECoG samples supports the concept of periodic discharges as a marker of the epileptic lesion.⁸ Low-amplitude fast activities also were less reliable because establishing evolution or duration was ambiguous.

The RNS System is a therapeutic rather than a diagnostic device. Although continuously monitoring, the RNS System does not continuously record. One can review only what was detected and stored. In practice, the neurostimulator may make hundreds of detections a day—based on patterns that precede that patient’s seizures—and provide brief (usu-

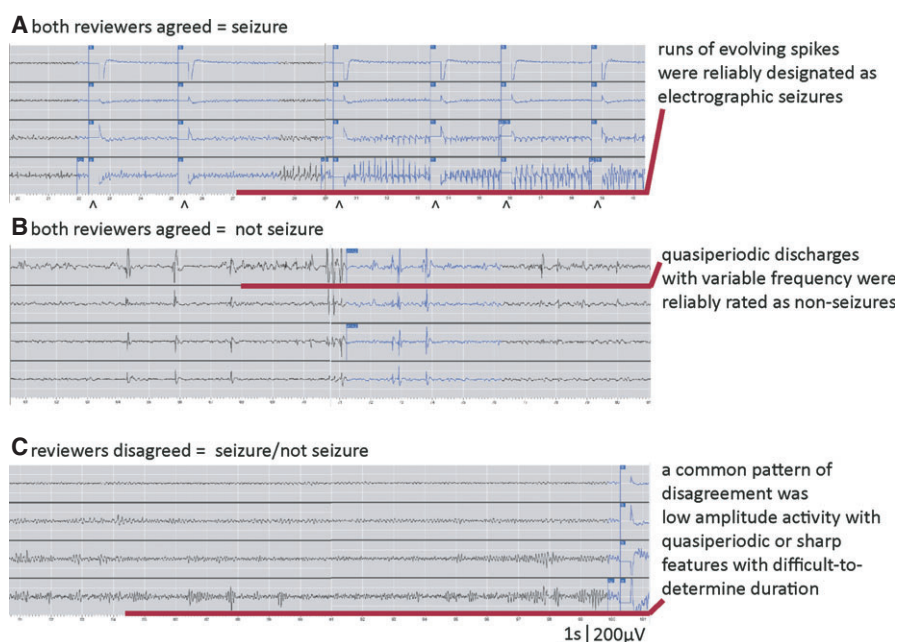


Figure 2.

Representative ECoG recordings demonstrating (A) interrater agreement in assignment of electrographic seizure with clearly evolving rhythmic activity; (B) interrater agreement designating nonseizure activity consisting of variable runs of slowly occurring periodic discharges; and (C) interrater disagreement with one reviewer assigning seizure and the other nonseizure in lower amplitude activities with quasi-periodic bursts with difficult-to-determine durations or evolution.

Epilepsia © ILAE

ally 100 msec) bursts of stimulation, amounting to <6 min per day. Those that are preferentially stored are episodes suggestive of an electrographic seizure. To assess whether these ECoG episodes can be correlated with the patient's clinical seizure frequency, it will be necessary to define the clinical significance of each of these types of detections.

Physicians in the RNS System trial³ strived to set parameters to achieve detection of typical seizures. The finding that a minority of patients accounted for agreement rates <75% meant that programming parameters allowed recording of activity that could be reliably interpreted in the majority. In other words, most events stored by the RNS System occupy clear ends of the interictal-ictal spectrum on physician interpretation. Subsequent studies using the recording capabilities of the neurostimulator—for example, evaluations of detection frequency, temporal distribution, or location—can be undertaken with greater confidence. Further work will be needed to reliably use ECoG studies recorded by the RNS System as surrogate markers for electrographic seizures.

ACKNOWLEDGMENTS

We thank Martha Morrell, MD, for her editorial suggestions on discussions of capabilities of the RNS system. Mark Quigg receives research support from the National Institutes of Health/National Institute of Neurological Disorders and Stroke (NIH/NINDS) and from Barron Associates.

DISCLOSURE OF CONFLICT OF INTEREST

Felice Sun and Emily Mirro are employees of NeuroPace. Nathan B. Fountain, Mark Quigg, and David C. Spencer participated in clinical trials of the RNS System. The remaining authors have no conflicts of interest. 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.

REFERENCES

1. Penfield W, Jasper HH. *Epilepsy and the functional anatomy of the human brain*. London: Churchill; 1954.
2. Osorio I, Frei MG, Giftakis J, et al. Performance reassessment of a real-time seizure-detection algorithm on long ECoG series. *Epilepsia* 2002;43:1522–1535.
3. Morrell MJ. Responsive cortical stimulation for the treatment of medically intractable partial epilepsy. *Neurology* 2012;77:1295–1304.
4. Heck CN, King-Stephens D, Massey AD, et al. Two-year seizure reduction in adults with medically intractable partial onset epilepsy treated with responsive neurostimulation: Final results of the RNS System Pivotal trial. *Epilepsia* 2014;55:432–441.
5. Durazzo TS, Spencer SS, Duckrow RB, et al. Temporal distributions of seizure occurrence from various epileptogenic regions. *Neurology* 2008;70:1265–1271.
6. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378–382.
7. Beniczky S, Hirsch LJ, Kaplan PW, et al. Unified EEG terminology and criteria for nonconvulsive status epilepticus. *Epilepsia* 2013;54(Suppl. 6):28–29.
8. Yoo J, Rampal N, Petroff OA, et al. Brief potentially ictal rhythmic discharges in critically ill adults. *JAMA Neurol* 2014;71:454–462.