Neural network analysis of preoperative variables and outcome in epilepsy surgery

Jeffrey E. Arle, M.D., Ph.D., Kenneth Perrine, Ph.D., Orrin Devinsky, M.D., and Werner K. Doyle, M.D.

Department of Neurosurgery, Hospital of The University of Pennsylvania, Philadelphia, Pennsylvania; and Department of Neurology and Neurosurgery, New York University, New York, New York

Because appropriate patient selection is essential for achieving successful outcomes after epilepsy surgery, the need for more robust methods of predicting postoperative seizure control has been created. Standard multivariate techniques have been only 75 to 80% accurate in this regard. Recent use of artificial intelligence techniques, including neural networks, for analyzing multivariate clinical data, has been successful in predicting medical outcome. The authors applied neural network techniques to 80 consecutive patients undergoing epilepsy surgery in whom demographic, seizure, operative, and clinical variables to predict postoperative seizures data were obtained.

Neural networks were able to predict postoperative seizures in up to 98% of cases. Student's t tests or chi-square analysis performed on individual variables revealed that only the preoperative medication index was significantly different (p = 0.02) between the two outcome groups. Six different combinations of input variables were used to train the networks. Neural network accuracies differed in their ability to predict seizures using all data (96%); all data minus electroencephalography concordance and operative side (93%); all data except intra- or postoperative variables such as tissue pathology (98%); all data excluding pathology, intelligence quotient (IQ) data, and Wada results (84%); only using demographics and tissue pathology (65%); and only using IQ data (63%).

Analysis of the results reveals that several networks that are trained with the usual accepted variables characterizing the typical evaluation of epilepsy patients can predict postoperative seizures with greater than 95% accuracy.

Key Words * epilepsy * neural network * seizure * epilepsy surgery * temporal lobe surgery

Although predicting medical outcomes is an increasingly important need in all areas of the medical field, it is often difficult to obtain reliable clinical data in a controlled setting. Most clinical problems involve multiple variables with varying influence on a clinical state that changes within and across patients over time. If only a single variable is examined, much of its underlying significance in describing, evaluating, and/or treating a disease may be missed. Defining and studying outcome from surgery for the treatment
of epilepsy is an excellent example of this problem. Epilepsy is defined as electrical discharges that cause seizures and that recur in cortical sites. The characterization of factors in the cortical substrate that initiate and sustain a seizure is still unclear, and the gross phenotype of seizures varies across a number of categories. Thus, the prognosis of a patient with epilepsy, which is often grossly attributable to seizure frequency and based on any one of the variables that might describe an epileptic patient, remains unpredictable. Comprehensive outcome measures are also more difficult to define in this patient population,[23] although most components of the several, extant standardized outcome measurement instruments are either directly or indirectly related to the presence or absence of seizures and seizure severity.

That single variables can be prognostic is sometimes the case in medicine. The link between certain genes and particular diseases such as in the multiple diseases associated with mitochondrial DNA mutations makes this clear.[6] Several studies have been performed to investigate multiple variables for prediction of outcome after epilepsy surgery.[2,4,5,8,13,18] In a comprehensive review of the literature through 1985,[8] however, it was found that almost all of the variables had a different prognostic significance than in at least one of the other prior studies in which that same variable was examined. When the investigators used multivariate techniques such as multiple regression and cut-off analysis with up to 28 variables, a prognostic accuracy of approximately 80% was found in predicting seizure outcome after surgery. In summarizing data from their study, only factors such as a patient's sex (in which there was never a correlation), whether discharges were specifically localized in the anterior midtemporal lobe (always positively correlated), and whether the patient had an emotional adjustment disorder (always negatively correlated) were consistent in their relationship to outcome across multiple studies in which these variables were examined.[13]

This inconsistency in the prognostic significance of individual variables exemplifies the chimerical nature of this clinical problem. Analysis of more recent studies in which attempts to predict outcome following epilepsy surgery has corroborated the differences in the predictive power of single variables. Linear multivariate methods are less helpful, and it is increasingly more likely that many pieces of the puzzle are predominantly nonlinear in nature.[13] Although attempts have been made to combine disparate studies on the same topic by performing metaanalysis, multiple regression (including logistical regression), or discriminant analysis,[12] none of these techniques has resulted in sufficiently high predictive power. We have utilized a newer approach also used in a recent study by Grigsby, et al.,[12] in which neural networks are used to account for the potentially nonlinear variables in outcome after epilepsy surgery.

**CLINICAL MATERIAL AND METHODS**

Inclusion in this study was based on patients having undergone the following criteria: resective epilepsy surgery between March 1, 1994, and January 1, 1997, magnetic resonance (MR) imaging, a complete neuropsychological evaluation, a sodium-amytal (Wada) test, long-term video electroencephalography monitoring (VEEG) to capture ictal events, invasive chronically implanted subdural electrode VEEG monitoring prior to resection, and psychiatric evaluation as indicated. All patients who were included in network training and testing, however, had a minimum of 6 months of follow-up review from surgery; verbal, performance, and total IQ scores obtained using Wechsler Adult Intelligence Scale-Revised; evaluations of intraoperative electrocorticography (iECoG); separate right and left memory scores from the Wada test; adequate knowledge of their preoperative and follow-up antiepileptic medications and dosages; and a full pathological evaluation of the resected tissue.
Twenty patients from a potential pool of 100 did not meet inclusion criteria and were excluded from subsequent analyses. In nine of these patients inadequate neurological follow-up information was obtained, in six there were incomplete neuropsychological profiles, and in five patients either an incomplete or no Wada study was performed. There were 80 patients in whom complete data sets were obtained, and these were used to develop neural networks as described below. Neural networks were studied to predict one outcome measure: whether or not a patient had any seizures after resection at any time during the entire follow-up period.

**Magnetic Resonance Imaging Evaluation**

Coding of the MR image scans prior to surgery was based on evaluation of the neuroradiological readings and the detailed notes of the neurosurgeon (W.K.D.). Three categories were encoded: 1) whether there was laterality (none, uni-, or bilateral) of any of the findings, 2) whether or not there was hippocampal atrophy or increased signal in the mesial structures, and 3) whether or not there were any other abnormalities on the image. This method of coding the MR images resulted in having four binary nodes in the networks: two nodes that coded for the three laterality options (00, 01, 11) and one node each (present/absent - 0,1) for the other two categories.

**Electroencephalography Evaluation**

All available EEG evaluations, including both scalp, VEEG, and chronic invasive studies, were reviewed to determine coding for EEG laterality of ictal onsets. Codes were assigned in three categories as follows: 0, all unilateral; 1, greater than 75% unilateral; and 2, all bilateral.

**Neuropsychological Evaluation**

All patients underwent neuropsychological evaluations before deciding on the surgical course. Results from the WAIS-R verbal, performance, and full-scale IQ tests were uniformly obtainable among the 80 patients, although some required estimations[16,28] because not all subtests had been obtained. Other neuropsychological tests were not included in the database because they were not administered to all patients. Values from the tests were used directly for neural network training.

**Seizure Frequency**

Values for seizure frequency were estimated from clinic notes, admission notes, and other independent sources (such as neurosurgeon's notes) when available. Although it was difficult to rely solely on the family as a completely accurate source because of memory problems, their ability to recognize or keep an account of a patient's seizures, different types of seizures, incomplete charting, clustering of seizures, and changes in seizure patterns over time was nonetheless a potentially useful value for network training. Small differences between seizure-frequency estimates and the actual seizure frequency are unlikely to affect weights within any given network. This is related to the assumption, generally required for a network to generalize at all, that the function the network "solves" between inputs and outputs is smooth.[22] The estimated number of seizures per week was the value used directly in the networks. Additionally, the number of years the patient had experienced seizures prior to surgery was used as an input variable.

**Medication Index**

The medication index (MI) was created to address the issue of whether the overall amount of
antiepileptic medication could be viewed as an indication of severity of seizure disorder and, thereby, possibly an indication of postoperative prognosis. To accomplish this, pre- and postoperative antiepileptic medications were scaled first as a percentage of an arbitrary "high" daily dose for each particular medication (Table 1). The overall score for a given combination of medications was then obtained by adding together these scaled values for each of the patient's medications. In this manner, medication regimens between patients could be compared by using a single index of their "relative" antiepileptic medication dose. For instance, a patient who received 1000 mg three times daily of carbamazepine and 500 mg daily of phenytoin (an MI of 1.5), would be equivalent to a patient who received 2500 mg daily of felbamate, 2000 mg daily of vigabatrin, and 10 mg daily of clobazam (also an MI of 1.5). The MI, computed immediately preoperatively and at the most recent postoperative visit, was used directly in network training. Some patients may have been receiving doses of certain medications that were even above the amounts in this table. The corresponding value added in to their MI in those cases would be greater than 1.0. These arbitrary "high" dose values are not meant to be related necessarily to blood levels, suggested dosing regimens, or any other clinical correlate.

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ARBITRARY &quot;HIGH&quot; DOSES OF VARIOUS ANTI-EPILEPTIC MEDICATIONS</strong></td>
</tr>
<tr>
<td>Medication</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>phenytoin</td>
</tr>
<tr>
<td>carbamazepine</td>
</tr>
<tr>
<td>oxcarbamazepine</td>
</tr>
<tr>
<td>phenobarbital</td>
</tr>
<tr>
<td>valproate</td>
</tr>
<tr>
<td>lamotrigine</td>
</tr>
<tr>
<td>gabapentin</td>
</tr>
<tr>
<td>felbamate</td>
</tr>
<tr>
<td>vigabatrin</td>
</tr>
<tr>
<td>topiramate</td>
</tr>
<tr>
<td>primidone</td>
</tr>
<tr>
<td>ethosuximide</td>
</tr>
<tr>
<td>clobazam</td>
</tr>
<tr>
<td>clonazepam</td>
</tr>
</tbody>
</table>

**Intraoperative Electroencephalography Evaluations**

Recordings from the immediate intraoperative postresection period varied according to the exact location and amount of tissue removed, as well as in the region in which strip or grid electrodes were placed for monitoring. Typically, electrodes were placed at each of the resection margins, although other areas such as frontal cortex were also studied if they had been active preoperatively. The final recordings, after all resection procedures were completed, were used to determine whether or not epileptiform activity still existed. Efforts were made to separate injury spike activity from "true" epileptiform activity, and this potential difference was usually noted in the final reading of the iEEG. For networks, the variable was encoded as residual epileptiform activity (present/absent - 0,1).

**Pathological Categorization**

The final pathology of the tissue resected in each patient was encoded using three separate binary variables, each of which was evaluated for its presence or absence. These variables included: 1) cortical
dysplasia, 2) hippocampal sclerosis or cell loss, and 3) the finding of a tumor or other structural lesion. Combinations of these possible entities could thus also be taken into account, (for example, cortical dysplasia in the presence of a tumor). Extrahippocampal gliosis was not included as a separate category because it occurred in the vast majority of all tissue specimens.

**SURGICAL SITE(S)**

The location of tissue that was resected in each patient was divided into one of three categories: temporal lobe only (61 patients), predominantly temporal lobe plus some extratemporal (four patients), and predominantly extratemporal (15 patients). Thus, we included mostly those patients who underwent temporal lobe resections but who were able to allow the networks to learn variants that involved extratemporal regions as well. In addition, a variable (coded as 0 or 1) was used to denote the side of the resection, and a binary variable (0 or 1) was used to code for whether or not multiple subpial transections (MSTs) were performed.

**Neural Network Development**

Neural networks are typically created by allowing a collection of input nodes to take on the values of the variables found in the data set. These input nodes are then connected to one or more "hidden" layers of nodes via weights. Typically, each node of one layer connects via a separate weight to every node of the next layer (a feedforward net). Sometimes hidden layers may also project back to prior layers (a feedforward and feedback net). The value of any given hidden node, in the simplest forms of networks, is determined by summing the values of its inputs multiplied by each of their respective weights and then by performing a nonlinear type of function on the result (for example hyperbolic tangent function). At some point the last hidden layer projects to an output node or set of nodes. In these studies, only one binary output node was used to code for the presence or absence of seizures postoperatively. When the known real output (0 or 1) is compared with the obtained value of the output in the network, an error value is generated. This error is then distributed back through the network by adjusting all of the weights between nodes. The input values are presented again and a new error is generated, back-propagated through the network by adjusting the weights, and so on until the error that the network produces is arbitrarily close to the known result. When all of the data have been shown to a network in this way, the network will eventually "learn" to provide the appropriate output for any of the training input sets. The network will also generalize, in that it should be able to provide the appropriate output for input patterns that it has not seen before, assuming the system generating the data was the same.

Our networks were developed using commercially available software (NeuralWorks; Neuralware Inc., Pittsburgh, PA) that allows a wide range of network structures and learning algorithms. Although several companies design software to perform the appropriate calculations for neural network learning, more important is the ability to develop a data set that enhances training and allows for so-called "cross validation" of a test set of the real data; in this way, networks can be found that generalize well and are not fixed at having learned only the training data. Programs that allow this were written by one of the authors (J.E.A.) in MS-Fortran version 5.1 (Microsoft Corp., Redmond, WA).

Typical techniques for network development and refinement have been described in a prior study.[1] Briefly, back-propagation algorithms (the most common learning method) were analyzed in small networks with one or two hidden layers. Variations in network parameters (such as the type of learning rule used, momentum terms, activation functions, and network size) were adjusted until a "best" network for a given set of inputs was found. All of a network's predictions were obtained by a method (cross
validation or bootstrapping) in which each network structure was trained using most of the training data and then tested on the small number of original data left aside. The particular patients' data that was left aside for testing each time was randomly selected. Such training and testing, each time using a different test set, was performed for each network 5 times, with the resulting network's predictions of each test set averaged together. As such, a certain network may have been 100% accurate for 4 of the 5 tests after training, but then only 90% accurate on the last set, making it overall 98% accurate. This method is considered effective for network validation and has been used in multivariate analyses as well.[2,22] By performing cross validation and averaging over five tests, a second layer of validation further eliminates variability in a network's performance on a single random set of the test data.

In addition, the number of cases in the training set of data was increased by randomly altering individual cases so as to create "new" cases. All of these artificial training cases were based on original case data but increased the overall size of the training set to enhance network learning and generalization. This method has been used effectively in many other neural network studies,[14] and it is often essential to eliminate entrenchment of the network in the training set and the frequent overtraining that so often occurs in clinical data. All test sets used in cross validation, however, were composed only of actual data obtained in the series.

These methods were used to create networks for six different patterns of variables. As noted, the parameters for learning rules and the network size and structure were adjusted to optimize each network's predictive power. The six networks developed to predict outcome were based on either 1) all variables (as described previously), 2) all variables except the EEG data and the side on which operation is performed, 3) all variables except data obtained from intraoperative sources (that is, iECoG, tissue pathology, and whether multiple subpial transections were performed or not), 4) only EEG, MR imaging, age, sex, years with seizures, and the preoperative MI data, 5) only tissue pathology, age, and sex data, or 6) only neuropsychological data (verbal, performance, and full-scale IQ scores). These groupings were chosen to help delineate what degree of evaluation might be adequate to predict outcome.

RESULTS

Only four patients in the postoperative seizure group had a single seizure that was attributable either to alcohol abuse, acute medication withdrawal, or a severe fever from a viral infection and, thus, would have been arguably classified as Engel Class I despite having had a seizure. Of the 20 patients originally with incomplete data sets, there was enough information obtained in 11 patients to determine follow-up seizure activity. Five (45%) of these 11 had experienced at least one seizure during the follow-up period, which is similar to the percentage found in the 80 patients who met inclusion criteria.

Table 2 provides data from results of the student's t test and chi-square analyses performed on many of the individual variables. These were conducted to clarify whether certain individual variables had significance in predicting outcome independently. Some additional variables, which were derived from the original data but not used in neural network development, were also examined. These additional variables included "Estimated Total Seizures per Lifetime," "Verbal-Performance IQ Difference," and "Total Memory Score." At the set significance level alpha = 0.05, none of the variables was significantly different between groups of those who had experienced seizures and those who did not, except both the preoperative MI and the follow-up MI. There was a large mean difference between the groups with regard to estimated total lifetime seizures, but the wide standard deviations within the groups led to an insignificant p value of 0.11.
Resulting structures of the "best" networks are shown in Figure 1 (a-f). One network was found for each of the six sets of input variables which had the highest level of predictive accuracy in cross-validation testing. The network using all available data (Fig. 1a) predicted the incidence of a postoperative seizure with 96% accuracy. Using all data except intraoperative or postoperative variables (e.g. tissue pathology) was even better at 98% accuracy (Fig. 1b). Using all data except EEG concordance and operative side was 93% accurate (Fig. 1c). If all the data were used excluding pathology, IQ data, and Wada results the network was 84% accurate (Fig. 1d). Using only demographics and tissue pathology (Fig. 1e) resulted in only 65% accuracy, however, and using IQ data alone (Fig. 1f) had the least accuracy at 63%.
Fig. 1. Structures of the "best" networks developed for each of the six input patterns. a) Twenty-five inputs with two hidden layers of 30 and 15 nodes, respectively. b) Twenty-two inputs with 15 and seven nodes in the hidden layers. c) Twenty inputs with 13 and seven nodes in the hidden layers. d) Thirteen inputs with 13 and seven nodes in the hidden layers. e) Six inputs with 13 and seven nodes in the hidden layers. f) Three inputs with a single hidden layer of seven nodes. All nets have both feedforward and feedback connections, except for that depicted in "f" which has only feedforward connections.

**DISCUSSION**

This study differs from prior outcome studies on several accounts. The overall accuracy of prediction by using neural networks (when utilizing a large enough input space) is superior to prior models in predicting both successes and failures on an individual patient basis. In most prior multivariate studies, 70 to 85% accuracy is achieved in predicting validation data.[2,8,13] The networks described here ranged in accuracy from 63 to 98% depending on which input variables were used. This enhanced predictive ability of at least some of the networks is likely caused by the nonlinear nature of network learning and processing. Although feedforward nets with one hidden layer are similar to projection pursuit regression statistical methods and although probabilistic neural nets are identical to kernel discriminant analysis, feedforward and feedback networks that have more than one hidden layer and nonlinear activation functions (hyperbolic tangent function), as were used in this study, have only superficial similarities to other more traditional multivariate statistical analyses.[22]

Additionally, in this study we enrolled consecutive patients from the same evaluation center, involving the same epileptologists and surgeon, over a relatively brief, recent time period. Although the authors of
other studies, recognizing the importance of these issues, have attempted either to maintain this same consistency or to correct for it when it does not exist, no prior study has achieved this degree of internal consistency. Whereas such consistency may not contribute, for example, to overall outcome results (such as seizure-free plus occasional seizure percentages across all patients in larger nontumor epilepsy series are typically reported between 40-80%[9,29]), such internal consistency does allow a more robust assessment of the method of analysis. It is less clear, for instance, what surgical techniques have been used or how patients have been evaluated when they receive antiepileptic drug combinations prior to surgery when such data might span a 5-year or longer period and when such a time period preceded the study by 10 or more years. Combinations of such variables may even be 80% predictive of outcome, but they still leave unresolved exactly what methods or criteria made up the variables analyzed, making it difficult to rely on the conclusions.

Our results are consistent with those found recently by Grigsby, et al.,[12] who used similar neural network analysis methods. In their study they analyzed a similar number of patients, pooled from three epilepsy surgery centers, and found predictive abilities of the networks to be similar to those demonstrated here (approximately 95%). They also compared their networks to more typical linear discriminant analysis, where prediction was only in the 80% range, which was similar to results obtained in prior studies in which those methods were used.

Although it demonstrates the ability to predict outcome accurately after epilepsy surgery, this study has potential deficiencies as well. These are found in two particular areas: those related to the nature of the data and its collection, and those related to the inherent use of neural networks for prediction. The data collection process in this study was slightly weakened by the incomplete data sets in 20% of the 100 overall cases. Whereas the outcome data in terms of seizure occurrence does not seem to be greatly affected (similar percentages of seizures were found in excluded patients' outcomes, when available), such a deficit suggests the general problem of initiating and maintaining an epilepsy surgery database. Rarely ever clear is knowledge of the patient's compliance, true day-to-day functioning, and medication awareness, in combination with the attempts to quantitate numbers of seizures during follow-up review, type and severity of seizures, and related symptoms of language difficulties or affective changes.

In addition, several authors have suggested problems inherent in using "static" measures in this population. For example, although certain variables may individually have predictive value, Dodrill, et al.,[7] have pointed out the increasing likelihood of finding a positive or negative correlation as the number of analyzed variables in any single study increases. Moreover, seizure frequency during the follow-up period may vary over time.[10] Some patients (up to 38% in a latter study) were free of seizure several years prior to follow-up examination, although they experienced a few seizures in the 1st postoperative year. In light of our criteria these patients would have been considered failures in this study. This suggests that the criterion for success might be pushed to an extreme: becoming both seizure and medication free. In contradistinction, perhaps the time limit for the follow-up period should be standardized at 3 years instead of 3, 6, or 12 months; however, some authors have shown that the probability of remaining without seizures after being seizure-free for 1 postoperative year is well over 85%,[9, 23] and the probability of continuing to experience seizures when seizures occur within the 1st year postoperatively is also high.[29] This difficulty in more precisely understanding the natural postoperative seizure history in patients who have undergone resection has been noted elsewhere.[24,29]

To make the best presurgery predictions of outcome for patient and family guidance, optimally as few and as reliable variables as possible should be used. In this sense, we have tried to select variables and
use them in such a way as to achieve this goal. We have omitted other variables that might be useful in outcome prediction, such as early risk factors,[11,27] positron-emission tomography in hypometabolism,[19,21,26] more detailed MR imaging analysis,[15,17,26] the presence and type of aura and subclinical seizures,[25] and more detailed breakdowns of EEG or iECoG findings[3,20], among others. Because equivalent significance among most of these factors is not found in all studies, further work toward understanding how these additional factors should be defined more uniformly and physiologically related to seizure activity needs to be done.

In terms of neural network analysis, several important issues exist. Although exhaustive testing with many network structures and parameters is performed to find a best network, there is no standardized algorithm for finding such a network. Ultimately, the resulting networks are determined based on the experience of the investigator and the nature of the data. This methodology leaves open the possibility that either the "best" network may not have been found or that a number of networks would work just as well for any given set of input variables. Randomized input variables, however, analyzed by using the resulting best networks in a similar study (unpublished data), provided predictions that were no better than chance, suggesting that these networks had in fact learned relevant generalized details of the real data set and, perhaps more importantly, that the data set was itself coherent enough to be generalizable. Moreover, although smaller or more accurate networks might be obtained by further adjustment of parameters, the most predictive networks here were quite accurate (> 95%), and smaller networks in which fewer input variable were used were, perhaps appropriately, less able to predict outcomes.

In addition, neural network analysis makes it more difficult than multivariate analyses to isolate the relative contributions of any individual input variable. Weights from any one input node contribute to all of the nodes in the next hidden layer, blurring the contribution to the outcome of that one variable. It might be possible to find a method for extracting the total weights, for example, from each input, at least to the first hidden layer. This might provide some indication of whether a particular input variable is always heavily added into the next layer (either positively or negatively). That information, however, becomes diffused with a second hidden layer and especially further diluted if there are feedback circuits in the network, as there were in many of the networks studied here.

CONCLUSIONS

In this study we have demonstrated the potential of using neural networks to analyze multivariate clinical outcomes after surgery for epilepsy. Based on widely accessible data predictive abilities of the best networks were above 95% accuracy. Further studies in which neural network approaches were used to facilitate operative planning and patient discussions appear promising. It is likely that if the database is large enough, individual institutions will be able to tailor networks for their own needs using their own databases. The results of the only other study in this area[12] support this view, as their findings were similar to those found here. However, research in this area could also focus on using a given network trained at one center on another center's data. Such cross-institutional prediction could perhaps then more fully suggest underlying bases for the prediction of outcome after epilepsy surgery.

Acknowledgments

The authors wish to thank Drs. Cornelia Santchi, Steven Pacia, Douglas Miller, and Suash Shamma.


Manuscript received December 7, 1998.

Accepted in final form February 2, 1999.

Address reprint requests to: J. E. Arle, M.D., Ph.D., Department of Neurosurgery, Hospital of the University of Pennsylvania, 3400 Spruce Street, Philadelphia, Pennsylvania 19104.