Published in final edited form as:

Brain Stimul. 2015 ; 8(6): 1025–1032. doi:10.1016/j.brs.2015.06.003.

Machine learning approach to optimizing combined stimulation and medication therapies for Parkinson's disease

Reuben R. Shamir¹, Trygve Dolber¹, Angela M. Noecker¹, Benjamin L. Walter^{2,3}, and Cameron C. McIntyre^{1,2}

¹Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA

²Department of Neurology, Case Western Reserve University, Cleveland, OH, USA

³Neurological Institute, University Hospitals Case Medical Center, Cleveland, OH, USA

Abstract

Background—Deep brain stimulation (DBS) of the subthalamic region is an established therapy for advanced Parkinson's disease (PD). However, patients often require time-intensive postoperative management to balance their coupled stimulation and medication treatments. Given the large and complex parameter space associated with this task, we propose that clinical decision support systems (CDSS) based on machine learning algorithms could assist in treatment optimization.

Objective—Develop a proof-of-concept implementation of a CDSS that incorporates patient-specific details on both stimulation and medication.

Methods—Clinical data from 10 patients, and 89 post-DBS surgery visits, were used to create a prototype CDSS. The system was designed to provide three key functions: 1) information retrieval; 2) visualization of treatment, and; 3) recommendation on expected effective stimulation and drug dosages, based on three machine learning methods that included support vector machines, Naïve Bayes, and random forest.

Results—Measures of medication dosages, time factors, and symptom-specific preoperative response to levodopa were significantly correlated with postoperative outcomes ($p < 0.05$) and their effect on outcomes was of similar magnitude to that of DBS. Using those results, the combined machine learning algorithms were able to accurately predict 86% (12/14) of the motor improvement scores at one year after surgery.

Corresponding Author: Cameron C. McIntyre, Ph.D., Department of Biomedical Engineering, Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH 44106, ccm4@case.edu, Phone: +1-216-368-5869.

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.

Conflicts of Interest:

CCM authored intellectual property related to this article, is a paid consultant for Boston Scientific Neuromodulation, and is a shareholder in the following companies: Surgical Information Sciences, Inc.; Autonomic Technologies, Inc.; Cardionomic, Inc.; Enspire DBS, Inc.; Neuros Medical, Inc.

Conclusions—Using patient-specific details, an appropriately parameterized CDSS could help select theoretically optimal DBS parameter settings and medication dosages that have potential to improve the clinical management of PD patients.

Keywords

Parkinson's disease; deep brain stimulation; clinical decision support system

1. Introduction

Deep brain stimulation (DBS) of the subthalamic region is an established treatment for management of the motor symptoms of advanced Parkinson's disease (PD) (Obeso *et al.*, 2001; Weaver *et al.*, 2009). Following the surgery, the neurologist is faced with the challenge of balancing both the stimulation and medication therapies for the patient to maximize benefit and minimize side effects. This complex process is currently driven by clinical experience and typically incorporates guidelines developed from previous clinical studies (Bronstein *et al.*, 2011; Castrioto *et al.*, 2013). However, treatment optimization often requires multiple time-consuming follow-up visits because of the extremely large treatment parameter space. In addition, DBS can be associated with side effects generated by the unwanted spread of stimulation to non-target regions, which is dependent on the patient-specific location of the electrode(s) in the brain (Butson *et al.*, 2011).

The relationship between subthalamic DBS electrode locations and treatment outcomes has been extensively studied (e.g. Maks *et al.*, 2009; Welter *et al.*, 2014; Eisenstein *et al.*, 2014). These kinds of results, coupled with patient-specific imaging data and electrical stimulation models, have been used to create clinical decision support systems (CDSS) designed to help customize DBS parameter settings to the patient (McIntyre *et al.*, 2007; Frankemolle *et al.*, 2010). Further, the first generation of commercial DBS CDSS are now available in Europe (e.g. Optivise by Medtronic (MN, USA) or GUIDE DBS by Boston Scientific (MA, USA)). However, current CDSS platforms only provide guidance on electrical stimulation, ignoring the pharmacological side of PD patient management. Therefore, we set out to define the foundation for a more comprehensive CDSS that couples stimulation and medication, while also considering the patient-specific symptomology (i.e. tremor, rigidity, bradykinesia) (Fig. 1). Our proposed CDSS extends the typical clinical database functionality, as well as the current CDSS systems that visualize stimulation volumes, with three additional functions: 1) retrieval of similar patient datasets with respect to symptoms, time factors (i.e. age, time since surgery), medications, and stimulation setups; 2) visualization of past patient outcomes, and; 3) recommendation on expected effective treatments according to machine learning methods. Most importantly, the database and the CDSS can be continuously updated with retrospective and/or prospective data to improve the CDSS prediction accuracy over a large spectrum of patients.

2. Methods

2.1 Methods overview

We implemented a prototype CDSS for combined stimulation and medication treatment, using a clinical dataset of 10 PD patients (89 postoperative visits) who underwent bilateral DBS electrode placement in the subthalamic region. Our goal was to examine if incorporation of patient-specific symptoms and medications into a machine learning algorithm would better predict the treatment outcomes in comparison to considering the stimulation parameter settings alone. Our underlying assumption was that patients with similar symptoms that are given similar treatments would result in similar outcomes when the treatment parameters and similarity measures are appropriately defined.

Our first step was to create patient-specific DBS computer models that documented the anatomical location of the DBS electrodes in the patient brain, as well as the stimulation volume generated by their clinically defined stimulation parameter settings. This process included integrating MRI and CT data with intraoperative microelectrode recording data to characterize the patient anatomy. Each patient's anatomical model was then coupled with an electrical model that estimated their volume of tissue activated (VTA). This enabled correlation analyses between stimulation variables and clinical outcomes.

The next step was to develop a metric that measured the similarity between two patients with respect to their PD symptoms, clinical history, levodopa equivalent daily dosage (LEDD), and its delivery timing. These results were then be coupled with the overlap between the VTA and a predefined therapeutic target volume. The resulting similarity metric facilitated retrieval of the most relevant previous office-visit charts from multiple patients to provide a retrospective reference for visualization (Fig. 2).

The final step was to use the information available in the database to create a recommender system. Our goal was to use computer optimization to predict the most promising treatment strategy for a given patient, incorporating the key factors related to therapeutic outcomes. We relied on machine learning methods to achieve this goal (Fig. 3). The following methods sections (and supplementary material) describe our dataset, the similarity measure, the extraction of the relevant features, and our recommender system.

2.2 DBS patient data

Initially we identified a collection of 78 PD patients who underwent DBS surgery at University Hospitals between 2008–2013 (UH; Cleveland, OH). Of them, we were able to retrieve office visit charts from 53 patients that remained active in the UH neurology and neurosurgery departments. Of the 53 patients, 12 were excluded because postoperative follow-up was less than nine months; 8 were excluded because their DBS was unilateral; 13 were excluded because of an image related issue (i.e. missing image, large brain shift, or poor registration); 10 were excluded because of inconclusive or missing clinical notes. Therefore, data from the 10 remaining patients was incorporated in this study (Table 1).

The study was approved by the Institutional Review Board of University Hospitals and Case Western Reserve University School of Medicine. All patients met accepted selection criteria

for DBS and signed informed consent for the surgery. The third subsection (motor score) of the unified PD rating scale (UPDRS-III) was assessed preoperatively both off (>12 hours) and on dopaminergic medication (~1 hour after medication administration). Postoperatively, UPDRS-III was assessed at a total of 89 follow-up visits of the 10 patients, each visit under one of the following four setups: 1) on-meds on-stim; 2) on-meds off-stim; 3) off-meds on-stim, or; 4) off-meds off-stim). The relative improvement of motor symptoms on-medication in the preoperative state, and on/off-medication on/off-stimulation in the postoperative state were defined as follows to avoid false correlations that may arise using the non-normalized UPDRS scores (Zaidel *et al.*, 2010):

$$100 \times \frac{PRE_{off} - PRE_{on}}{PRE_{off}} \text{ and } 100 \times \frac{PRE_{off} - POST_{comb}}{PRE_{off}}, \quad (1)$$

respectively, where PRE_{off} , PRE_{on} , and $POST_{comb}$ represent the UPDRS-III score preoperative off-medication, preoperative on-medication and a postoperative combination of on/off-medication and on/off-stimulation, respectively. To compare specific symptom relative improvement (Eq. 1) the following subsections of the UPDRS-III (motor) section were investigated as well. 1) speech (section 18; max 4); 2) tremor (sections 20–21; max 28); 3) rigidity (section 22; max 20); 4) limb bradykinesia (sections 23–26; max 32), and; 6) axial akinetic symptoms (sections 19 and 27–31; max 24). When some of the sections of a specific symptom were missing from the patient record, their values were estimated by a linear interpolation. When all of the sections in a specific sub-score were missing, the entire office visit was excluded.

Medication—Levodopa equivalent daily dosage (LEDD) was computed from each patient's medications records as suggested by Tomlinson *et al.* (2010) and the relative change in LEDD was defined as follows:

$$100 \times \frac{LEDD_{pre} - LEDD_{post}}{LEDD_{pre}} \quad (2)$$

where $LEDD_{pre}$ and $LEDD_{post}$ represent the LEDD before and after surgery, respectively. The time since last PD medication intake was also recorded.

Deep brain stimulation—Each brain hemisphere of each DBS patient has a unique electrode placement relative to their neuroanatomy, as well as a unique stimulation parameter setting selected from a large number of possible options. Detailed patient-specific DBS models exist to account for these variables between patients. In the context of this study, we propose that a single common metric is also needed to compare patients and simplify analysis. We elected to use the overlap of volume of tissue activated (VTA) for the specific stimulation parameter setting used in the patient with a *target zone*. The target zone was defined relative to the Harvard-Oxford brain atlas (Desikan *et al.*, 2006) with representation of the subthalamic nucleus (Keuken *et al.*, 2013). The target zone location was empirically derived based on the results of our earlier studies (Butson *et al.*, 2011;

Frankemolle *et al.*, 2010). Our target zone was defined as an ellipsoid that intersects the dorsal aspect of the STN, as well as the Zona-Incerta, which are both considered effective stimulation targets for the management of the motor symptoms of advanced PD (Plaha *et al.*, 2006; Welter *et al.*, 2014). The following steps were followed for each patient to estimate the VTA with respect to the target zone (Fig. 4).

1) Atlas to preoperative MRI registration: The midline, anterior commissure (AC), and posterior commissure (PC) were identified on the patient's preoperative MRI, as well as the atlas image. A rigid transformation was then computed to match the defined AC/PC and midline axes of the atlas with the patient MRI (Miocinovic *et al.*, 2007). Then, the atlas anatomical volumes, including the target zone, were overlaid on the patient MRI and manually fitted to the anatomy using a 3D affine transformation (e.g. 3 translation, 3 rotation and 3 scaling parameters).

2) Atlas refinement with microelectrode recordings: Each patient-specific model was constructed within the context of the stereotactic frame coordinate system (Fig. 4). This allowed for incorporation of intraoperative microelectrode recording (MER) data defining the location of STN neurons (Maks *et al.*, 2009). When the atlas volumes, fitted to the anatomy alone, and the MER data points did not align exactly, we attributed this discrepancy to brain shift (Roberts *et al.*, 1998). In such cases we performed linear translation of the atlas volumes to better correspond with the MER data.

3) Electrode position: The location of the DBS electrode was extracted from a postoperative CT acquired a few hours after surgery. This estimate of the electrode position was compared to the surgically intended location in stereotactic coordinates. When the distances between the centers of these two estimated electrodes locations were below 2mm for both hemispheres we considered them consistent and included their associated results in our study. Information from 10 patients for which the electrode locations were consistent is presented at Table 1. Postoperative CT images that were acquired several weeks after the surgery were available for two of the patients (patients 3 and 5). For these patients, the final electrode position used in their DBS model was computed based on that image.

4) Volume of tissue activated: The VTA was computed based on the stimulation settings documented in each postoperative follow-up visit. The VTA is a metric to estimate the spatial extent of axonal activation generated by DBS for a given parameter setting. The specific methods for VTA calculation used in this study have been previously described in detail (Chaturvedi *et al.*, 2013). Briefly, an artificial neural network (ANN) was trained to match results from detailed finite element DBS electric field models coupled to multi-compartment cable models of myelinated axons (a.k.a field-neuron models). The DBS electric field models assumed an isotropic tissue medium and a high resistance encapsulation sheath surrounding the DBS electrode. The axon models assumed a fiber diameter of 5.7 μm . The ANN VTA predictions were derived from models that have been validated in both human (Chaturvedi *et al.*, 2010) and monkey (Miocinovic *et al.*, 2009) experiments; however, they do represent a relatively simplistic estimation of stimulus spread. The

advantage of using the ANN VTA predictor, over the traditional field-neuron modeling approach, is that computational resources are minimal and results are generated instantly.

5) VTA overlap with target zone: The overlap between the VTA and the target zone was computed and its percentage from the total volume of the target was defined as follows:

$$100 \times \frac{|\text{target_zone} \cap \text{VTA}|}{|\text{target_zone}|} \quad (3)$$

The average of the right and left overlap volumes was computed to represent a single measure for each clinical visit of each patient. This simplified single metric was then used for defining our correlations.

Typical MRI size was $256 \times 256 \times 190$ with voxel size of $1 \times 1 \times 1 \text{ mm}^3$. Typical CT image size was $512 \times 512 \times 40$ with voxel size of $0.36 \times 0.36 \times 2.4 \text{ mm}^3$. 3D-Slicer (Fedorov *et al.*, 2012) was used for the MRI-CT and CT-CT registrations. Cicerone (Miocinovic *et al.*, 2007) was used for the atlas/MRI fitting, electrode fitting, as well as computation of VTAs and their overlap with the target zone.

2.3 Retrieval of similar patient data

To retrieve data from similar patients we first normalized selected features by computing their Z-score over all postoperative visits as follows:

$$z_{pi} = \frac{(x_{pi} - \mu(x_{pi}))}{\sigma(x_{pi})} \quad (4)$$

where the index '*i*' enumerates the features (e.g. improvement in tremor, rigidity, etc.) and the index '*p*' enumerates the different postoperative office visits. x_{pi} is a vector of selected feature values, and $\mu(x_{pi})$ and $\sigma(x_{pi})$ are the mean and standard deviation of x_{pi} , respectively. Then, a normalized '*signature*' vector $Z_p = (z_{p1} \dots z_{pn})$ was defined for each of the 89 postoperative visits of the 10 PD patients. The similarity between two postoperative visits was then defined as the root mean square (RMS) between their computed signatures as follows:

$$\text{similarity}(Z_p, Z_q) = \sqrt{\frac{\sum_{i=1}^n (z_{pi} - z_{qi})^2}{n}} \quad (5)$$

Conceptually, this type of similarity measure could be used in a CDSS by identifying previous example patients with common features and successful outcomes. When a new patient would visit the physician after DBS surgery, the CDSS would compute the patient's signature Z-vector. Then, all previous postoperative visits in the database could be sorted in order of their similarity.

2.3 Recommender system

The concept we propose for a recommender system is that the physician would input the known parameters from the patient into the CDSS that then loops over thousands of possible treatment parameters, estimating an outcome for each iteration (Fig. 3). After covering a broad range of treatment options, the system would identify clusters of treatment parameters that are associated with good outcomes and provide a recommendation for an optimal treatment (Fig. 5). Clustering effective treatment predictions is expected to overcome some of the errors that can be associated with using single prediction methods.

Selecting the specific prediction methods are of great importance to the success of the recommender system. Following an investigation of correlation and linear regression measures with the actual motor outcomes, we tested several machine learning methods for the prediction of treatment outcomes. The details of the correlation and linear regression methods, as well as their corresponding results are reported in the supplementary material (Figs. S1 and S2). The general results of these analyses were: 1) postoperative improvement of specific symptoms was correlated with specific subsets of measures; 2) the overlap of the VTA and the target zone was significantly correlated with the motor improvement, and; 3) combining motor outcome measures and patient-specific stimulation measures together with a linear formula generated a metric that was highly correlated ($r = 0.73$; $p < 0.001$) with the clinical outcomes:

$$-11.2m + 6.5r - 6.2h + 2.5o + 50.4 \approx \text{motor improvement} \quad (6)$$

where ' m ' denotes months since surgery, ' r ' denotes rigidity preoperative improvement with levodopa, ' h ' denotes hours since last medication dosage, and ' o ' denotes average right/left overlap of VTA and target zone. The coefficient of determination was $R^2 = 0.53$. It indicates that equation 6 predicts 53% of the variance in the variable *motor improvement*.

We then conducted a retrospective experiment to test the ability of the system to accurately classify clinical outcomes. Fourteen postoperative visits performed approximately 1 year following DBS surgery were selected as test cases. In each test case, the system was used to predict the motor improvement generated by the specific input parameters (patient data, medication data, and stimulation data). We followed a leave-one-out approach where all of the clinical information in the database was incorporated to train a machine learning classifier, except the specific visit being evaluated, as well as the visits that followed it, for that specific patient. This simulates a typical postoperative office visit where earlier office-visit records of the same patient are available in addition to those of other patients.

The classifier inputs are 'fixed parameters' that are set by the caregiver and the outputs are 'treatment parameters' that should be optimized to improve outcome (Fig. 3). In our test, the fixed parameters were age, months since surgery, preoperative motor response to levodopa (e.g. UPDRS-III and its subscales), the preoperative levodopa daily dosage, and VTA/target zone overlap (assuming that DBS therapy had already been optimized). The treatment parameters were the postoperative levodopa equivalent daily dosage and the medication administration times.

The UPDRS-III outcomes from all of the postoperative visits were classified into three categories: 1) non-responsive – less than 35% improvement after combined DBS and medication therapy with respect to the preoperative off medication state; 2) moderate response – 35 to 65% improvement, and; 3) high response – greater than 65% improvement. Since pre-operative improvement of 30% or more from levodopa alone was an inclusion criterion for the DBS surgery, an improvement of 35% or less under the combined treatment presents insignificant benefit to the patient. Note that our classification is by post-operative visit and not by patient. This is an important point as the same patient may be associated with different symptoms during the day or over longer periods of time depending on the specific treatment administered to the patient, the timing of medications, and disease progression, among other factors.

The classifier was then used to predict the expected improvement in the 14 test cases: non-responsive (<35%), moderate response (35%–65%), or high response (>65%). The actual clinical outcomes measured for each of the 14 test cases and the algorithm predicted outcomes were compared. Three common machine learning methods were evaluated: 1) Naïve Bayes classifier (NB; uniform distribution of priors); 2) Support Vector Machine (SVM; with Gaussian radial basis function kernel and least squares method for finding the separating hyperplane), and; 3) Random Forest (RF; with ensemble of 50 decision trees). Considering only the overall motor score may not account for the variability among patients as two patients with similar overall motor score may have very different symptoms. Therefore, we extended the above machine learning methods to predict each of the sub-symptoms outcomes and then aggregate the results into one prediction. We refer to this method as the symptoms aggregate (SA) prediction.

Any machine learning classifier is subject to errors that are related to the assumptions of the method and to the variability of the data. However, if multiple methods generate the same classification, it is likely that some reasonable evidence exists in the dataset to support the classification. Therefore, we estimated the accuracy of the three machine learning methods for consistent predictions. All of these machine learning analyses were performed using Matlab (version R2012b, MathWorks, MA, USA) on a standard PC.

3. Results

Reading and extracting the relevant information from patient charts can be a cumbersome and a time consuming task. Therefore, we suggest presenting a visual summary of the most relevant information available in the database (Fig. 2). A hypothetical new patient with median feature values of age (64), months since surgery (12), preoperative motor response to levodopa (61%), the preoperative levodopa daily dosage (1660 mg), and VTA/target zone overlap (20%) was selected to illustrate this concept. Figure 2 shows 2D projections of data from previous patients in the database with various levels of similarity to the hypothetical new patient. The results suggests that similar patients responded well with low LEDD of 300–400 mg/day (in combination with DBS) compared to higher LEDD values, and that motor improvement lasted about 4 hours for each dose.

The goal of our CDSS was to employ machine learning strategies to prospectively identify mathematical relationships that quantify the general trends in Fig. 2, or what experienced clinicians know based on years of training. Therefore, we tested various machine learning algorithms in their ability to make predictions on the motor outcome of a combined stimulation and medication treatment plan in specific patients. Table 2 summarizes the observed prediction accuracies for the different machine learning algorithms. Since there are three categories of classification (e.g. non-responsive, moderate response, and high response) a random guess of the outcomes would result in an accuracy of 33%. The direct prediction accuracies for the overall motor outcomes were 71% (10/14), 64% (9/14) and 64% (9/14) using the SVM, Naïve Bayes, and RF methods, respectively. The three direct-prediction algorithms agreed on the outcomes of 8 out of the 14 postoperative visits. The accuracy in this subset of test cases was 88% (7/8).

Higher accuracies for the direct-prediction algorithms were observed when only considering specific symptoms. For example, tremor and speech outcomes were predicted with accuracy of 100% (14/14) and 93% (13/14) respectively, with the SVM method. Limb bradykinesia and axial akinetic symptoms were predicted with accuracy of 86% (12/14) using the RF classifier.

Predicting the overall motor outcomes as a weighted sum of the individual symptom predictions using the best method for each symptom (systems aggregate (SA) method) improved the prediction accuracy to 86% (12/14). However, SA prediction accuracy was only 64% when the same-patient records from their earlier clinical visits were excluded. It was further reduced to 43% when only five patients were incorporated instead of nine. These results highlight the need for a large dataset to facilitate the machine learning algorithms ability to make accurate predictions.

Machine learning methods provide the ability to estimate the outcome of a singular set of inputs. This general tool becomes useful when put into the context of an overall CDSS (Fig. 3). The concept is to use the classification tool to broadly sample the complex parameter space of different medication and DBS dosages via extensive simulation, prior to any direct clinical testing on the patient (Fig. 3). Figure 5 (middle column) presents the classification results from this type of parameter space sampling for the hypothetical new patient described in the first paragraph of the Results section. In this example, thousands of different input parameter sets were automatically defined and subsequently classified by the algorithm (Fig. 5, upper row). The convex hull of the treatment parameters that generated a high response prediction were computed and its center provides the recommended output of the CDSS (Fig. 5, lower row).

The classification method also facilitates investigation of on the mutual interaction between stimulation and medication (Fig. 5, columns). Simulating various overlap values between the VTA and target zone alters the predicted optimal medication therapy. The example in Figure 5 shows that enhancing the overlap between the VTA and the target was associated with a wider range of effective medication treatments.

4. Discussion

The concepts of using advanced computational systems to assist in complex clinical care decisions have existed since the 1970s (Schwartz et al., 1970). However, clinical adoption and application of these ideas have been far slower than many expected, with reasons that range from limited utility to lack of computational power. Yet, in today's clinical world of electronic medical records and cloud computing resources, the limitations associated with clinical access to computational tools no longer apply. Therefore, the onerous is on the computer science world to identify and develop applications with realistic potential for actual clinical utility. In the specific world of deep brain stimulation (DBS), we identified the time consuming task of balancing stimulation and medication dosages as a clinical challenge that could benefit from computational assistance. However, achieving that long-term goal first requires the development of the database infrastructure and initial algorithms that a clinical decision support system (CDSS) could eventually be based upon. This paper presents a first-generation proof-of-concept demonstration of how such a system could be created.

Our results suggest that the outcomes of a combined pharmacologic-DBS treatment can be predicted relatively well, especially considering that data from a very small number of patients was used to build the prototype prediction algorithms. This provides confidence that a robust and clinically viable CDSS could be adequately trained and parameterized using high quality data from recently completed large-scale DBS clinical trials (e.g. Schuepbach et al., 2013). One interestingly finding from our analysis was that the prediction accuracy varied among the different symptoms. For example, compare the 57% and 100% accuracies that were observed for predicting rigidity and tremor outcomes, respectively (Table 2). One possible explanation for this difference is that some symptoms are measured more precisely in comparison to others, and/or that the inter- and intra-observer variability may be different for different symptoms. Therefore, we propose that incorporation of quantitative measurement devices for specific symptoms could play an important role in improving the prediction accuracy and efficacy of any forthcoming CDSS. Thankfully, quantitative measurement systems customized for PD symptom monitoring are becoming readily available as smart phone applications and/or wearable sensors (e.g. Daneault et al., 2013; Mera et al., 2013).

Our multi-parameter correlation analysis reveals important possible relations between specific PD symptoms and information that is available at time of the postoperative office visits. For example, older PD patients were associated with less improvement in limb bradykinesia in comparison to younger patients, consistent with the positive correlation previously noted between the patient's age and severity of bradykinesia (Lee et al. 2011). The axial-akinetic symptoms worsened the most as time passed after the surgery, also in line with previous reports (Castrioto et al. 2011). In addition, the total UPDRS III preoperative motor response to levodopa was not correlated with the postoperative outcomes when the data was appropriately normalized, as reported in previous studies (Zaidel et al. 2010). However, preoperative responses of specific symptoms (i.e. tremor and limb bradykinesia) to levodopa were associated with postoperative improvements. This result may be explained by the hypothesis that various sub-types of Parkinson's diseases are associated with different

spatial characteristics or temporal progression of pathology (Thenganatt et al. 2014). If true, this suggests that a differential symptom-specific evaluation of the outcomes would be the most accurate approach for developing future therapy prediction algorithms.

Our regression analysis suggests that both stimulation and medication parameters have a similar magnitude of importance to the measured overall motor outcomes (Eq. 6). The correlation analyses demonstrated that the postoperative reduction in levodopa equivalent daily dosage (LEDD), as well as the timing of medication administration were significantly related to clinical benefit. In addition, the overlap of the stimulation volume with the target zone was significantly correlated with the motor outcomes. In turn, the regression analysis resulted in a linear formula that had a higher correlation with the actual motor improvement than any single measure. These results stress the importance of integrating both medication and stimulation variables together when managing late stage PD treatment.

In the current study, only 10 out of 78 potential patients were included in our analysis because of limitations in the current clinical data collection and/or archiving processes. This illustrates the need for a prospectively designed clinical database system, customized to the management of movement disorders patients, which is capable of handling the diverse datasets necessary for a CDSS to function. Such a system would need to interface with multiple caregivers associated with multiple departments (radiology, neurosurgery, neurology) to ensure completeness. Obviously, this is no simple task, requiring significant institutional investment. However, we propose that such an investment is justified by the expense and effort associated with using DBS technology in clinical practice. Academic database systems of this type are currently under development at numerous DBS centers, including our own, and we foresee these tools evolving hand-in-hand with CDSS technology.

Our goal with this general line of research is to develop computational tools that help balance the DBS and medication therapies to optimize control of PD symptoms (Fig. 5). The presented recommender system, or future variants based on this concept, have potential to facilitate the selection of effective patient-specific treatment strategies. For example, our results suggest that larger VTA/target overlap is associated with increased medication efficacy time. This preliminary finding motivates further investigation on the mutual effects of DBS and medication on motor outcomes. In turn, identification of these kinds of new hypotheses represent one of the many benefits of approaching PD clinical management with database tools and machine learning assistance.

While our prototype efforts at PD CDSS development appear promising, this study suffers from a number of limitations that should be noted. First, the data used to construct the model was all retrospective and derived from a small number of unblinded measures acquired from DBS patients whose care was already being optimized by an experienced clinical team. These issues are compounded by the fact that all of the patients were implanted at a single center and managed by a single neurologist. This provided consistency in our dataset and reduced statistical variance, potentially limiting the general applicability of the specific coefficients listed in our first-generation prediction equations. Nonetheless, the statistical

significance and correlation coefficients of our analysis were strong, suggesting the concepts presented may be applicable to the broader DBS patient population.

The second major limitation of this study was the simplicity of the analysis. We focused on standard clinical measures of motor outcome (i.e. UPDRS-III) taken at various times in the patient history. However, these are typically qualitative measures that fail to capture the complex detail of specific symptoms. In addition, our analysis ignored the non-motor symptoms of PD. Therefore, it is likely that incorporation of a wider set of symptoms, evaluated with more detailed quantitative metrics, coupled to non-linear regression or advanced machine learning methods, could further enhance prediction accuracy.

In summary, our results highlight the important interplay between stimulation and medication in PD patient outcomes. While clinical decision support systems are becoming available to assist in DBS parameter selection, the overall success of DBS therapy is likely dependent on finding the optimal balance between both stimulation and medication. We propose that machine learning based systems are capable of predicting this interplay and suggesting treatment strategies that maximize clinical outcomes. Such concepts represent an exciting opportunity for the field of “Big Data Analytics” to directly impact PD medicine, because the predictive power of the CDSS grows as the size and scope of the database used for parameterization grows. However, this hope for the future is intimately dependent upon the rigor and quality of the clinical measurements used to populate the database.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the National Institutes of Health (NIH R01 NS047388)

References

- Bronstein JM, Tagliati M, Alterman RL, Lozano AM, Volkmann J, Stefani A, et al. Deep brain stimulation for Parkinson disease: an expert consensus and review of key issues. *Arch Neurol*. 2011; 68:165. [PubMed: 20937936]
- Butson CR, Cooper SE, Henderson JM, Wolgamuth B, McIntyre CC. Probabilistic analysis of activation volumes generated during deep brain stimulation. *Neuroimage*. 2011; 54:2096–2104. [PubMed: 20974269]
- Castrioto A, Lozano AM, Poon Y-Y, Lang AE, Fallis M, Moro E. Ten-year outcome of subthalamic stimulation in Parkinson disease: a blinded evaluation. *Arch Neurol*. 2011; 68:1550–1556. [PubMed: 21825213]
- Castrioto A, Volkmann J, Krack P. Postoperative management of deep brain stimulation in Parkinson's disease. *Handb Clin Neurol*. 2013; 116:129–146. [PubMed: 24112890]
- Chaturvedi A, Butson CR, Lempka SF, Cooper SE, McIntyre CC. Patient-specific models of deep brain stimulation: influence of field model complexity on neural activation predictions. *Brain Stimul*. 2010; 3:65–67. [PubMed: 20607090]
- Chaturvedi A, Luján JL, McIntyre CC. Artificial neural network based characterization of the volume of tissue activated during deep brain stimulation. *J Neural Eng*. 2013; 10:056023. [PubMed: 24060691]

- Daneault JF, Carignan B, Codère CÉ, Sadikot AF, Duval C. Using a smart phone as a standalone platform for detection and monitoring of pathological tremors. *Front Hum Neurosci*. 2013; 6:357. [PubMed: 23346053]
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006; 31:968–980. [PubMed: 16530430]
- Eisenstein SA, Koller JM, Black KD, Campbell MC, Lugar HM, Ushe M, Tabbal SD, Karimi M, Hershey T, Perlmuter JS, Black KJ. Functional anatomy of subthalamic nucleus stimulation in Parkinson disease. *Ann Neurol*. 2014; 76(2):279–295. [PubMed: 24953991]
- Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012; 30:1323–1341. [PubMed: 22770690]
- Frankemolle AMM, Wu J, Noecker AM, Voelcker-Rehage C, Ho JC, Vitek JL, et al. Reversing cognitive-motor impairments in Parkinson's disease patients using a computational modelling approach to deep brain stimulation programming. *Brain*. 2010; 133:746–761. [PubMed: 20061324]
- Keuken MC, Bazin P-L, Schäfer A, Neumann J, Turner R, Forstmann BU. Ultra-high 7T MRI of structural age-related changes of the subthalamic nucleus. *J Neurosci*. 2013; 33:4896–4900. [PubMed: 23486960]
- Lee MS, Lyoo CH, Ryu YH, Lim HS, Nam CM, Kim HS, Rinne JO. The effect of age on motor deficits and cerebral glucose metabolism of Parkinson's disease. *Acta Neurol Scand*. 2011 Sep; 124(3):196–201. [PubMed: 20880270]
- Maks CB, Butson CR, Walter BL, Vitek JL, McIntyre CC. Deep brain stimulation activation volumes and their association with neurophysiological mapping and therapeutic outcomes. *J Neurol Neurosurg Psychiatry*. 2009; 80(6):659–666. [PubMed: 18403440]
- McIntyre CC, Miocinovic S, Butson CR. Computational analysis of deep brain stimulation. *Expert Rev Med Devices*. 2007; 4:615–622. [PubMed: 17850196]
- Mera TO, Burack MA, Giuffrida JP. Objective motion sensor assessment highly correlated with scores of global levodopa-induced dyskinesia in Parkinson's disease. *J Parkinsons Dis*. 2013; 3(3):399–407. [PubMed: 23948993]
- Miocinovic S, Lempka SF, Russo GS, Maks CB, Butson CR, Sakaie KE, et al. Experimental and theoretical characterization of the voltage distribution generated by deep brain stimulation. *Exp Neurol*. 2009; 216:166–176. [PubMed: 19118551]
- Miocinovic S, Noecker AM, Maks CB, Butson CR, McIntyre CC. Cicerone: stereotactic neurophysiological recording and deep brain stimulation electrode placement software system. *Acta Neurochir Suppl*. 2007; 97:561–567. [PubMed: 17691348]
- Obeso JA, Olanow CW, Rodriguez-Oroz MC, Krack P, Kumar R, Lang AE. Deep-brain stimulation of the subthalamic nucleus or the pars interna of the globus pallidus in Parkinson's disease. *N Engl J Med*. 2001; 345:956–963. [PubMed: 11575287]
- Plaha P, Ben-Shlomo Y, Patel NK, Gill SS. Stimulation of the caudal zona incerta is superior to stimulation of the subthalamic nucleus in improving contralateral parkinsonism. *Brain*. 2006; 129:1732–1747. [PubMed: 16720681]
- Roberts DW, Hartov A, Kennedy FE, Miga MI, Paulsen KD. Intraoperative brainshift and deformation: a quantitative analysis of cortical displacement in 28 cases. *Neurosurgery*. 1998 Oct; 43(4):749–758. [PubMed: 9766300]
- Schuepbach WM, Rau J, Knudsen K, Volkmann J, Krack P, Timmermann L, Hälbig TD, et al. EARLYSTIM Study Group. Neurostimulation for Parkinson's disease with early motor complications. *N Engl J Med*. 2013 Feb 14; 368(7):610–622. [PubMed: 23406026]
- Schwartz WB. Medicine and the computer. The promise and problems of change. *N Engl J Med*. 1970; 283(23):1257–1264. [PubMed: 4920342]
- Thenganatt MA, Jankovic J. Parkinson disease subtypes. *JAMA Neurol*. 2014; 71(4):499–504. [PubMed: 24514863]

- Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Mov Disord.* 2010; 25:2649–2653. [PubMed: 21069833]
- Weaver FM, Follett K, Stern M, Hur K, Harris C, Marks WJ, et al. Bilateral deep brain stimulation vs best medical therapy for patients with advanced Parkinson disease: a randomized controlled trial. *JAMA.* 2009; 301:63–73. [PubMed: 19126811]
- Welter M-L, Schüpbach M, Czernecki V, Karachi C, Fernandez-Vidal S, Golmard J-L, et al. Optimal target localization for subthalamic stimulation in patients with Parkinson disease. *Neurology.* 2014; 82:1352–1361. [PubMed: 24647024]
- Zaidel A, Bergman H, Ritov Y, Md ZI. Levodopa and subthalamic deep brain stimulation responses are not congruent. *Mov Disord.* 2010; 25:2379–2386. [PubMed: 20824733]

- Development of a clinical decision support system for the treatment of Parkinson's disease patients implanted with deep brain stimulation systems.
- Machine learning algorithms incorporate patient-specific details to identify a theoretically optimal balance between stimulation parameter settings and medication dosages.
- Conceptual implementation of "Big Data Analytics" into the clinical management of Parkinson's disease.

Figure 1.
The concept of a database driven clinical decision support system (CDSS).

Figure 2.

Visualization of treatment outcomes of previous patients with respect to the simulated new patient. The example summarize the outcomes as a function of levodopa equivalent daily dosage (LEDD) and of its timing. Three thresholds for the similarity measure s were applied (1.0, 0.8 and 0.7), where a smaller s corresponds to greater similarity within the database, but fewer comparative samples.

Figure 3.

A recommender system for the CDSS. The clinician enters the patient's known information at the time of the postoperative visit. In this example we assume that theoretically optimal DBS parameter settings have already been defined (e.g. via existing clinical tools such as GUIDE DBS). Alternatively, the DBS parameters can be incorporated as additional variables in the system. Then treatment parameters, such as levodopa equivalent daily dosage (LEDD) and medication intake times, are randomly drawn and a prediction of outcomes is computed. Once the treatment parameters space is covered with enough samples, clusters of predicted effective treatment are computed and recommended.

Figure 4.

Computation of the *volume of tissue activated* (VTA) with respect to the preferred *target zone*. (a) Co-registration of the patient's preoperative MRI with a brain atlas. (b) Co-registration of the patient's preoperative MRI and postoperative CT to the preoperative CT with the stereotactic frame, thereby establishing a common coordinate system. (c) Definition of the DBS electrode position relative to the intraoperative microelectrode recordings and the anatomical volumes. Green dots – STN recording points, Green volume – STN, Yellow volume – Thalamus, Red volume – VTA, Grey volume – target zone.

Figure 5.

Examples of machine learning predicted outcomes (upper row) and recommended treatment (lower row) for the simulated new patient. The predicted outcomes were clustered into three categories: non-responsive, moderate response, or high response. The convex hull of the high response predictions is computed to define a predicted effective range of treatment parameters. The recommender system defines the center of the convex hull as the preferred treatment. Changing the stimulation parameters to increase the overlap of the volume of tissue activated and the target zone (i.e. DBS overlap) generates more medication options with high response.

10 advanced Parkinson's disease patients participated in this study (106 follow up office-visits; of them in 89 the volume of tissue activated overlapped with the preferred target zone and are further studied). Age refers to the patient's age at time of DBS surgery. Unified Parkinson's disease rating scale, part III (UPDRS-III) preoperative scores are presented. Off medication refers to >12h since last medication, on medication state was typically assessed one hour after medication intake.

Table 1

Patient #	Sex (M/F)	Age (year)	Follow up #visits	Follow up #months	Preoperative UPDRS-III off meds.	Postoperative UPDRS-III On meds.
01	F	64	3	13	48	33
02	M	38	10	24	48	17
03	F	74	10	47	54	31
04	F	71	10	20	51	25
05	M	63	20	38	53	20
06	M	71	9	27	22	9
07	M	63	6	7	46	31
08	F	54	4	6	36	17
09	M	64	26	54	32	13
10	M	69	8	19	23	12

Table 2

Percentages of correct predictions of the combined medication-DBS treatment outcomes ($N=14$ office visits; $N=8$ for consistent prediction method). Various machine-learning algorithms were applied to predict the overall and specific motor symptoms improvement according to the UPDRS-III.

Method Symptom	SVM	NB	RF	Consistent prediction	Symptoms aggregate
Overall motor	71%	64%	64%	88%	86%
Speech	93%	64%	86%	N/A	N/A
Tremor	100%	64%	100%	N/A	N/A
Rigidity	57%	57%	50%	N/A	N/A
bradykinesia (limbs)	78%	71%	86%	N/A	N/A
Akinetic (axial)	64%	50%	86%	N/A	N/A