

JUDGING A BOOK BY ITS COVER: SEIZURES VERSUS ITS MIMICKERS

Interrater Reliability of EEG-Video Monitoring. Benbadis SR, LaFrance WC Jr, Papandonatos GD, Korabathina K, Lin K, Kraemer HC, NES Treatment Workshop. *Neurology* 2009;73(11):843–846. **OBJECTIVE:** The diagnosis of psychogenic nonepileptic seizures (PNES) can be challenging. In the absence of a gold standard to verify the reliability of the diagnosis by EEG-video, we sought to assess the interrater reliability of the diagnosis using EEG-video recordings. **METHODS:** Patient samples consisted of 22 unselected consecutive patients who underwent EEG-video monitoring and had at least an episode recorded. Other test results and histories were not provided because the goal was to assess the reliability of the EEG-video. Data were sent to 22 reviewers, who were board-certified neurologists and practicing epileptologists at epilepsy centers. Choices were: 1) PNES, 2) epilepsy, and 3) nonepileptic but not psychogenic (“physiologic”) events. Interrater agreement was measured using a κ coefficient for each diagnostic category. We used generalized κ coefficients, which measure the overall level of between-method agreement beyond that which can be ascribed to chance. We also report category-specific κ values. **RESULTS:** For the diagnosis of PNES, there was moderate agreement ($\kappa = 0.57$, 95% confidence interval [CI] 0.39–0.76). For the diagnosis of epilepsy, there was substantial agreement ($\kappa = 0.69$, 95% CI 0.51–0.86). For physiologic nonepileptic episodes, the agreement was low ($\kappa = 0.09$, 95% CI 0.02–0.27). The overall κ statistic across all three diagnostic categories was moderate at 0.56 (95% CI 0.41–0.73). **CONCLUSIONS:** Interrater reliability for the diagnosis of PNES by EEG-video monitoring was only moderate. Although this may be related to limitations of the study (diagnosis based on EEG-video alone, artificial nature of the forced choice paradigm, single episode), it highlights the difficulties and subjective components inherent to this diagnosis.

COMMENTARY

Psychogenic nonepileptic seizures (PNES), also termed pseudo-seizures, are defined as paroxysmal changes in behavior that resemble epileptic seizures but are without organic cause and are not accompanied by the ictal, peri-ictal, and interictal changes that characterize epilepsy. PNES is often suspected in patients with a history of somatization, abuse, psychiatric comorbidity, and when spells are characterized by unusual features, such as emotional triggers, prolonged duration, stop/start quality, or pelvic thrusting. However, the diagnosis of PNES can be difficult to make based on clinical criteria alone. The standard of care for diagnosing PNES remains the recording of a patient’s typical event with video-EEG to confirm the absence of electrographic changes on the ictal EEG recording during the spell, while simultaneously evaluating semiology for either PNES behaviors or a lack of ictal characteristics.

The incidence of PNES is reported to be 1.5 per 100,000 per year that is approximately 4% of the incidence of all epilepsy (1). Yet, PNES accounts for between 25% and 45% of all video-EEG admissions to epilepsy monitoring units (2). In this Benbadis et al. investigation, 22 reviewers analyzed 22 consecutive single events recorded with video-EEG for the purpose of assessing interrater reliability. The reviewers were blinded to a detailed clinical history and other events documented on the patient (3). Although there was substantial concordance for the diagnosis of epilepsy, overall there was actually “low agreement for physiologic nonepileptic events and moderate agreement for

psychogenic nonepileptic events.” This study helps to underscore the inherent difficulties associated with the diagnosis of PNES in video-EEG monitoring units based on the evaluation of a single event in isolation. Perhaps, the interrater reliability would have been improved considerably with assessment of multiple events.

Because epilepsy is often diagnosed by a change in the EEG, it was no surprise to find substantial concurrence for an epilepsy diagnosis. There should have been even high interrater reliability, given that there were 22 raters, and if all of the EEG samples had had an unequivocal, easily interpreted change. However, scenarios in which the significance of a video-EEG change may not be readily apparent make correctly diagnosing epilepsy versus PNES difficult. In particular, it can be difficult to discriminate between PNES and simple partial seizures, because only approximately 20% of patients have an EEG accompaniment, and complex partial seizures arise from deep convexities within the frontal lobe, making them difficult to record on scalp electrodes (4).

One issue that arises from the study pertains to the identification of the semiologic characteristic of the EEG event that led to each of the reviewers’ diagnosis. Recently, Hoerth and colleagues systematically analyzed clinical signs that may be useful for discriminating between the diagnosis of epilepsy and PNES in video monitoring studies (5). The isolated behaviors they identified included pelvic thrusting, ictal stuttering, ictal eye closure, and the teddy bear sign. Table 1 lists the diagnostic accuracy of these clinical signs. Ictal stuttering and teddy bear sign were often the most significant predictors of psychogenic nonepileptic events. Among the EEG reviewers, those who had a moderate level of interrater reliability for the psychogenic

TABLE 1. *Diagnostic Accuracy of Clinical Signs during Spells for Distinguishing PNES from Epileptic Seizures*

| CLINICAL SIGN | SENSITIVITY | SPECIFICITY |
|------------------|--------------|--------------|
| Pelvic thrusting | 78% (59–97) | 76% (64–88) |
| Ictal stutter | 9% (3–14) | 100% |
| Eye closure | 64% (56–71) | 74% (69–79) |
| Teddy bear sign | 5.2% (3–7.5) | 99% (99–100) |

Table 1 adapted from Hoerth (5).

nonepileptic events, perhaps, represent a bias towards analyzing specific semiological component of the clinical signs to inform a particular diagnosis. In contrast, other reviewers may have primarily relied on the clinical gestalt, when other variables such as medical history and heterogeneity of multiple spells, were not available to factor into the diagnosis.

Moreover, the low interrater reliability for physiologic nonepileptic events is not surprising, given the authors' explanation that these events were an infrequent diagnostic category and more importantly, that the physiologic event could have represented a number of other neurologic conditions. Perhaps, a neurologist or other neurological subspecialist, who is accustomed to diagnosing paroxysmal events and has had a significant amount of exposure to diagnosing numerous spells types, could better identify episodes of physiological nonepileptic events that mimic epilepsy (e.g., movement disorders) than could this group of "board-certified neurologists and practicing epileptologists at epilepsy centers."

In summary, the study by Benbadis et al. is a reminder that there is no one clinical factor, whether video-EEG, history, or an examination, that is going to 100% accurately identify a patient with a given diagnosis. The authors were quite correct

in stating that the diagnosis of PNES represents the art of medicine, as different lines of converging evidence need to be assimilated in order to arrive at its correct diagnosis. This study also underscores the fact that video monitoring of a single event cannot be utilized in isolation, without also evaluating the clinical history and other records. Doing so, is akin to judging a book by its cover and not by reading what is inside.

Clearly much more work is needed in order to better standardize video-EEG interpretation and perhaps find better biomarkers to confirm a diagnosis of PNES with high specificity and sensitivity. Benbadis and colleagues have done excellent work in advancing the knowledge of interrater reliability for interpretation of individual nonepileptic events. Consistent use of video-EEG monitoring, along with itemizing the clinical signs that help assess whether an event is epileptic or nonepileptic, may allow for the development of more sophisticated models that accurately represent the true complexity of these events.

by Joseph I. Sirven, MD

References

1. Krumholz A, Hopp J. Psychogenic seizures. *Semin Neurol* 2006;26:341–350.
2. Parra J, Iriarte J, Kanner AM. Are we overusing the diagnosis of psychogenic non-epileptic events? *Seizure* 1999;8:223–227.
3. Benbadis SR, LaFrance WC Jr, Papandonatos GD, Korabathina K, Lin K, Kraemer HC, for the NES Treatment Workshop. Interrater reliability of EEG-video monitoring. *Neurology* 2009;73:843–846.
4. Lieb JP, Walsh G, Babb T. A comparison of EEG seizure patterns recorded with surface and depth electrodes in patients with temporal lobe epilepsy. *Epilepsia* 1976;17:137–160.
5. Hoerth MT, Wellik K, Demaerschalk B, Draskowski J, Noe K, Sirven JI, Wingerchuk D. Clinical predictors of psychogenic nonepileptic seizures. *The Neurologist* 2008;14:266–270.