Intraoperative Decisionmaking with MER for STN DBS in PD and the Potential Relationship to Patient Selection

J.E. Arle*,1, J. Zani2 and J.L. Shils1

1Department of Neurosurgery, Lahey Clinic, 2Department of Neurology, Lahey Clinic, USA

Abstract: Microelectrode recording (MER) techniques generally require experience and nuanced technical ability. In order for appropriate decisions in electrode placement to be made during DBS procedures, multiple variables need to be differentially weighed including patient-to-patient variability and differences encountered during each MER recording track. Moreover, most often the appropriate weightings of such variables are not well known. As such, MER for DBS is typical of a class of decision making exercises involving multiple variables of unknown weight, common in surgical environments where Class I evidence is often lacking.

This study identifies 9 categories of information used by surgeon and neurophysiologists during DBS in the STN for Parkinson’s Disease, quantifies their occurrence, PPV, and NPV in 274 MER tracks across 75 consecutive patients and explores how they contribute to overall outcome in these patients. Although MER optimizes the ability to place the STN electrode accurately in these cases, it is not predictive in avoiding three categories of adverse outcome (gait and falling problems, worsened speech problems, or new psychiatric or cognitive problems). One conclusion as a result of this finding is the possibility that overall outcome, in terms of avoidance of unwanted side effects of DBS surgery, is unrelated to the decision making during surgery with MER, but is instead likely related to the ability to diagnose PD accurately at the outset.

Keywords: Intraoperative decisionmaking, MER, STN, Patient selection.

INTRODUCTION

Deep Brain Stimulation (DBS) in the subthalamic nucleus (STN) has become a standard treatment option for medically-refractory Parkinson’s Disease (PD) [1]. Results have been found to be equivalent or better, and more durable, than lesioning [2]. Moreover, the risk profile and cost basis of DBS surgery have both been favorably compared to medication management for PD [3, 4]. DBS in the STN, in particular, is generally performed with the adjunctive use of microelectrode recording (MER) to best locate the physiologic target in the STN, potentially minimizing power requirements for stimulation. MER techniques, however, generally require experienced and nuanced technical ability. In order to make appropriate decisions in electrode placement, multiple variables need to be differentially weighed—from patient-to-patient variability to differences encountered during each MER recording track—while most often the appropriate weightings of such variables are not well known. As such, MER for DBS is typical of a class of decision making exercises involving multiple variables of unknown weighting, common in surgical environments where Class I evidence is often lacking.

This paper seeks to identify the most appropriate variables contributing to decision making during MER for STN DBS, quantify them, and explore how they contribute to overall outcome in these patients. One interesting result of this study raises the possibility that overall outcome, in terms of avoidance of unwanted side effects of DBS surgery, is unrelated to the decision making during surgery with MER, but is instead likely related to the ability to diagnose PD accurately at the outset.

METHODS

Data was examined from 75 consecutive patients who underwent STN DBS for PD, encompassing a total of 274 MER tracks. These cases were all performed over a 9 year period involving the same surgeon (JEA) and neurophysiologist (JLS). The electrode implanted was also identical in all cases (Medtronic model 3387). One patient had irretrievable data, and one patient had only one of two sides included in the analysis as the second side was performed with a different neurophysiologist. All patients had undergone preliminary testing that involved a full neuropsychological work-up, videotaping and pre-operative UPDRS scoring, and full assessment by a movement disorders specialist in our program.

The analysis of MER tracks was made according to a set of guidelines that involved the following categories:

**Negative**

- AC-PC plane >5mm off image plane
- Vascular or Movement Artifact Complicating recordings
- Electrode or Equipment issues
- Patient exam or diagnosis concerns

*Address correspondence to this author at the Department of Neurosurgery, Lahey Clinic, Burlington, MA 01805, USA; Tel: 781-744-8644; Fax: 781-744-5228; E-mail: jeffrey.arle@lahey.org*
Positive

- STN >4mm clearly
- Normal thalamic cells and recording
- Clear Borders between zones everywhere
- >1 Kinesthetic cell appreciated
- Tremor Cells found

Such retrospectively analyzed categories were determined by evaluating prospectively acquired intraoperative physiology notes. Important aspects of grading each track are given in Table 1. Each category was given a value of 1 (present or good) or 0 (not present or bad). What is conceptually important in such a method is the understanding that in this analysis we are seeking the aspects of decision making for determining electrode placement, or potentially placing another recording track. As such, the confidence derived by the sureness of such data contributes most to the ability of the surgeon and physiologist to appropriately weight the variables, and thus optimize localization of the permanently implanted electrode. So, for example, unless the thalamus was ‘perfect’ in character and boundary (i.e. all borders were clearly defined), or the STN length was solidly >4.0 mm (without question marks in the cells or STN/SNr boundary concerns), then these were graded 0, and not 1. This, again, is because unless one has almost an absolute confidence in a variable, it contributes an unknown quantity to the decision process.

A Bayesian-type of decision-tree can thus be constructed using all of these variables at each decision node. In this paper, we considered the specific decisions at two points: either placing the electrode or making a second recording following the first track, or either placing the electrode or making a third or more recording (see Fig. 1). Ultimately, 2x2 matrices could be made for each variable as if independent from all others, and full assessment of prior probabilities could be calculated (see Table 2 for an example). Such data represents the potential prior probabilities of anyone performing STN DBS with MER, as no other data for these particular variables is known to exist.

All patients were diligently followed for a minimum of 2 years in this study post-operatively, programmed as optimally as possible, and had UPDRS measures followed up until and through the period when optimal programming had occurred. On the other hand, adverse outcomes (AO) were considered outcomes that adversely affected the patient while existing independently of the results on the cardinal features of medically-refractory PD (dyskinesia, bradykinesia, rigidity, and tremor). They included a high fall frequency or gait problems beyond their initial baseline function, wors-

Table 1. Classifiers for the MER Recording Tracts. Commentary Provides Rationale for Including the Category and what Role it Could Play in Affecting the Confidence of the Surgeon and Neurophysiologist for Targeting

<table>
<thead>
<tr>
<th>Item</th>
<th>Classified as a 1</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC-PC &lt; 85° relative to the vertical</td>
<td>If the AC and PC points were within two images of each other then this considered a 1.</td>
<td>The greater the Z difference in the AC and PC points the greater the potential for depth changes on the final target.</td>
</tr>
<tr>
<td>Thalamus</td>
<td>At least two single units that were greater than 0.5 mm apart</td>
<td>If only background and multi-units were noted then it was placed in this category. Also if no activity at all was noted it was categorized as a 0.</td>
</tr>
<tr>
<td>STN &gt; 4.0 mm of definite STN</td>
<td>STN/SNr questionable at times</td>
<td>If only some nuclei had definitive borders, then the track was classified as a 0. This criteria was chosen since we use borders as mapping points and lack of good borders increases potential error in mapping.</td>
</tr>
<tr>
<td>Borders</td>
<td>Good border in all nuclei</td>
<td>If no kinesthetic are noted in a tract then there is no information as to whether the electrode is in the sensory motor region</td>
</tr>
<tr>
<td>Kinesthetic</td>
<td>2 or more kinesthetic cells in the STN</td>
<td>No vascular artifact. Vascular artifact can potentially obscure kinesthetic and cell classification</td>
</tr>
<tr>
<td>Tremor Cells in the STN</td>
<td>If any ‘tremor cell’ was recorded in the STN.</td>
<td>In our institution the sound of the MER recordings is the most critical element in classification. Thus any equipment issues that reduce the quality of the recordings constitute a potential reduction in classification accuracy</td>
</tr>
<tr>
<td>Vascular Artifact</td>
<td>Any vascular artifact anywhere in the tract. This included pulse sounds and modulation of the single units in direct relation to the EKG.</td>
<td>All of these items can obscure the recordings of single units and thus reduce the confidence in single unit classifications</td>
</tr>
<tr>
<td>Poor Equipment</td>
<td>Auditory feedback, excessive (noise that obscures the signal) line noise, excessive electrical noise, low impedance electrodes.</td>
<td>In our institution the sound of the MER recordings is the most critical element in classification. Thus any equipment issues that reduce the quality of the recordings constitute a potential reduction in classification accuracy</td>
</tr>
<tr>
<td>Questionable Exam</td>
<td>Anesthesia usage, last medication within 12 hours, questionable patient responses, prior pallidotomies, previous brain surgeries, diagnosis</td>
<td>In our institution the sound of the MER recordings is the most critical element in classification. Thus any equipment issues that reduce the quality of the recordings constitute a potential reduction in classification accuracy</td>
</tr>
</tbody>
</table>
ened speech problems, and psychiatric or cognitive changes that were not seen prior to surgery. Such outcomes are thought to be relatively uncommon, difficult to treat, and of significant distress for the patient and family, potentially outweighing any benefits achieved in the 4 cardinal features of medically-refractory PD. We were specifically interested in knowing whether intraoperative variables per se, and decisions made intraoperatively in placing the electrode, were ultimately relatable to these adverse outcomes.

Table 2. Example of One of our 2x2 Contingency Tables for the Greater than 4 mm of STN Category. PPV = 17 / 50 = 0.34, NPV = 5 / 6 = 0.833. These were computed for every category for every track. Totals for all of these predictive values are given in Table 3

<table>
<thead>
<tr>
<th>Adverse Outcomes</th>
<th>No Adverse Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>17</td>
</tr>
<tr>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
</tr>
</tbody>
</table>

RESULTS

Overall improvements measured by Part III of the UPDRS were approximately 50%, and medication reduction in DA equivalents was approximately 40%. These basic results are similar to most large series of DBS STN for PD [5]. This is emphasized because we are concerned with analyzing MER decision making in detail, and its potential effects on outcome. As such, it is helpful that outcomes in our series as measured in a standard way (e.g. UPDRS) are at least equivalent to those found by other groups. The typical OR time for placement for a single side STN electrode was 60-90 minutes, 120-150 for bilateral electrode placement, and 45-60 minutes to place and connect both battery generators, which includes intubation, positioning, and redraping. The average number of MER passes (now in over 200 cases including other indications) was 1.6 per side and the average number of single discernible cells per track was 7-10. Fig. (2) shows every MER track move and its location relative to the original target. There were no symptomatic or fatal hemorrhages, but 4 asymptomatic (<2cm) hemorrhages in this series. Of the 274 total MER tracks considered, 48 resulted

Fig. (1). General decision making tree for MER in STN DBS. ‘I’ indicates a decision to implant the permanent electrode, while ‘NI’ indicates it was not implanted after that track. ‘AO’ abbreviates ‘adverse outcome’.

Fig. (2). All moves of subsequent MER tracks after first track whether implanting or not (note that recordings at x=0, y=0 are left out because all tracks started at this point). Moves generally clustered in the anterior and posterior directions, with a move of 2mm either way being the most frequent.
Table 3. Positive and Negative Predictive Values for 9 categories of MER findings in STN DBS. Highlighted are particularly high or low values – though they are all in the opposite direction to be useful. For example, not having a normal thalamus is predictive for not having an adverse outcome, as defined in the text.

<table>
<thead>
<tr>
<th>Category</th>
<th>PPV-1</th>
<th>PPV-2</th>
<th>NPV-1</th>
<th>NPV-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC-PC&lt;85°</td>
<td>0.333</td>
<td>0.316</td>
<td>0.681</td>
<td>0.623</td>
</tr>
<tr>
<td>Normal Thalamus</td>
<td>0.395</td>
<td>0.396</td>
<td>0.923</td>
<td>0.708</td>
</tr>
<tr>
<td>Length of Recorded STN &gt;4</td>
<td>0.340</td>
<td>0.460</td>
<td>0.833</td>
<td>0.864</td>
</tr>
<tr>
<td>Nuclear Borders</td>
<td>0.286</td>
<td>0.357</td>
<td>0.657</td>
<td>0.636</td>
</tr>
<tr>
<td>&gt;2 Kinesthetic cells in the STN</td>
<td>0.318</td>
<td>0.456</td>
<td>0.667</td>
<td>0.808</td>
</tr>
<tr>
<td>Tremor Cells in the STN</td>
<td>0.333</td>
<td>0.200</td>
<td>0.679</td>
<td>0.627</td>
</tr>
<tr>
<td>Vascular and Movement Artifact</td>
<td>0.600</td>
<td>0.333</td>
<td>0.706</td>
<td>0.638</td>
</tr>
<tr>
<td>Equipment Failures</td>
<td>0.182</td>
<td>0.286</td>
<td>0.644</td>
<td>0.621</td>
</tr>
<tr>
<td>Pt issues</td>
<td>0.222</td>
<td>0.500</td>
<td>0.660</td>
<td>0.652</td>
</tr>
</tbody>
</table>

in an AO ultimately in the patient. In fact, a total of 25 of the 75 (33%) patients had an AO in the sense described above in the Methods section.

Analysis of the MER tracks themselves, specifically in terms of the categories outlined in the Methods section, and calculating their relationships to these AOs, yielded positive and negative predictive values as found in Table 3. As can be appreciated from Table 3, there were no positive or negative predictive values that correlated with appropriate positive or negative categories. Curiously, in fact, there were findings such as high NPVs for first or second tracks with STN>4mm. This means that if there was <4mm STN, then there was likely not to be an AO. Such unexpected correlations in predictability were found because although there were AOs found in 33% of the patients, only 48 of 274 MER tracks were associated to a patient with an AO and most tracks had STN>4mm. As such, the calculations in Bayesian 2x2 contingency tables can make some ‘predictive’ values quite unexpected.

DISCUSSION

The analysis of MER in these DBS cases has shown not only the ability of MER to help in guiding optimal placement of the electrode for treating the cardinal features of medically-refractory PD (bradykinesia, rigidity, tremor, and dyskinesia), but also that important characteristics of the intraoperative decision making using MER do not predict whether a given patient will have other adverse outcomes in terms of gait disturbances and fall problems, speech difficulties beyond their pre-operative, or psychiatric or cognitive worsening. While some might consider this finding of little import, we posit the opposite, in that the findings of MER specifically characterize the local physiology of the patient and physiologically-optimized placement of the electrode for therapeutic benefit. In our experience, the electrode cannot be placed more accurately in the sensory-motor STN than by using MER in conjunction with imaging and sound surgical technique. As such, there is no other way to account for the adverse outcomes in these patients than by making the assumption, based partly on this data, that the only possibility for why at least a large portion of these patients have adverse outcomes is that they are not all clearly idiopathic PD patients. (An alternative we will not consider in this account is that the circuitry between PD patients is significantly different from one to another such that even though they might all have idiopathic PD, they still may respond to electrical stimulation in the same anatomic location in different ways.)

What is the evidence that idiopathic PD could be an unreliable diagnosis, even in the most experienced hands? In this study, for example, we had every patient fully evaluated and managed by fellowship-trained Movement Disorders specialists, which is similar to what occurs in many centers. Reviewing available literature on this subject reveals that up to 35% of patients are misdiagnosed – even in the hands of specialists – this number ranges from about 8%-35% depending on the study. An excellent review of this problem can be found in Tolosa et al. [6]. A prospective study showed that 65% are diagnosed correctly after 5 years of symptoms and only 76% are correct after 12 years of symptoms [7]. Despite rigid inclusion criteria uniformly applied, and improved accuracy in diagnosis achieved, review from the UK brain bank data found 10% of PD diagnoses from life would have needed to be revised after post-mortem examination.

Alternative diagnoses that may confound diagnostic accuracy include Diffuse Lewy Body Disease (DLB), vascular Parkinsonism, Progressive Supranuclear Palsy (PSP), Multi-system Atrophy (MSA), Corticobasal degeneration (CBD), Essential Tremor (ET), and drug-induced Parkinsonism. The possibility of substantial overlap of symptom manifestations is high, and most tests of differentiation have been shown to have little discriminating value [8]. Moreover, most potential tests have not been verified by post-mortem validation. The Dopamine (DA) challenge test may be positive in MSA (20%), and its negative predictive value in de novo patients is only between 40-60% [9, 10]. Various clinical neurophysiologic measures have shown little predictive ability, such as long latency reflex times [11-13], heart rate variabili-
Autonomic function testing and urodynamic studies are somewhat useful in separating MSA from PD [21], but there is still substantial overlap, especially further along in the disease processes. Olfactory disorders affect 90% of PD patients while they are noticeable in many fewer MSA, PSP, and vascular parkinsonism patients [24, 25] and thus evaluation of olfactory function may hold some promise. With imaging, CT has been unhelpful and is usually normal. MRI, particularly DWI, has shown promise particularly with regional diffusion coefficients in the putamen and 3-D DWI might be more specific and sensitive in early disease states [26-28]. SPECT, particularly DAT-SPECT, using the DaTSCAN in clinically uncertain Parkinsonian syndromes study groups demonstrated initial agreement between clinical diagnosis and imaging in 69 of 77 (90%) of patients while in 8 additional patients a diagnosis was inconclusive [29]. Out of the 14 patients with inconsistent results or disagreement, later scans were able confirm diagnosis in 11 of them (2 dropped out), thus leaving an 11/16 diagnostic accuracy in this small study. MiIBG-SPECT has also shown potential in differentiating PD from MSA in a study by Druschky et al., where 30 patients (10 PD and 20 MSA) were investigated [30]. In this study MiIBG uptake was significantly lower in MSA patients as compared to PD patients and was also significantly lower in both PD and MSA patients as compared to normals [30]. 18F-DOPA PET has become widely believed to be a reasonable surrogate marker of PD [31, 32], where 30 patients (10 PD and 20 MSA) were investigated [33]. In this study, MiIBG uptake was significantly lower in MSA patients as compared to PD patients and was also significantly lower in both PD and MSA patients as compared to normals [30]. 18F-DOPA PET has become widely believed to be a reasonable surrogate marker of PD [31, 32]. In this study, MiIBG uptake was significantly lower in MSA patients as compared to PD patients and was also significantly lower in both PD and MSA patients as compared to normals [30]. 18F-DOPA PET has become widely believed to be a reasonable surrogate marker of PD [31, 32]

Decision making in surgery is an important field of study using both interviews [37] and artificial neural networks [38], as even a simple Medline search of “Surgery”, “Decision making” and “Decision Making” found 2140 results. Our analysis of using informational probabilities during MER for performing STN DBS in PD has shown that although many of the posterior probabilities can be refined reliably, thus making a Bayesian predictive approach feasible, even the ability to know such data in real time and execute the technique appropriately does not necessarily prevent adverse findings or guarantee success. Clearly, there are conditions beyond the scope of control of the individual surgeon in any given procedure. Typically, posterior odds/probabilities are not actually known, data used in analysis is not always consistently accurate, and the context of data interpretation by practitioners varies (prior expectations, mental/emotional state, bias/agenda, constraints of resources, differences in valuation or risk aversion). How much each of factor contributes to ‘correct’ decisions, and how reliable is that amount? Perhaps ‘time’ artificially allows for multiple decision points, which give many more opportunities to change a decision if necessary before consequences are realized. Perhaps, a database and algorithm can provide real-time feedback to guide decisions. One illustrative example of this was from a study by deDombal et al. [39] using a computer program to synthesize data and decide on diagnoses for acute abdominal pain. With data on over 700 prior patients and studying prospectively 552 patients, probabilities were generated for all alternative diagnoses. Predictions for each case were compared with attending physicians’ diagnoses. Physicians were correct between 42 and 81 percent of the time – the computer was correct in 91 percent of cases. Rates of appendix perforation dropped dramatically, as did false positive appendectomy surgery. Interestingly, the frequency of erroneous decisions rose again after the program ended.

In conclusion, we determined reasonable estimates of posterior probabilities for many characteristics in MER tracks during STN DBS and their ability to provide PPV or NPV to the occurrence of certain adverse outcomes from the procedure, namely balance/gait/fall problems, speech difficulties, and new or exaggerated psychiatric or cognitive worsening, without the compromise of benefit to the typical features of PD. We suggest from this that intraoperative MER-guided decisions have little ability to predict adverse outcomes, or guarantee success, other than to assure us that the electrode is as optimally positioned in the STN as possible. Decision making in surgery in general, and in DBS surgery specifically, is still underanalyzed, and inquiry at present raises more questions than answers - can it be improved? More importantly, however, is the suggestion from this work that patient selection in DBS for PD seems to remain problematic and perhaps has more pertinence than ever.

REFERENCES


