Clinical diagnoses and EEG interpretation

GEORGE W. WILLIAMS, PHD; RONALD P. LESSER, MD; J.B. SILVERS, PHD; ABRAHAM BRICKNER, PHD; MARLENE GOORMASTIC, MPH; KATHY J. FATICA, BS; DONALD W. KLASS, MD

■ A sample of 62 electroencephalographers in the United States evaluated 10-second samples of eight electroencephalograms. The evaluations were performed with and without knowledge of the clinical history. Evaluations consisted of multiple choice questions related to electroencephalographic observations, clinical diagnosis, and requests for additional tests such as computerized tomography and cerebrospinal fluid studies. The results indicate that clinical history influences interpretation, with considerable variation among readers in the number and type of additional tests requested.

□ INDEX TERMS: DIAGNOSTIC ERRORS; ELECTROENCEPHALOGRAPHY □ CLEVE CLIN J MED 1990; 57:437–440

ACCURACY and reliability in interpreting an electroencephalogram (EEG) may be a function of several variables, including reader experience and the availability of other clinical information. These possibilities were examined by asking a group of electroencephalographers to interpret several EEGs before and after clinical information was made available.

■ *See Chilcote (pp 477–479)*

Little and Raffel[1] studied the question of intra-rater variability in EEG interpretation by having three readers each evaluate 100 tracings on two occasions, from several days to several months apart. The readers were given clinical data available at the time the tracings were made and used a three-point rating system (normal, borderline, or abnormal) to interpret the tracings. In no instance was a tracing interpreted as normal during one reading and abnormal in the other. The authors concluded that experienced electroencephalographers are consistent in their judgment of at least the degree of abnormality of a tracing.

Woody[2] examined 15 pairs of electroencephalograms, composed of records of 15 subjects with behavioral problems and records of 15 normal counterparts. Except for chronological age and sex, all identifying characteristics were removed, eg, name, group membership, and case history. The electroencephalograms were coded and placed in random sequence. The electroencephalographer judged each of the records and then rejudged them approximately 8 months later, this time with the EEGs placed in a new random order. The interpretations varied greatly, leading the author to conclude that electroencephalographers were inconsistent in their ability to duplicate their findings.

The authors of these two studies reached different conclusions about the extent of intraobserver variability in EEG interpretation. However, the designs of the two stu-

From the Departments of Biostatistics and Epidemiology (G.W.W., M.G., K.J.F.) and Services Research (A.B.), The Cleveland Clinic Foundation; Department of Neurology, The Johns Hopkins School of Medicine, Baltimore, Maryland (R.P.L.); Health Systems Management Center, Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio; and Department of Neurology, The Mayo Clinic, Rochester, Minnesota (D.W.K).

Address reprint requests to G.W.W., Department of Biostatistics and Epidemiology, The Cleveland Clinic Foundation, One Clinic Center, 9500 Euclid Avenue, Cleveland, Ohio 44195

dies were quite different, with differing degrees of clinical information available to the electroencephalographers.

In the present study, we examined the impact of clinical information on EEG interpretation, along with interobserver variability in the clinical diagnosis and in requests for additional diagnostic tests.

BACKGROUND

In an earlier study,[3] EEG readers were randomly selected from among nationally recognized experts and certified and noncertified members of EEG societies. Selection was limited to those who were currently active in reading EEGs (ie, they interpreted five or more EEG records per week). Of 177 individuals determined to be eligible, 100 (56%) actually participated.

All who agreed to participate received a questionnaire and 12 EEG recordings chosen to represent a wide variety of clinical situations. The questionnaire asked about training and experience and contained multiple-choice questions about the 12 recordings. These questions were in four categories relating to (1) age of the patient, (2) EEG finding, (3) artifact, and (4) consciousness of the patient. The recordings consisted of photocopied reductions of 10-second samples of EEGs using the International 10-20 Electrode Placement System.

Each of the 12 recordings had a "correct" response against which participants' responses in the four categories were measured. The percent of participants indicating the patient's actual age ranged from 14% to 94%. The percent of participants indicating agreement with the authors' evaluations ranged from 65% to 100% for consciousness, 16% to 100% for artifact, and 19% to 88% for EEG finding, depending on the recording being evaluated.

In addition, the participants' rate of reporting the "correct" response was examined in terms of various respondent characteristics such as EEG board certification, age, percent of time in clinical EEG work, and number of recordings interpreted annually.

The percent of "correct" responses was significantly higher for: (1) national experts and certified EEG society members, (2) respondents who spent more time in clinical EEG work, (3) respondents with longer full-time EEG training, and (4) respondents from laboratories that produce more recordings.

The respondent's age, year of completion of full-time EEG training, and number of recordings interpreted annually were only weakly associated with the percent of "correct" responses. Group membership and board certification appeared to be most related to the percent of "correct" responses. The study indicated that even today there is considerable variability in EEG interpretation and that the variability is influenced by specific reader characteristics.

METHODS

Participants in the original study[3] were invited to participate in a follow-up study, in which the readers would be sent a subset of 8 of the original 12 EEG recordings along with clinical histories for each one. They would be asked to answer questions about EEG observations and about the clinical diagnosis and further studies to be performed.

Of the 100 readers who participated in the original study, 94 expressed an interest in the follow-up study, and 62 returned the questionnaire. Participants recorded a 4-digit identification number of their own choice on the questionnaires for both the original and the follow-up study. These numbers were used to link data from the two studies and to assure anonymity for the respondents. All data were reviewed before computer analysis was performed. The period between the completion of the questionnaires for the original and follow-up studies ranged from a few days to 8 months.

RESULTS

Reader characteristics

In the follow-up study, the average age of the respondents was 46 years. The average full-time EEG training was 11 months, with full-time EEG training completed in 1968 on the average. Sixty-one percent of the respondents were EEG board-certified; the average year of certification was 1973. Twenty-three percent of the respondents' time was, on average, spent in clinical EEG work. The respondents were equally distributed among private, group, and university practices. The average number of recordings produced by the respondents' laboratories annually was 1,917, and the average number of recordings interpreted annually was 949. These demographic characteristics are comparable to those of the entire group responding to the original study.

Reader responses

The modal response, or the response given most frequently by the participants in their interpretations of the EEG recordings, ranged from 52% to 98%. When these results were compared with the authors' evaluation (including analysis of the entire EEG), the modal response agreed with the authors' evaluation except for one recording. We compared the EEG findings reported before clinical information was made available with those reported after information was made available, and the percent of participants reporting the same findings in both cases ranged from 50% to 89% (*Table 1*).

Requests for additional tests

Most respondents would order additional EEG studies and CT scans in most of the studied cases—a notable observation since five of the eight records were considered by the authors to be normal. However, there was never complete agreement on the need to order such additional tests; the percent of respondents ordering the tests ranged from 26% to 93% for an additional EEG and 27% to 88% for a CT scan, depending upon the recording being evaluated. Those who would request cerebrospinal fluid (CSF) studies ranged from 2% to 55%. A similar range of respondents, from 2% to 51%, indicated that they would not order any of the indicated tests. The type of information most frequently requested was confirmation of the primary diagnosis, but this reason was given by no more than 69% of the respondents for any case. Many reasons were frequently indicated for requesting a CT scan, but no single reason was given by more than 55% of the respondents for any case.

Reasons for additional EEGs

Special EEG methods, including sleep records, prolonged records, and video techniques, were frequently requested, although the particular method varied depending upon the recording. In most instances, additional EEGs were requested in order to confirm the primary diagnosis. The next most frequent reason for additional EEGs was to further specify the primary diagnosis and to obtain information for planning therapy. Rarely was the additional EEG used to identify the secondary diagnosis or for other reasons. Additional EEG studies were expected to provide more than one kind of information.

Reasons for CT scans and CSF studies

The reasons cited most frequently for requesting a CT scan included confirmation of the primary diagnosis, further specification of the primary diagnosis, and identification of the secondary diagnosis. Although many participants hoped to gain information for planning therapy from the CT scan, this was not the most frequently indicated reason. Few participants hoped to gain other types of information from the CT scan, although many of them expected it to yield multiple types of information.

Reasons frequently cited for requesting CSF studies included further specification of the primary diagnosis, identification of the secondary diagnosis, and obtaining information for planning therapy. For the three recordings for which CSF studies were most frequently requested, specification of the primary diagnosis was the reason most often indicated. Few participants hoped to confirm the primary diagnosis or to gain other types of information from CSF studies.

TABLE 1
EEG FINDINGS WITH AND WITHOUT A CLINICAL HISTORY FOR EIGHT RECORDINGS

	Recordings							
	1	2	3	4	5	6	7	8
Percent in total agreement	88	61	77	84	75	89	66	50
Percent in agreement on normal, abnormal	88	62	84	92	78	90	66	62
Of those found normal without clinical history, percent found normal with clinical history	96	68	89	0	84	93	30	58
Of those found normal with clinical history, percent found normal without clinical history	91	66	87	–	79	96	70	94

DISCUSSION

Interpretation of 10-second EEG samples is a challenge that differs from the ordinary activity of the clinical electroencephalographer. A 10-second sample is only a fraction of the length of a minimal waking EEG record. Longer samples are needed when the phenomena under study are transient, and when estimates of the subject's age and level of consciousness are sought. However, it is reasonable to assume a strong association between readers who accurately interpret 10-second samples and readers who accurately interpret full-length recordings. Hence, the results in this study would probably also apply to similar studies utilizing full-length recordings.

Limitations

It is conceivable that the selection of 10-second samples could significantly influence the results. Moreover, the terminology provided for the possible answers on the questionnaire may not be universally accepted (eg, hypnagogic hypersynchrony, electrodecremental pattern); hence, electroencephalographers could argue that some questions could not be answered reliably given the choice of answers.

The nonresponse rate in this study was an important factor despite extensive efforts to keep it to a minimum. This nonresponse may have introduced bias into this study since participants may not have been representative of their particular groups. Moreover, reasons for nonparticipation may vary from one group to another. Participants were encouraged to keep copies of their responses to the original and follow-up studies. Hence, agreements between the two evaluations may be biased compared to independent assessments.

Conclusion

These results indicate that EEG interpretations can differ before and after clinical information is available. As many as one third of the respondents changed their interpretation between normal and abnormal after clinical information was made available for a given case. Whether or not a reader was board certified did not seem to influence the interpretation of clinical information. The extent of variation in clinical diagnosis was reflected in the fact that the percent of participants indicating the modal response ranged from 38% to only 73%. Moreover, EEG readers vary considerably in their clinical diagnoses and the number, type, and rationale for additional tests. These results document the intuitively reasonable conclusion that the extent of other information available can influence the interpretation of EEGs.

ACKNOWLEDGMENT

The Cleveland Foundation generously provided grant support for this study as part of a larger grant to The Cleveland Clinic Foundation entitled "Medical Technology Assessment Models." Cindi Elber and Barbara Moster have provided invaluable assistance in the logistics of this project.

REFERENCES

1. Little SC, Raffel SL. Intra-rater reliability of EEG interpretations. J Nerv Ment Dis 1962; **135**:77–81.
2. Woody RH. Intra-judge reliability in clinical electroencephalography. J Clin Psychol 1966; **22**:250–254.
3. Williams GW, Luders HO, Brickner A, Goormastic M, Klass DW. Interobserver variability in EEG interpretation. Neurology 1985; **35**:1714–1719.