Published in final edited form as:

Epilepsy Behav. 2013 April ; 27(1): 148–153. doi:10.1016/j.yebeh.2012.12.030.

Inter-rater Agreement for Diagnoses of Epilepsy in Pregnant Women

Shahram Khoshbin¹, Amy Herring², Gregory L. Holmes³, Donald Schomer⁴, Daniel Hoch⁵, Elizabeth C. Dooling⁶, Eileen P. G. Vining⁷, and Lewis B. Holmes⁸

¹Department of Neurology, Brigham and Women's Hospital

²Department of Biostatistics, Harvard School of Public Health (currently, Department of Biostatistics, University of North Carolina)

³Department of Neurology, Children's Hospital (currently the Department of Neurology, The Geisel School of Medicine at Dartmouth)

⁴Department of Neurology, Beth Israel Deaconess Medical Center

⁵Epilepsy Center, Massachusetts General Hospital

⁶Department of Neurology, Division of Pediatric Neurology, Massachusetts General Hospital

⁷John M. Freeman Pediatric Epilepsy Center, The Johns Hopkins Medical Institutions

⁸Medical Genetics Unit, MassGeneral Hospital for Children, Boston, MA

Abstract

We report on inter-rater agreement in assessing the types of seizures exhibited by one hundred mothers ascertained in a study of the teratogenicity of maternal epilepsy and antiepileptic drugs. A summary of each woman's medical record, and a one-page report of her responses to questions about her epilepsy, were reviewed independently by six neurologists, three in pediatric neurology and three in adult neurology. Agreement was measured by the kappa statistic and log-linear modeling techniques. The adult neurologists agreed with each other 59 percent of the time, with agreement higher when all three used information from the patients' records, such as an EEG, rather than when depending on the woman's responses to questions about her epilepsy. Pediatric neurologists agreed with each other 44% of the time and tended to rely more heavily, than adult neurologists, on information in the patients' records, such as an EEG or a prior diagnosis.

Keywords

types of seizures; adult and pediatric neurologists; log linear modeling

© 2012 Elsevier Inc. All rights reserved.

Corresponding author: Lewis B. Holmes, M.D., Medical Genetics Unit, MassGeneral Hospital for Children, 175 Cambridge Street, Boston, MA 02114, Tel: 617-726-1742, Fax: 617-724-1911, holmes.lewis@mgh.harvard.edu.

Dedicated to the late Elizabeth A. Harvey, M.P.H., Ph.D., who made the original observation with S.K. and organized the analysis presented in this manuscript.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

Studies of inter-rater agreement or inter-examiner reproducibility help to identify those elements used most often in decision-making. In the study of the teratogenicity of maternal epilepsy, potential confounding factors identified have included: the type of epilepsy in the mother, the type and frequency of seizures during pregnancy, the antiepileptic drug(s) used, including dose and maternal blood level, socioeconomic status and other exposures, such as alcohol and other medications.

Previous studies of inter-rater agreement for the type of epilepsy (1–4) have focused on: 1) the clinical interpretation of EEGs [1]; 2) the value of a semi-structured interview of the patient with epilepsy [2]; 3) the degree of agreement on seizure types diagnosed from verbatim descriptions of seizure manifestations [3]; 4) the comparison of the degree of agreement by neurologists in evaluating the same information separately [4, 5].

A variety of statistical methods have been used to describe agreement between evaluators. Cohen's kappa statistic (6) has been the most widely-used of these methods. Assuming that the assessments of two raters are independent, kappa is a measure of the degree of agreement that is present beyond that expected by chance alone. However, the kappa statistic has many limitations in both its applicability and interpretations (7). One alternative to the kappa statistic is to use the log-linear modeling approach (8, 9). This regression-based approach is quite flexible and makes it possible not only to assess the effects of covariates (10) and multiple raters (11) on agreement, but also allows the investigator to describe the patterns of agreement more precisely than is done by using the kappa statistic alone.

We present the findings in the comparison of the epilepsy diagnosis from the review of the same histories and laboratory studies for 100 women, selected from a cohort study (12), by six neurologists, three of whom specialize in epilepsy among children and three, in adults.

MATERIALS AND METHODS

This project was carried out as part of a study of the teratogenicity of maternal epilepsy and anticonvulsant drugs that was conducted at five Boston-area hospitals between August 1, 1986 and December 31, 1993 (12). Three groups of newborn infants were enrolled: 1) those whose mothers had taken an anticonvulsant drug(s) during pregnancy; 2) women (called "seizure history" hereafter) who reported having had epilepsy previously, but were not taking anticonvulsant drug during this pregnancy, and 3) unexposed controls. The enrolled mothers were interviewed with regard to their seizures and any use of anticonvulsant drugs during pregnancy. Each woman was asked to sign a written release so the records of her previous neurologic evaluations and obstetrical management could be obtained. A concerted effort, usually requiring additional telephone calls and letters, was made to obtain these records.

The study epileptologist (S.K.) reviewed the information obtained and made a diagnosis as to the type of seizures each enrolled woman had had, using the 1981 International Classification of Epilepsy (13). The fact that a woman had epilepsy was considered confirmed if one of the following conditions was met: a) her medical records contained a specific diagnosis of epilepsy; b) the report of her EEG showed the presence of one of these abnormalities: spike and wave, epileptiform or paroxysmal tracing; c) her neurological records contained no firm diagnosis, but suggested that she had had epilepsy and noted that she was treated for epilepsy with an anticonvulsant drug; d) no neurological records were available, but other medical records stated that she had taken an anticonvulsant drug to prevent seizures; e) no neurological records were available, but other medical records stated

that she had taken anticonvulsant drugs for at least two years and the woman stated that the treatment was for epilepsy; f) the woman claimed that she had had epilepsy and the study epileptologist agreed that her description of her signs and symptoms in the questionnaire administered of these episodes were consistent with epileptic seizures.

In December, 1991, part-way through the seven-year study, it was noted that the research staff had diagnosed a higher portion (83%) of the seizures to be partial complex seizures, a rate much higher than the rates reported (32 to 52%; [14] for the general population of individuals with epilepsy. This observation led to the hypothesis that possibly an epileptologist who treats adults primarily interprets the history and findings in establishing a diagnosis differently from an epileptologist who evaluates primarily children. To test that hypothesis, we recruited five additional epileptologists, three who evaluate primarily children (G.L.H., E.C.D. and E.P.G.V.) and two (D.L.S. and D.H.) who evaluate primarily adults, so there were three adult neurologists and three pediatric neurologists participating in this project. At that time 231 women who had taken anticonvulsant drugs during pregnancy and 97 seizure history (no drug) women had been enrolled, interviewed and classified. We asked each epileptologist to review a random sample of 100 records, 67 from women who had taken anticonvulsants and 33 women with seizure history (no anticonvulsant drug). An additional ten cases were selected non-randomly as sample cases to familiarize the reviewers with the types of cases involved and the varying amount of information available on each case. Their assessment of these sample cases was not included in the analysis of the findings.

Each additional epileptologist was sent, first, 10 cases for practice. For each there was a one page summary of the woman's answers to questions about her seizures, and the findings in the records from her neurologic record. These instructions were given:

"The purpose of this project is not to see if physicians can be trained to arrive at the same diagnosis. It has been shown that a very high degree of agreement is attainable, given a set of decision rules and complete records (14–16). Our purpose is to see how often neurologists arrive at the same diagnosis, using the interpretation of the International Classification of Epileptic Seizures (13) that they use in their own practices.

"You will find that the records available in many cases are not as complete as you might wish, making your task more difficult. We realize that if you could examine the patient yourself or if additional information were available, your diagnosis might change; your task is to make the best choice you can, given the data the examining physician included in his/her case notes.

"In this study, primary generalized seizures and partial seizures are considered to be mutually exclusive, despite the fact that some neurologists feel that primary generalized seizures are associated with increased risk for partial seizures, and that the two often go together. We have to ask you to make your "best guess" as to which diagnosis best describes what is going on with the patient.

"For the purpose of this study, you can classify a woman's episodes as (1) not seizures; (2) partial seizures (simple or complex), with or without generalization; (3) primary generalized seizures. Only primary generalized seizures and partial seizures are considered mutually exclusive diagnoses; that is, a woman could have both seizures and non-seizure episodes (i.e. syncope, migraine), but not both primary generalized and partial seizures."

The participating neurologists were asked for their comments and suggestions about this methodology, which were discussed as part of the description of their task. Thereafter, they were sent 100 cases to review.

The medical records obtained on the 100 women whose diagnoses were reviewed varied from 2 to 35 pages of materials, with an average of 7.4 pages. There were notes from a neurologist on 96 women, at least one EEG on 87 and at least one head CT scan on 39 women. Only written information was provided to the neurologists. The neurologists did not examine the patients or review the EEGs or CT scans.

STATISTICAL METHODS

The findings were analyzed using an extension of the basic log-linear agreement model, described by Agresti (8, 9) and suggested by Tanner and Young (10) and Graham (11), to characterize agreement between two raters with respect to an ordered outcome variable. To describe the basis agreement between the two groups of neurologists, two log-linear models were used, one for adult neurologists and one for pediatric neurologists. Fitting a log-linear model involves organizing the data into multidimensional tables and modeling the log of the number of observations in each “cell” of the table. The most basic decision by each neurologist was choosing among three categories of diagnosis: no seizures, partial seizures (with or without secondary generalization), and generalized seizures. To tabulate these decisions a $3 \times 3 \times 3$ table with 27 cells was established. For example, the (1,1,1) cell contained the number of cases in which all three neurologists chose category one, a diagnosis of “no seizures” (Appendix 1). Similarly, the (1,2,3) cell contained the number of patients for whom the first neurologist diagnosed category one (no seizures), the second diagnosed category two (partial seizures), and the third diagnosed category three (generalized seizures).”.

The log-linear models used terms that specified both the diagnosis made by each of the three neurologists and the sources of information used previously. For example, if we were interested in only one covariate, such as the use of the EEG by all three neurologists, we would form $3 \times 3 \times 3$ tables, one for those subjects for whom all three neurologists reported use of the EEG, and one for the remainder of the subjects (Appendix 2).

Several sources of information were used to explain the beyond-chance agreement among the neurologists: the EEG, the patient’s medical history, previous physician diagnoses, and the findings in the physical examination. In addition, the neurologist’s level of confidence in the diagnosis, as well as their thinking that certain cases were “classic” for the diagnosis was assessed.

RESULTS

In order to assess agreement, the seizures were classified in one of four non-overlapping categories: 1) generalized seizures; 2) partial seizures that became generalized; 3) partial seizures that did not become generalized; or 4) no seizures at all. The degree of agreement among the adult neurologists and among the pediatric neurologists was tabulated separately (Table 1). The adult neurologists concluded on average, that 35 percent of the women had partial seizures, which became generalized secondarily, 21 percent had partial seizures that did not generalize, 31 percent of the women had generalized seizures, and 13 percent had no seizures (Table 1). The pediatric neurologists, on average, decided that 27 percent were diagnosed with partial seizures that became generalized secondarily, 11 percent were diagnosed with partial seizures that did not generalize, 40 percent of the women had generalized seizures, and 22 percent did not have seizures.

When the two classes of partial seizures (with and without becoming generalized) were collapsed into one category, the degree of agreement was tabulated for the three diagnoses: generalized, partial, or no seizures. If restricted to these three diagnoses, the adult neurologists diagnosed more generalized seizures (40 percent) than partial seizures (38 percent).

The adult and pediatric neurologists differed also according to reliance on sources of information in making the diagnosis (Figure 1). The pediatric neurologists tended to rely more heavily on the EEG, patient history, prior physician diagnosis, and physical exam than did the adult neurologists.

As for overall agreement (Table 2), adult neurologists agreed on the seizure diagnosis 59 percent of the time; specifically, 16 percent of the cases were diagnosed as generalized seizures, 38 percent as partial seizures, and 5 percent as no seizures (Table 3). Lack of agreement was present in the remaining 41 percent.

Overall, agreement rates among the three pediatric neurologists tended to be lower than occurred among the three adult neurologists. The pediatric neurologists agreed 44 percent of the time (Table 2). When the diagnosis categories of generalized, partial, and no seizures were used, all three neurologists diagnosed generalized seizures in 22% of cases, diagnosed partial seizures in 20%, and diagnosed no seizures in 2% (Table 3). Lack of agreement was present in the remaining 56 percent. The most common (19 percent) pattern of lack of agreement among pediatric neurologists was when one or two diagnosed no seizures and the other diagnosed partial seizures. The next most common pattern of lack of agreement occurred when one or two diagnosed generalized seizures and the other diagnosed either no seizures (16 percent of the time) or partial seizures (15 percent of the time).

When we examined agreement among the three diagnoses of generalized, partial, or no seizures among adult neurologists, log-linear models suggested that there was significant ($p=.03$) based on log-linear model agreement beyond that which is expected by chance alone. For example, the adult neurologists agreed 59 percent of the time (Table 2), a rate much higher than would be expected by chance alone (21 percent) [Table 1].

Among the adult neurologists, exact agreement occurred most often for the diagnosis of partial seizures (Table 3). For that diagnosis, adult neurologists agreed 38 percent of the time. The neurologists agreed 16 percent of the time on the diagnosis of generalized seizures and agreed 5 percent of the time on the diagnosis of no seizures. We can compare these percentages of exact agreement with the expected percentages of agreement by chance, as presented in Table 3. Using log-linear models, we found that agreement that seizures were generalized was significant statistically ($p<.01$) from log-linear model, and although agreement in the other diagnosis categories was present, it was not statistically significant.

Since the pediatric neurologists agreed 44 percent of the time compared to the 12 percent agreement expected by chance, the corresponding kappa statistic for agreement was 0.36, which indicates only fair agreement (9).

AGREEMENT BY SOURCE OF INFORMATION USED

Four different sources of information were considered, which affected the level of agreement between the neurologists: the woman's EEG finding, the medical history provided, when the findings were considered "classic" for the epilepsy diagnosis and when the neurologist was most "confident" of the diagnosis. When all three adult or pediatric neurologists used information from the EEG in making a diagnosis, agreement was greater (Table 2). In fact, for the group of all 100 women, adult neurologists agreed on the fact that

they had epilepsy 59 percent of the time. Moreover, for the 33 women in whose evaluation an EEG was used by the adult neurologists in making the diagnoses, agreement increased to 79 percent of the time (Table 3). This increase in agreement has high statistical significance ($p < 0.0001$) from the log-linear model. For the specific diagnosis of partial seizures, agreement was present 38 percent of the time in the group of all 100 cases. This agreement increased to 45 percent of the time when we consider only those cases in which all three neurologists used the EEG ($p < 0.01$) [Appendix 5]. For the diagnosis of generalized seizures, agreement was present 16 percent of the time in the group of all 100 women and increased to 33 percent when we consider women for whom an EEG report was available ($p < 0.01$).

In addition, if two out of three of the adult neurologists thought that the seizures were “classic” for the diagnosis, agreement was also increased from 59 percent in the entire group to 82 percent ($p = 0.05$) [Table 3]. Specifically, for the diagnosis of partial seizures, agreement was increased from 38 percent to 82 percent ($p = 0.05$). Finally, as one might expect, agreement was increased when a higher level of confidence was reported by the group of neurologists as a whole. In general, agreement was increased from 59 percent for the entire group to 87 percent ($p < 0.01$). In the case of the “no seizure” diagnosis, agreement was increased from 5 percent to 8 percent ($p = 0.05$); for the case of partial seizures, agreement was increased from 38 percent to 50 percent ($p = 0.01$), and for the case of generalized seizures, agreement was increased from 16 percent to 29 percent ($p < 0.01$). Although the use of the patient’s history did not affect agreement significantly, in general, it actually decreased agreement ($p < 0.01$) in the generalized diagnosis from 16 percent to only 13 percent (Table 2). That is, neurologists were less likely to agree that seizures were generalized in those women whose histories were used as opposed to those women whose medical histories were not used for diagnostic purposes.

We measured also agreement between neurologists that was not necessarily exact. The adult neurologists agreed about whether some type of seizure had occurred in 83% of the cases. This agreement was significant statistically ($p = 0.03$) from the log-linear model.

For pediatric neurologists, the use of certain sources of information did not have such a great influence on agreement. Although none of these covariates included in the model either increased or decreased agreement in general, the physician’s confidence increased significantly agreement in the case of generalized seizures from 22 percent to 28 percent ($p = 0.04$) from log-linear model.

One explanation for why none of the covariates affected agreement significantly in the original pediatric model is the fact that the pediatric neurologists relied more heavily on the EEG, history, prior diagnosis, and physical exam than the adult neurologists (Figure 1). In addition, pediatric neurologists were more likely to classify seizures as “classic for the diagnosis” than adult neurologists were. So, although none of the covariates alone contributed significantly to agreement, it is possible that the overall agreement was affected by the high reliance of pediatric neurologists on these diagnostics, in general.

As in the adult neurologist model, we measured agreement among neurologists that was not necessarily exact. The neurologists agreed about whether any type of seizure had occurred only 59 percent of the time. This agreement was not significant statistically beyond agreement expected by chance.

In order to assess agreement more exactly, we fit a model that took into account four categorizations of the seizures by separating the partial seizure category into two groups: 1) partial seizures that generalized; 2) partial seizures that do not generalize. Among adult neurologists, log-linear models suggested that the exact agreement term was highly

significant ($p < 0.01$) from log-linear model. In addition, exact agreement for each diagnosis was highly significant except in the case of partial seizures that became secondarily generalized; here, a slight increase in agreement was not found to be statistically significant.

In this expanded model for adult neurologists with the more specific partial seizure diagnosis, a slightly different subset of covariates was found to be significant in improving agreement. As in the first model, agreement was increased when all three neurologists used the EEG ($p = 0.04$) and when a high confidence level was reported ($p = 0.01$). However, use of patient history decreased overall agreement with marginal significance ($p = .06$).

When four diagnostic categories were used, the EEG and high physician confidence level increased significantly agreement only in the diagnosis of generalized seizures ($p < 0.01$ and $p = 0.05$), respectively. The decrease in agreement when patient history was used was found to be significant only in the diagnosis of generalized seizures ($p < 0.01$). When two of the three physicians thought that the seizures were classic for the diagnosis, agreement was significantly increased ($p = 0.03$) in the diagnosis of partial seizures that become secondarily generalized.

When the pediatric log-linear model was extended to allow for the four category classification of seizures, the model changed only slightly. There still was significant exact agreement ($p < 0.01$). In addition, the log-linear model suggested that agreement was significant statistically in concluding that the diagnoses were partial seizures that did not generalize ($p < 0.01$) and partial seizures that became secondarily generalized ($p < 0.01$). Once again, physician confidence increased significantly agreement ($p = 0.04$) in the case of generalized seizures.

It is notable that adult neurologists were more likely to diagnose partial seizures and pediatric neurologists, to diagnose more generalized seizures (Table 1). In the case of pediatric neurologists agreement was significant statistically for the diagnosis of partial seizures; in the case of adult neurologists, agreement was significant statistically for the case of generalized seizures. When the adult neurologist concluded that the woman had had a seizure and was more uncertain of the specific type, the diagnosis of partial seizures might be more likely due to the adult neurologist's tendency to diagnose more partial seizures. On the other hand, the pediatric neurologists appeared to be more careful about diagnosing partial seizures and might tend to call those episodes that are more difficult to diagnose generalized seizures.

DISCUSSION

A strength of this project was the fact that the neurologists used case material from a series of patients enrolled consecutively in a study of the fetal effects of maternal epilepsy and anticonvulsant drugs, not theoretical case histories. This comparison of the interpretations of six experienced neurologists showed that the three adult neurologists agreed 59% of the time. They disagreed most often over whether or not the woman's seizures were partial or generalized. Pediatric neurologists showed less agreement than the adult neurologists, but they still had a significant degree of agreement.

These findings are similar to those reported previously in which experienced neurologists evaluated the same clinical information with the goal of classifying the patient's epilepsy (3–5). For example, Van Donselaar et al (4) reported the findings by an expert committee that evaluated 207 children with a possible first seizure. They concluded that 156 (75%) had a definite seizure disorder. There was no consensus among the experts on the diagnosis for 51 (24.6%) of the 207 children. The findings in the EEG did not help very much to clarify the diagnoses. In another study, Camfield and Camfield (5) estimated that individual experts

disagreed about syndrome classification in 30% of cases. The authors concluded: “There is a huge need for a more accurate syndrome scheme that uses unequivocal definitions”.

Bodensteiner et al (3) reported the comparison of the classification by four observers (one senior neurologist and three neurology residents) of the same verbatim descriptions of seizure manifestations transcribed from medical records. The overall agreement in classifying seizure types was poor, whereas the classification of specific types of epilepsy was fair to excellent for most types. They concluded that the use of specific criteria for the categorization of symptoms would improve reliability and recommended the development of more explicit criteria for the diagnosis of specific seizure types.

Two factors shown to improve consistency of the diagnoses made are the passage of time and the addition of more extensive monitoring. Hauser et al (17) noted that only 50% of patients with epilepsy are diagnosed within the first six months of the onset of symptoms. It took 5 years for 85% of the patients to be diagnosed. Berg et al (18) showed that reassessment two years after the initial diagnosis showed a change in 13.5% of the “consensus” diagnoses.

Shinnar et al (19) re-evaluated 9 years later the classification of the epilepsy syndrome for 182 children and changed the syndrome classification for 18%.

Foley et al (20) compared the findings in routine interictal electroencephalograms (EEG) with the findings in long-term computer-assisted outpatient EEG monitoring in 84 children and adolescents. The seizure diagnosis and classification was the same in 19%, but differed in 63% of the patients.

Only 92% of the time, adult neurologists relied on the woman’s history compared to 100% of the time for pediatric neurologists. This may be attributed to the difference between relying on the adult for self-reporting in comparison to relying on the history provided by the parents of the child with epilepsy. The reliability of self-reported information about epilepsy was shown in studies of twins by Corey et al (21) to be under-estimated. By contrast, the parents are more likely to provide a better history.

The fact that between 11 and 33% of these women were considered not to have epilepsy was a surprising conclusion. It emphasizes the fact that psychogenic non-epileptic seizures, a potential alternative diagnosis for these women, are very common. The prevalence of psychogenic non-epileptic seizures is between 1/50,000 and 1/3,000, or 2 to 33 per 100,000, and between 10–20% of individuals referred to epilepsy center have non-epileptic seizures (22), thus making this a significant neurologic condition.

The fact that the level of agreement was not higher reflects the fact that diagnosing seizures or classifying epilepsy syndromes is not an exact science. This inexactness reflects the myriad behavioral and EEG features of epilepsy. It may also reflect weakness or inadequacy of the classification system.

Acknowledgments

We appreciate the splendid cooperation of the enrolling mothers and their doctors who made this study possible.

References

1. Houfek EE, Ellingson RJ. On the reliability of clinical EEG interpretation. *J Nerv Ment Dis.* 1959; 128:425–437. [PubMed: 13665365]

2. Ottman R, Hauser A, Stallone L. Semistructured Interview for Seizure Classification: Agreement with Physicians' Diagnoses. *Epilepsia*. 1990; 31(1):110–115. [PubMed: 2406127]
3. Bodensteiner JB, Brownsworth RD, Knapik JR, Kanter MC, Cowan LD, Leviton A. Interobserver Variability in the ILAE Classification of Seizures in Childhood. *Epilepsia*. 1988; 29:123–128. [PubMed: 3258236]
4. van Donselaar CA, Geerts AT, Meulstee J, Habbema JDF, Staal A. Reliability of the diagnosis of a first seizure. *Neurology*. 1989; 39:267–271. [PubMed: 2915800]
5. Camfield P, Camfield C. Childhood epilepsy: What is the evidence for what we think and what we do? *J Child Neurol*. 2003; 18:272–287. [PubMed: 12760431]
6. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960; 20:37–46.
7. Melia, BM.; Diener-West, M. Modeling inter-rater agreement for pathologic features of choroidal melanoma. In: Lange, N.; Ryan, L., et al., editors. *Case Studies in Biometry*. New York: John Wiley and Sons; 1994.
8. Agresti A. a model for agreement between ratings on an ordinal scale. *Biometrics*. 1988; 44:539–548.
9. Agresti, A. *Categorical Data Analysis*. New York: John Wiley; 1990.
10. Tanner MA, Young MA. Modeling agreement among raters. *J Am Stat Assoc*. 1985; 80:175–180.
11. Graham P. Modeling covariate Effects in Observer Agreement Studies: The Case of Nominal Scale Agreement. *Statistics in Medicine*. 1995; 14:299–310. [PubMed: 7724915]
12. Holmes LB, Harvey EA, Coull BA, Huntington KB, Khoshbin S, Hayes AM, Ryan LM. The teratogenesis of anticonvulsant drugs. *N Engl J Med*. 2001; 344:1132–1138. [PubMed: 11297704]
13. International Classification of Epilepsy 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. [PubMed: 6790275]
14. Hauser, WA.; Hesdorffer, DC. *Epilepsy: Frequency, Causes and Consequences*. Demos; NY: 1990.
15. Sandis RJ, Koch GG. A one-way components of variance model for categorical data. *Biometrics*. 1977; 33:671–679.
16. Dorsey BL, Nelson RO, Hayes SC. The effects of code complexity and of behavioral frequency on observer accuracy and interobserver agreement. *Behavioral Assessment*. 1986; 8:349–363.
17. Hauser WA, Kurland LT. The epidemiology of epilepsy in Rochester, Minnesota, 1935 through 1967. *Epilepsia*. 1975; 16:1–66. [PubMed: 804401]
18. Berg AT, Shinnar S, Hauser WA, Alemany M, Shapiro ED, Salomon ME, Crain EF. A prospective study of recurrent febrile seizures. *N Engl J Med*. 1992; 327:1122–1127. [PubMed: 1528207]
19. Shinnar S, O'Dell C, Berg AT. Distribution of epilepsy syndromes in a cohort of children prospectively monitored from the time of their first unprovoked seizure. *Epilepsia*. 1999; 40:1378–1383. [PubMed: 10528932]
20. Foley CM, Legido A, Miles DK, Chandler DA, Grover WD. Long-term computer-assisted outpatient electroencephalogram monitoring in children and adolescents. *J Child Neurol*. 2000; 15:49–55. [PubMed: 10641611]
21. Corey LA, Kjeldsen MJ, Solaas MH, Nakken KO, Friis ML, Pellock JM. The accuracy of self-reported history of seizures in Danish, Norwegian and U.S. twins. *Epilepsy Res*. 2009; 84:1–5. [PubMed: 19128944]
22. Benbadis SR, Hauser AW. An estimate of the prevalence of psychogenic non-epileptic seizures. *Seizure*. 2000; 9:280–281. [PubMed: 10880289]

Highlights

- We observed how “adult” neurologists classify type of epilepsy in pregnant women.
- We observed how “pediatric” neurologists classify type of epilepsy in children
- Adult neurologists agreed on the type of epilepsy more than the pediatricians
- Both groups showed more agreement when EEG results were available

Figure 1. Findings relied upon the adult neurologists in comparison to pediatric neurologists
Over EEG, the columns show that 60% of the adult neurologists and 80% of the pediatric neurologists used the EEG findings in establishing the type of epilepsy.

Table 1

Frequencies of seizure types made by adult and pediatric neurologists.

Neurologist	N*	PPN**	PG	G
Adult Neurologist 1	0.14	0.20	0.37	0.29
Adult Neurologist 2	0.11	0.19	0.44	0.26
Adult Neurologist 3	0.14	0.24	0.25	0.37
<u>Adult Average</u>	0.13	0.21	0.35	0.31
Pediatric Neurologist 1	0.33	0.03	0.33	0.31
Pediatric Neurologist 2	0.08	0.23	0.15	0.54
Pediatric Neurologist 2	0.25	0.07	0.33	0.35
<u>Pediatric Average</u>	0.22	0.11	0.27	0.40

Legend: N= no seizures; PN=partial seizures that did not generalize; PG=partial seizures that generalize, G=generalized seizures.

* Under N, 0.14, 0.11 and 0.14 mean that the three adult neurologists agreed that the women did not have seizures in 14%, 11% and 14% of the women, respectively; the pediatric neurologists concluded that 33%, 8% and 25% of these women did not have seizures, respectively.

** Under PN, the three adult neurologists decided that 20%, 19% and 24% of the women had PN (partial seizures that did not generalize), and 3%, 23% and 7% of the pediatric neurologists concluded that these women had PN.

Table 2

Patterns of diagnosis between neurologists when different sources of information were used.

		{-----lack of agreement-----}		{-----lack of agreement-----}		{-----lack of agreement-----}		{-----lack of agreement-----}	
	# Cases	Agree	P v. N	G v. N	P v. G	N v. P	P v. G	N v. P	P v. G
Overall Agreement									
Adult	100	0.59 *	0.11 **	0.04 ***	0.24	0.02	0.24	0.02	0.02
Pediatric	100	0.44	0.19	0.16	0.15	0.06	0.15	0.06	0.06
Agreement when EEG used									
Adult	33	0.79	0.03	0.00	0.18	0.00	0.18	0.00	0.00
Pediatric	46	0.50	0.15	0.15	0.13	0.07	0.13	0.07	0.07
Agreement when history used									
Adult	80	0.59	0.13	0.01	0.24	0.03	0.24	0.03	0.03
Pediatric	99	0.44	0.18	0.16	0.15	0.06	0.15	0.06	0.06
Agreement when findings were "classic" for diagnosis¹									
Adult	11	0.82	0.00	0.00	0.18	0.00	0.18	0.00	0.00
Pediatric	23	0.65	0.04	0.13	0.09	0.09	0.09	0.09	0.09
Agreement with group high confidence²									
Adult	100	0.59	0.11	0.04	0.24	0.02	0.24	0.02	0.02
Pediatric	100	0.44	0.19	0.16	0.15	0.06	0.15	0.06	0.06

N=no seizures, P=partial seizures, G=generalized seizures.

* = adult neurologists agreed 59% of the time;

** = adult neurologists diagnosed partial seizures 11% of the time while others diagnosed no seizure.

*** = adult neurologists diagnosed generalized seizures 4% of the time while others diagnosed no seizure.

¹ They marked whether the seizures described were "classic" for diagnosis.

² Each neurologist was asked how confident he or she was on a scale of 1 to 5 in the diagnosis. Used cut-off of 50% above or below.

Table 3

The frequency and percentage of agreement between neurologists for specific diagnoses.

		Observed Exact Agreement***			Expected Exact Agreement****		
	Number of women	N*	P	G	N	P	G
Overall Agreement							
Adult Neurologist (Proportion)	100	5*** (0.05)	38 (0.38)	16 (0.16)	0.2*** (0.03)	17.6 (0.18)	2.8 (0.03)
Pediatric Neurologist (Proportion)	100	2 (0.02)	20 (0.20)	22 (0.22)	0.7 (0.01)	5.4 (0.05)	5.9 (0.06)
Agreement when EEG used							
Adult Neurologist (Proportion)	33	0 (0.00)	15 (0.45)	11 (0.33)	0.0 (0.00)	6.2 (0.19)	2.3 (0.07)
Pediatric Neurologist (Proportion)	46	2 (0.04)	7 (0.15)	14 (0.30)	0.4 (0.01)	1.2 (0.03)	4.8 (0.10)
Agreement when history used							
Adult Neurologist (Proportion)	80	5 (0.06)	33 (0.41)	10 (0.13)	0.2 (0.00)	17.7 (0.22)	1.3 (0.02)
Pediatric Neurologist (Proportion)	99	2 (0.02)	20 (0.20)	22 (0.22)	0.6 (0.01)	5.4 (0.05)	6.0 (0.06)
Agreement when classic for the diagnosis							
Adult Neurologist (Proportion)	11	0 (0.00)	9 (0.82)	1 (0.09)	0.0 (0.00)	5.9 (0.54)	0.0 (0.00)
Pediatric Neurologist (Proportion)	23	1 (0.04)	6 (0.26)	8 (0.35)	0.1 (0.00)	1.1 (0.05)	2.5 (0.11)
Agreement with high confidence							
Adult Neurologist (Proportion)	24	2 (0.08)	12 (0.50)	7 (0.29)	0.0 (0.00)	4.1 (0.17)	1.1 (0.05)
Pediatric Neurologist (Proportion)	76	2 (0.03)	14 (0.18)	21 (0.28)	0.3 (0.00)	3.1 (0.04)	7.1 (0.09)

* = N = no seizures; P=partial seizures; G=generalized seizures.

** = The number (5) is the number of adult neurologists who agreed that the woman did not have seizures; the number in parentheses (0.05) signifies the percentage (5%) of women who were classified as not having seizures.

*** = "Observed exact agreement" is the likelihood of agreement expected by chance alone. For example, we expected adult neurologists to agree on the diagnosis of partial complex seizures 17.6% of the time, but they actually agreed 38% of the time.

*** = "Expected exact agreement" is the likelihood of agreement based on the interpretations of each neurologist in each group as listed in Table 1. For example, the three adult neurologists agreed on the diagnosis of no epilepsy in 11%, 14% and 14% of the women. The calculation of agreement for all 100 women would be: $0.11 \times 0.14 \times 0.14 = 0.992 \times 100 \text{ women} = 0.2$.