Published in final edited form as:
Epilepsia. 1990 ; 31(1): 110–115.

Semistructured Interview for Seizure Classification: Agreement with Physicians' Diagnoses

Ruth Ottman^{*,†}, W. Allen Hauser^{*,†,‡}, and Lillian Stallone^{*}

^{*} *G. H. Sergievsky Center, Columbia University, New York, NY, U.S.A*

[†] *Epidemiology Division, School of Public Health, Columbia University, New York, NY, U.S.A*

[‡] *Department of Neurology, Columbia University, New York, NY, U.S.A*

Summary

A semi-structured interview was developed for classification of seizures in accordance with the 1981 International League Against Epilepsy (ILAE) criteria. The interview was administered over the telephone by trained lay interviewers. Interview-based diagnoses for 50 patients were compared with independent diagnoses by neurologists who also use the ILAE system for seizure classification. Interview diagnoses agreed with those of physicians for broad seizure-type classifications (i.e., partial vs. generalized onset) in 88% of patients and for detailed combinations of seizure type in 64% of patients. Nonchance agreement between the two sources, assessed by the κ statistic, was excellent for any partial onset ($\kappa = 0.83$), secondarily generalized ($\kappa = 0.81$), and primary generalized tonic-clonic ($\kappa = 0.76$) seizures. Agreement was fair to good for any generalized onset ($\kappa = 0.70$), simple partial ($\kappa = 0.56$), complex partial ($\kappa = 0.54$), and generalized nonconvulsive ($\kappa = 0.56$) seizures. Sensitivity ranged from 0.60 to 1.0 for partial onset seizures, and from 0.43 to 0.67 for generalized onset seizures. Specificity ranged from 0.60 to 0.87 for partial onset seizures, and was 1.0 for generalized onset seizures. Positive predictive value was 0.95 for any partial onset and 1.0 for any generalized onset seizure. These results suggest that this interview can produce accurate diagnoses of major seizure categories. Use of this instrument for clinical epilepsy research should facilitate conduct of large studies at a significant saving of both time and money.

Keywords

Seizures; Epilepsy; Epidemiology

Accurate seizure classification is one of the most important elements of clinical epilepsy research, but development of procedures for classification is difficult. Until recently, no uniform criteria for seizure nomenclature and classification were available. In 1970, the International League Against Epilepsy (ILAE) first proposed a classification scheme (Gastaut, 1970), which was revised in 1981 (Commission, 1981). This system is an important advance for epilepsy research, but many physicians treating patients with epilepsy are not familiar with it. Furthermore, even physicians familiar with the ILAE system may disagree more often than desired when classification is based on seizure descriptions in medical records. In one recent study (Bodensteiner et al., 1988), reliability of diagnoses based on 1981 ILAE criteria was assessed by analysis of agreement among four neurologists who classified childrens' seizures based on descriptions in medical records. Depending on the observer, 22–51% of descriptions were considered unclassifiable. Inter-rater agreement, assessed by the κ statistic, was rather

Address correspondence and reprint requests to Dr. R. Ottman at Sergievsky Center, Columbia University, 100 Haven Ave., Tower III, Apt. 26G, New York, NY 10032, U.S.A.
Ms. Stallone's present address is Department of Psychiatry, Yale University School of Medicine, New Haven, CT, U.S.A.

low (0.24–0.38) when all descriptions were included and increased somewhat (0.39–0.51) after descriptions that lacked sufficient detail or were otherwise unclassifiable were excluded. When specific seizure types were considered and only detailed descriptions were included, κ ranged from 0.28 to 0.90 for the best observer pair. This low reliability (reproducibility) has implications for validity (accuracy) since reliability is necessary (though not sufficient) to ensure validity.

Problems in reliable and valid seizure classification raise serious questions about strategies for clinical epilepsy research; e.g., it might be anticipated that adequate reliability could be assured only if each patient were thoroughly evaluated by a single neurologist. On the other hand, if alternative methods could be developed for reliable and valid seizure classification, a significant savings in both time and money might result, especially in studies involving large numbers of patients.

To this end, we developed a semistructured interview for classification of seizures in accordance with the 1981 ILAE criteria. We describe the interview and present data on agreement between interview- and physician-based diagnoses.

MATERIALS AND METHODS

Format and content of interview

The Seizure Classification Interview (SCI) was developed as part of the Epilepsy Family Study Proband Interview Form (Ottman et al., 1987), an instrument used for our ongoing study of the genetics of epilepsy. It contains both structured questions answered in discrete categories (e.g., “yes,” “no,” or “don’t know”), and open-ended questions answered in the patient’s own words. For example, information about auras preceding grand mal seizures is obtained in two ways. The first (open-ended) question asks “In your own words, can you describe how you feel before a big seizure starts?” The second (structured) question asks “Do you have any warning (or aura) that a big seizure is about to happen?” If the answer to the second question is “yes,” a description of the types of auras usually experienced is requested again.

The SCI contains a separate set of questions about each of four categories of seizures: “big seizures or grand mal seizures,” “small seizures,” “sudden jerking of part or all of your body,” and episodes in which “part or all of your body suddenly goes limp, causing you to fall or drop things.” The section on grand mal seizures is designed to distinguish between primary and secondary generalization, and includes questions on age at onset, history of aura, unilateral onset of convulsions, postictal unilateral weakness and numbness, and postictal aphasia. A verbatim description of what happens before, during, and after the seizure is also requested. Loss of consciousness is confirmed by history of postictal confusion and drowsiness and inability to recall events during the seizure. The section on small seizures includes questions on age at onset, seizure duration, changes in awareness and ability to communicate with surroundings, focal motor activity, automatisms, eye-fluttering, postictal confusion, and postictal drowsiness. A verbatim description of what happens before, during, and after a small seizure is also requested. If the patient has had more than one type of small seizure, this section is repeated for each type. The section on episodes of sudden jerking is aimed at ascertainment of myoclonic seizures, and includes questions on age at onset, extent of body involvement (i.e., part or all), unilateral or bilateral involvement, restriction of episodes to the same side, and usual timing (i.e., only before sleep, only before a big seizure, only on awakening). The section on episodes of going limp is aimed at ascertainment of atonic seizures and includes questions on age at onset, loss of consciousness, and postictal confusion.

Interview administration

The interviews were administered over the telephone by trained lay interviewers. The training sessions lasted 2 days and included an introduction to the study objectives, a lecture on clinical and epidemiologic aspects of seizure disorders, viewing of a videotape of different seizure types, general guidelines for interviewer conduct (i.e., neutrality, recording responses verbatim, not deviating from printed wording of questions), specific instructions for administering our interview (i.e., circling patients' answers, following "skip" instructions), and practice interviews among the trainees. No other experience or training was required, although all interviewers had at least some college education and some had additional professional training, such as nursing or social work.

Seizure classification

Seizures were classified by a research assistant who was medically trained in internal medicine but not neurology. She was instructed to use a set of guidelines we developed with reference to specific questions in the interview, and to consult the study neurologist (W.A.H.) in ambiguous cases.

Patient population

The study population comprised 50 patients: 5 were referred by neurologists specializing in epilepsy at our institution and 45 ascertained from voluntary organizations for epilepsy. Patients were eligible only if they were aged ≥ 18 years, English-speaking, capable of answering the questions in the interview (i.e., not severely mentally retarded, hearing impaired, or demented), and had a positive history of epilepsy (i.e., a lifetime history of two or more unprovoked seizures). During the interview, patients were asked for permission to contact their physicians for review of medical records. Those who gave permission received consent forms to be signed and returned. The signed consent forms were then sent to the physicians with requests for medical records. Patients were included in this study if the medical record had been received and the responsible physician was known to be an epilepsy specialist who used the 1981 ILAE classification system. Medical records of the five patients referred by physicians at our institution were obtained directly from our hospital record room.

Review of medical records

The records were reviewed for the neurologists' diagnoses of seizure type, and these diagnoses were recorded. Seizure *descriptions* in the records were not used to modify the diagnoses.

Analysis

Agreement between interview- and physician-based diagnoses was examined both by seizure *type* (regardless of how many different types each patient had) and by patient (for various combinations of seizure types). Kappa statistics were used to measure agreement between interview and physician-based diagnoses, after excluding agreement expected by chance (Fleiss, 1981). Values of κ range from -1 to $+1$, with values below zero representing lower than chance agreement. Interpretation of values above zero is somewhat arbitrary, but one analysis suggested that values below 0.4 be interpreted as "poor," those from 0.4–0.75 as "fair to good," and those >0.75 as "excellent" (Landis and Koch, 1977).

Agreement with physicians was also assessed by calculating sensitivity, specificity, and positive predictive value, assuming that the physicians' diagnoses were correct (Fleiss, 1981). For each seizure type, sensitivity was taken as the proportion of patients classified by the interview as having that type, among those so classified by physicians; specificity was taken as the proportion of patients classified by the interview as not having that type, among those

so classified by physicians; and positive predictive value was taken as the proportion of patients diagnosed by physicians with that type, among those so classified by the interview.

The seizure types examined included the broad categories of generalized and partial seizures and the following specific categories: simple partial, complex partial, secondarily generalized, primary generalized tonic-clonic, absence, myoclonic, and atonic. κ , sensitivity, specificity, and positive predictive value were not calculated for seizure types with prevalence lower than 10% in our series (according to physicians), since the estimates would be very imprecise for such seizure types. Because absence, myoclonic, and atonic seizures each had very low prevalence, they were grouped under the heading “generalized nonconvulsive seizures” for these calculations. Patients classified as unknown for a given seizure type by either the physician or interview were excluded from calculation of κ , sensitivity, specificity, and positive predictive value for that type.

RESULTS

Table 1 shows the distribution of seizure types for the 50 patients in the study according to each diagnostic source. The number of patients diagnosed with each seizure type is shown separately, regardless of the number of other types each patient might have. Two patients were diagnosed by physicians as having both generalized onset and partial onset seizures. The interview was unable to distinguish between generalized and partial onset seizures in three patients, whereas physicians were unable to make this distinction in only one patient. According to both sources, most patients had complex partial seizures and secondarily generalized seizures.

Table 2 shows agreement between the interview-and physician-based diagnoses for broad categories of seizure type (i.e., generalized vs. partial onset) and specific combinations of seizure type within the broad categories. Perfect agreement for detailed seizure combinations was found in 64% of patients. Agreement was higher for patients with partial (69%) than generalized onset (50%) seizures. Most of the disagreements were for a single seizure type within the broad classifications. Hence, agreement with physicians for broad seizure classifications was 88% overall (partial 95%; generalized 75%). Two patients diagnosed as partial onset by physicians were unclassifiable on the basis of interviews. In two patients diagnosed by physicians as having both primary generalized and partial onset seizures, the interview failed to identify the generalized onset seizures as independent phenomena (Table 2). In one remaining patient, neither the interview nor the physician was able to make a distinction.

Table 3 shows κ , sensitivity, specificity, and positive predictive value for each seizure type, calculated after excluding cases with unknown status for that type. Nonchance agreement, assessed by κ , was excellent for any partial (0.83), secondarily generalized (0.81), and primary generalized tonic-clonic (0.76) seizures. Agreement was fair to good for any generalized (0.70), simple partial (0.56), complex partial (0.54), and generalized nonconvulsive (0.56) seizures.

For partial onset seizures, sensitivity ranged from 0.60 to 1.0 and specificity ranged from 0.60 to 0.96. For all partial onset seizure types except simple partial, sensitivity was higher than specificity. For generalized onset seizures, sensitivity ranged from 0.43 to 0.67 and specificity was 1.0. Thus, in contrast to partial onset seizures, specificity was higher than sensitivity for primary generalized seizures. Positive predictive value ranged from 0.60 to 0.95 for partial onset seizures and was 1.0 for generalized onset seizures.

DISCUSSION

This study shows that our semistructured interview produces excellent agreement with physicians for diagnosis of any partial onset, secondarily generalized, and primary generalized tonic-clonic seizures, and fair-to-good agreement for all of the remaining assessed types (any generalized, simple partial, complex partial, generalized nonconvulsive).

The percentages of agreement shown in Table 2 are partly attributable to chance: Because most patients had partial onset seizures (82% according to either physicians or interview, Table 1), even a random allocation of patients to the partial category would produce a high level of agreement. Kappa statistics provide a more meaningful measure because they estimate agreement beyond chance.

Patients classified as “unknown” for a specific seizure type by either the interview or physician were excluded from calculation of κ , sensitivity, specificity, and positive predictive value. This seemed more appropriate than considering such patients as *unaffected* with that seizure type since it was known beforehand that the information was inadequate to make a diagnosis. In such cases, the two sources would need to be used together to make a diagnosis, and agreement between them is not very meaningful.

Agreement with physicians may be viewed as being either reliability (reproducibility) or validity (accuracy), depending on the degree of confidence placed in the physicians’ diagnoses. Physicians have obvious advantages over the interview: They are highly trained and experienced and have the opportunity to supplement the clinical history with laboratory data, neurologic examinations and, in some instances, direct observation of the patient’s seizures. On the other hand, use of a standardized interview for seizure diagnosis has advantages over use of physician diagnoses. The interview is administered uniformly to all patients and addresses aspects of the seizure history that might not be covered in all cases even by highly trained physicians.

When disagreements were observed, they were usually for a single seizure type in a patient whose broad classification and other seizure types were in agreement. In two cases of disagreement, the interview diagnosis may actually be correct. In one case, a clear description of simple partial seizures was given together with that of complex partial seizures, and the physician did not mention the former. In the other, the patient described complex partial seizures together with secondarily generalized seizures, and the physician did not mention the complex partial seizures. (Both instances may reflect omissions from the written record rather than true disagreements.) Two patients diagnosed by physicians with combinations of partial and generalized onset seizures were probably misclassified by the interview, but such diagnoses are difficult even for experienced clinicians. One of these patients had been evaluated at our institution, and the identification of both partial and generalized onset seizures was possible only through video monitoring.

Sensitivity, specificity, and positive predictive value reflect accuracy of the interview-based diagnoses, assuming that the physicians’ diagnoses are correct. Sensitivity was very high (above 90%) for complex partial and secondarily generalized seizures; there were few false negatives for either of these seizure types and they were seldom missed by the interview. Sensitivity was lower (60%) for simple partial seizures, which were missed in two of the five cases in which they were diagnosed by physicians. Overall, sensitivity was perfect for partial onset seizures: No case classified as such by a physician was misdiagnosed by the interview (although in two cases a diagnosis could not be made on the interview). Specificity for partial seizures was lower than sensitivity, since there were some false positives. For generalized seizures, the pattern of sensitivity and specificity was reversed. Specificity was perfect; there

were no false positives. However, sensitivity was only 60% overall because 4 of the 10 generalized cases were misdiagnosed as partial.

These findings suggest that the interview is more likely to misclassify a generalized patient as partial than to make an error in the opposite direction. However, positive predictive value, or the proportion of patients diagnosed by the interview with a given seizure type for whom that diagnosis was correct, was very high for both partial (0.95) and generalized onset (1.0) seizures (Table 3).

The higher level of agreement in our study than in that of Bodensteiner et al. (1988) is probably explained by the different methods used to obtain descriptions of the seizures. In the earlier study, descriptions were obtained from medical records that had been prepared by physicians with varying levels of training and expertise in epilepsy, and features critical for classification were omitted in many cases. In our study, critical features were unlikely to be missed either by the interview or the physicians, since the interview was specifically designed to cover them, and the physicians were epilepsy specialists familiar with the ILAE system. The two studies also differ in two other respects that might influence reliability. In the Bodensteiner et al. (1988) study, classifications by both observers were based on the same seizure descriptions, whereas in our study each classification was based on a different data source. In addition, the Bodensteiner et al. (1988) study involved classification of seizures in children, whereas ours involved classification of seizures in adults.

The levels of agreement observed in this study also compare favorably with levels observed in studies of other disorders. For psychiatric disorders, for example, κ ranged from 0.33 to 0.77 when normal clinical assessments were used (Spitzer and Fleiss, 1974), and from 0.29 to 0.82 when a structured interview was used (Helzer et al., 1977). Similarly, for diagnosis of clinical features of Hodgkin's disease, κ ranged from 0.29 to 0.59 with normal clinical assessments, and from 0.75 to 0.93 when specific criteria were used (Boyd et al., 1979).

The optimal approach for seizure diagnosis would take advantage of both sources, a structured interview for the clinical history supplemented by laboratory data and other findings from the medical record. With respect to our interview, supplementation with medical record data is especially important for identification of generalized nonconvulsive seizures, which tended to be misclassified as simple or complex partial seizures, thus reducing κ for both partial and generalized nonconvulsive seizures. This misclassification may reflect a lack of detail in the interview sections aimed at identifying these seizure types, whereas physicians can probe as much as necessary when they are suspected. Thus, in studies using the SCI for seizure diagnosis, medical records should be reviewed for evidence of absence, myoclonic, or atonic seizures. EEG reports, especially for recorded seizures, should be particularly useful in this regard.

In summary, our semistructured interview, designed for administration over the telephone by trained lay interviewers, produces accurate diagnoses of major seizure categories. Use of this instrument should facilitate conduct of large studies at relatively lower cost and shorter duration than has been possible in the past without compromising validity of seizure diagnoses.

Acknowledgements

We thank Dr. Bruce Link for comments on the manuscript, Dr. E. Jagielska-Wilk for assistance with seizure diagnoses, and Rose Mary Mahoney for clerical assistance. This study was supported by Grants No. NS20656 and NS22050 from the National Institutes of Health.

References

- Bodensteiner JB, Brownsworth RD, Knapik JR, Kanter MC, Cowan LD, Leviton A. Interobserver variability in the ILAE classification of seizures in childhood. *Epilepsia* 1988;29:123–8. [PubMed: 3258236]
- Boyd NF, Pater JL, Ginsburg AD, Myers RE. Observer variation in the classification of information from medical records. *J Chron Dis* 1979;32:327–32.
- Commission on Classification and Terminology of the International League Against Epilepsy. Proposal for revised clinical and electroencephalographic classification of epileptic seizures. *Epilepsia* 1981;22:489–501. [PubMed: 6790275]
- Fleiss, JL. Statistical methods for rates and proportions. 2. New York: John Wiley; 1981.
- Gastaut H. Clinical and electroencephalographic classification of the epilepsies. *Epilepsia* 1970;11:102–13. [PubMed: 5268244]
- Helzer JE, Clayton PJ, Pambakian R, Reich T, Woodruff RA Jr, Reveley MA. Reliability of psychiatric diagnosis II. The test/retest reliability of diagnostic classification. *Arch Gen Psychiatry* 1977;34:136–41. [PubMed: 843173]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74. [PubMed: 843571]
- Ottman, R.; Hauser, WA.; Stallone, L. Epilepsy Family Study Proband Interview Form. 1987. obtainable from the authors at a small fee for duplication
- Spitzer RL, Fleiss JL. A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry* 1974;125:341–7. [PubMed: 4425771]

TABLE 1
Distribution of seizure types for 50 patients studied

Seizure type	No. of Patients	
	Physicians	Interview
Partial onset		
Simple	5	5
Complex	39	37
Secondarily generalized	34	33
Any partial	41 (82%) ^a	41 (82%)
Generalized onset		
Tonic-clonic	9	6
Generalized nonconvulsive (absence, myoclonic, or atonic)	7	3
Absence	4	2
Myoclonic	2	1
Atonic	2	0
Any generalized	10 (20%) ^a	6 (12%)
Unknown onset	1 (2%)	3 (6%)

^aIncludes two patients with both generalized and partial onset seizures.

TABLE 2
Agreement between physician- and interview-based seizure type diagnoses

Physician's diagnosis	No. of patients			Interview diagnoses for patients who differed
	Interview in agreement			
	Total	Broad diagnosis ^a	Detailed diagnosis	
Partial onset				
CP only	6	6	5	SP, CP
SG only	1	1	0	CP, SG
SP, CP	1	1	1	—
SP, SG	1	1	0	CP, SG
SP, CP, SG	2	2	0	CP, SG (n = 1); SP, SG, CP? (n = 1)
CP, SG	28	26	21	SG only (n = 3) CP only (n = 1) SP, CP, SG (n = 1) Partial or generalized onset? (n = 2)
Total	39	37 (95%)	27 (69%)	
Generalized onset				
GTC only	3	3	2	GTC, Myo
GTC, Abs	3	1	1	CP, SG (n = 2)
GTC, Myo	1	1	1	—
GTC, Ato	1	1	0	GTC, Abs
Total	8	6 (75%)	4(50%)	
Both partial and generalized onset				
CP, SG, GTC, Abs, Ato	1	0	0	CP, SG
SP, CP, SG, Myo	1	0	0	SP, CP, SG
Uncertain	1	1	1	
Total	50	44 (88%)	32 (64%)	

CP, complex partial; SP, simple partial; SG, secondarily generalized; GTC, primary generalized tonic-clonic; Abs, absence; Myo, myoclonic; Ato, atonic.

^aBroad diagnosis, distinction between primary generalized and partial onset seizures.

TABLE 3
Kappa, sensitivity, specificity, and positive predictive value for specific seizure types

Seizure type	κ	Sensitivity	Specificity	Positive predictive value
Partial onset				
Simple	0.56	0.60	0.96	0.89
Complex	0.54	0.92	0.60	0.60
Secondarily generalized	0.81	0.94	0.87	0.94
Any partial	0.83	1.0	0.75	0.95
Generalized onset				
Tonic-clonic	0.76	0.67	1.0	1.0
Generalized nonconvulsive (absence, myoclonic, or atonic)	0.56	0.43	1.0	1.0
Any generalized	0.70	0.60	1.0	1.0