Behavioral effects of subthalamic deep brain stimulation in Parkinson's disease

M. Denheyer\textsuperscript{a}, Z.H. Kiss\textsuperscript{b}, A.M. Haffenden\textsuperscript{a, b, *}

\textsuperscript{a} Department of Psychology, Alberta Health Services, Calgary, Alberta, Canada
\textsuperscript{b} Department of Clinical Neurosciences, University of Calgary, Calgary, Alberta, Canada

\textbf{A R T I C L E  I N F O}

Article history:
Received 9 March 2009
Received in revised form 18 June 2009
Accepted 30 July 2009
Available online 5 August 2009

Keywords:
Executive dysfunction
Behavior
Apathy
Disinhibition
Cognition

\textbf{A B S T R A C T}

To date, few studies have utilized standardized measures to assess the neurobehavioral changes that can accompany deep brain stimulation (DBS) of the subthalamic nuclei (STN) for the treatment of Parkinson’s disease (PD); yet behavioral changes are the most debated among practitioners. We evaluated behavior with the Frontal Systems Behavior Scale (FrSBe), which includes a large-scale normative sample for self- and collateral ratings and is particularly relevant to PD with subscales assessing Apathy, Disinhibition, and Executive Dysfunction. Data were collected from 16 (11 males) PD patients. All FrSBe subscale scores increased significantly when retrospective preoperative scores and current (postoperative) scores were compared. Self- and collateral FrSBe ratings were not significantly correlated with each other, though for both scores at least half of the group met criteria for a clinically significant level of symptoms postoperatively. No significant correlations were seen for collateral current FrSBe ratings with cognitive or motor variables. Higher self-ratings of behavior characteristic of apathy were related to higher self-ratings of depressive symptoms, and to a smaller decrease in antiparkinsonian medications following surgery. We propose that the standardized assessment of behavioral aspects of executive dysfunction adds information that is largely dissociable from the motor and cognitive assessment of function in PD patients undergoing STN DBS. In future, prospective standardized measurement of behavior may allow for better prediction of which patients will experience significant behavioral issues postoperatively.

\textsuperscript{*} Corresponding author at: Foothills Medical Centre, Department of Psychology, 1403 29 Street NW, Calgary, Alberta, Canada T2N 2T9. Tel.: +1 403 944 1340; fax: +1 403 944 2060.
E-mail address: angela.haffenden@albertahealthservices.ca (A.M. Haffenden).

1. Introduction

Both Parkinson’s disease (PD) and its treatment impact frontal–striatal systems, circuitry that is critical for behavioral regulation (Frank, Scheres, & Sherman, 2007). While deep brain stimulation (DBS) of the subthalamic nuclei (STN) is one of the most effective treatments for PD motor dysfunction (for review see Limousin & Martinez-Torres, 2008), it is thought to adversely affect behavior and executive function in some patients (for review see Voon, Kubu, Krack, Houeto, & Tröster, 2006). Although the behavioral effects of STN DBS are widely debated both in the literature (see for example Tröster, 2008 commentary) and among practitioners, only the cognitive impact of STN DBS has been well-studied using standardized measures. To date, few reports have utilized standardized measures to assess neurobehavioral changes despite a call for such measurement nearly a decade ago (Trépanier, Kumar, Lozano, Lang, & Saint-Cyr, 2000).

Previous reports of behavioral executive dysfunction issues following surgery have been largely descriptive. One publication reported results of a standardized measure for a small selected sample of patients (Saint-Cyr, Trépanier, Kumar, Lozano, & Lang, 2000). A recent paper (Smeding et al., 2006) administered the Dysexecutive Questionnaire (DEX; Wilson, Alderman, Burgess, Emslie, & Evans, 1996); however, normative data are not available for this measure. The prevalence of papers on cognitive outcomes of STN DBS and the absence of papers focused on standardized assessment of behavior is particularly notable given that cognitive and behavioral measures purported to tap executive functions are readily dissociable (e.g., Reid-Arndt, Nehl, & Hinkebein, 2007; Shallice & Burgess, 1991; Vriezen & Pigott, 2002). Indeed, success on a standardized measure of problem solving or mental flexibility may fail to capture real-world behavioral consequences of executive dysfunction. Given the potential impact of behavioral changes on the functioning of the individual, and consequently their family, reliable and valid assessment of behavior is critical.

The choice of behavioral measures is far narrower than the choice of cognitive measures of frontal system functioning. Of the behavioral measures, The Frontal Systems Behavior Scale (FrSBe; Grace & Malloy, 2001) stands out as particularly relevant to PD. The FrSBe consists of 46 items, each of which falls into one of three subscales: Apathy, Disinhibition, or Executive Dysfunction. Zgaljardic and colleagues have previously emphasized how well these subscales map onto the frontal–striatal circuitry relevant to PD (Zgaljardic, Borod, Foldi, & Mattis, 2003). Psychometrically, the
FrSBe distinguishes itself from other measures with its large-scale normative sample for self- and family ratings and demonstrated validity for the assessment of behavior disturbances associated with damage to the frontal–subcortical circuits (Malloy & Grace, 2005). For these reasons we chose the FrSBe to provide quantitative assessment of post-operative behavioral functioning in PD patients who had undergone STN DBS surgery. Our primary objective was to explore the utility of the FrSBe in characterizing behavioral disturbance in this population. Secondly, we examined the relationship between FrSBe scores, mood, motor function, and cognitive measures. Finally, we sought to determine the relationship between self- and family reports of behavior.

2. Methods

2.1. Patients

A total of 23 patients underwent bilateral STN DBS for treatment of idiopathic PD in our clinic between January 2002 and December 2006. All patients underwent preoperative neuropsychological evaluations that are detailed below. Methods for bilateral insertion of DBS into the STN have been described previously (Hutchison et al., 1998). We performed staged implantation with the worse brain side operated on first, followed about 1 month later by second side surgery. Reasons for exclusion from this study were death unrelated to surgery (3), patient relocation (4), and incomplete neuropsychological data (2; e.g., non-English speaking patients); some patients had more than one reason for exclusion. Participants remaining were 16 (11 males) English speaking PD patients. Patient demographics and preoperative neuropsychological assessments were performed an average of 7 months (± 2 months) after this surgery. Although a range of scores are generated for the WCST, the percentage of error responses was chosen as the outcome measure because preliminary analysis demonstrated a high degree of correlation between all scores generated for this test and this score provided a measure of overall performance. The above-described cognitive scores were corrected for age and education using established normative samples and subsequently converted to z scores. An estimate of premorbid intelligence was made from the North American Adult Reading Test (NAART; Blair & Spreen, 1989). In addition, mood was assessed pre- and postoperatively using self-ratings. This study was undertaken with ethics approval to collect information on routine care for all neuromodulation patients. Patients provided informed, written consent.

2.2. Assessments

2.2.1. Motor function

Unified Parkinson’s Disease Rating Scale (UPDRS) evaluations were conducted in the Movement Disorders Clinic a few months prior to the first implantation (M = 3 ± 3 months) in both medication ON and practically OFF states (after withdrawal of medication for at least 8 h) and again 12 months postoperatively (M = 13 ± 1 months) in 4 states (Medication OFF/Stimulation OFF, Medication OFF/Stimulation ON, Medication ON/Stimulation OFF, Medication ON/Stimulation ON). Only best ON states in 4 states (Medication OFF/Stimulation OFF, Medication OFF/Stimulation ON, Medication ON/Stimulation OFF, Medication ON/Stimulation ON) were used for analysis, in order to examine the usual functional state of each patient, and evaluate the degree of improvement produced by the surgery. Scores from the UPDRS part II (activities of daily living: ADL) and UPDRS part III (motor) subscales were entered into the analysis. The levodopa equivalent dose was calculated as per the method of Chen, Garg, Lozano, and Lang (2001) using the medications taken at the time of the motor evaluations.

2.2.2. Cognitive function

Neuropsychological assessments were performed an average of 7 months (SD = 6 months) after the first surgery. Although a range of cognitive domains were assessed, for the present study we focused on measures of executive function due to the particular relevance of these functions to PD, STN DBS, and the FrSBe. Phonemic and semantic fluency were measured using the letter “FAS” and the category “animal,” respectively (Strauss, Sherman, & Spreen, 2006). Problem-solving aspects of executive function were assessed with the Wisconsin Card Sorting Test (WCST; Heaton, Chelune, Talley, Kay, & Curtis, 1993). Although a number of scores are generated for the WCST, the percentage of error responses was chosen as the outcome measure because preliminary analysis demonstrated a high degree of correlation between all scores generated for this test and this score provided a measure of overall performance. The above-described cognitive scores were corrected for age and education using established normative samples and subsequently converted to z scores. An estimate of premorbid intelligence was made from the North American Adult Reading Test (NAART; Blair & Spreen, 1989).

2.2.3. Behavioral function

FrSBe questionnaires were administered after completion of STN DBS surgery (average latency = 40 months, SD = 23, range = 13–78). Each of the 46 items on the FrSBe describes a behavior, and the frequency of the behavior is rated from 1 (Almost never) to 5 (Almost always). Each item falls onto one of three subscales: Apathy (14 items), Disinhibition (15 items), and Executive Dysfunction (17 items). The FrSBe is designed such that ratings are made retrospectively prior to an event (surgery in the present study) and as well as for current behavior. As well, the FrSBe has separate forms for patient self-reports and family reports with the same items, using either first or third person references to behaviors, respectively. Aside from one patient who had a sibling respond, spouses of patients provided the family report. Patient and family FrSBe reports were administered at the same time, but without communication between the patient and their family member during completion of the questionnaire. In addition, during the session in which they completed the FrSBe, patients also completed a BDI II to determine whether depression was contributing to item endorsement on the FrSBe. Using the normative data supplied by the test publisher for age, education, sex, and respondent (self or family member), all FrSBe ratings were converted to T scores. As per guidelines in the FrSBe manual, T scores equal to or greater than 65 are considered clinically significant in terms of the level of symptom severity; symptom severity increases with increasing scores.

2.3. Statistical analysis

Analyses were conducted using SPSS 15.0. All change variables were calculated as Post-STN DBS–Pre-STN DBS. A Shapiro–Wilks test identified several variables that were not normally distributed; a rank transformation (ranks) of the data was used for the correlational analysis. To balance the number of correlations conducted with the exploratory nature of the study, significance for the correlational analysis was set as p < .01 (r ≤ .623). For paired comparisons of pre- and post-surgical scores, paired t-tests were used to compare normally distributed variables and Wilcoxon signed rank tests were used for non-normally distributed variables with significance set at p < .05 after correction for multiple comparisons.

3. Results

3.1. Pre- and postoperative comparisons

3.1.1. Motor and cognitive variables

Changes in non-FrSBe variables are summarized in Table 2. Motor scores on the UPDRS improved when scored on treatment
(medication preoperatively, and medication plus stimulation postoperatively). Additionally, daily levodopa equivalent dose declined significantly following surgery. In contrast, no significant changes were detected in performance of ADLs (UPDRS II) or scores on the Beck Depression Inventory II completed during the postoperative neuropsychological assessment. For the cognitive variables, performance on the WCST test was unchanged, but both semantic and phonemic fluency scores declined following surgery. When corrected for multiple comparisons using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995), only the change in levodopa dose and phonemic fluency remained significant in this sample.

3.1.2. FrSBe data

All FrSBe subscale scores (T scores) increased significantly when pre- and postoperative scores were compared (Fig. 1A). For all subscale comparisons, t(15) > 3.04, with all p < .05 when corrected for multiple comparisons. Except for family ratings on the disinhibition subscale, mean postoperative values exceeded the cutoff for clinical significance of T ≥ 65. The only item from this subscale that showed consistently higher self-ratings as compared to family ratings addressed decreased taste and smell.

Caseness on the FrSBe (a T score equal to or greater than 65) was considered based on individual subscales (Fig. 1b) because meeting criterion for caseness on one subscale did not necessarily correspond to the same symptom severity on the other subscales. Corresponding to the increase in average scores, there was an increase in the number of cases meeting criterion for clinical significance from pre- to post-surgery. For family ratings, all patients meeting the criterion for caseness preoperatively also met the criterion for caseness postoperatively. For self-ratings, a single patient showed a decline in score on the apathy and executive dysfunction subscales such that they met the criterion for caseness preoperatively but not postoperatively.

Comparison between self- and family ratings for subscale T scores yielded no significant differences (all t(15) < 1.2, p > .05). McNemar’s test found no significant differences (all p > .05) when comparing the number of patients meeting caseness for self-ratings versus family ratings. Finally, the degree of agreement between self- and family ratings was assessed using a two-way mixed effects intra-class correlation (ICC model 3). Exact item agreement (i.e., the patient and family members endorsed an item using identical scores) was higher for current ratings (ICC = .4) as compared to retrospective ratings (ICC = .3). Based on this finding, subsequent correlational analyses were conducted on current ratings only to eliminate concerns that the retrospective nature of the preoperative ratings would introduce extraneous variability into the correlations.

3.2. Correlational analyses

The postoperative (current rating) FrSBe scores were entered into a nonparametric correlational analysis with the pre- and postoperative scores listed in Table 2. Also entered into the analysis were age at the time of surgery, level of education, the estimate of premorbid intelligence based on the North American Adult Reading Test, and the BDI II completed at the same time as the FrSBe. Because of the variable range of time that had elapsed since surgery within the group, we examined the correlation between the time since surgery and performance on all other measures. None of the correlations were significant, r(14) < -.55; in all cases, p > .01.

3.2.1. Relationships between FrSBe postoperative ratings and motor and cognitive variables

The self- and family postoperative FrSBe T scores revealed identical relationships between the FrSBe subscales. Executive Dysfunction correlated significantly with both Apathy (r(14) = .68 self; r(14) = .74 family; p < .01) and Disinhibition (r(14) = .67 self;
The BDI II score completed during the follow-up neuropsychological assessment showed a similar relationship ($r_s(14) = .52$, $p < .05$) but did not meet the more stringent alpha level imposed ($p < .01$) due to the number of correlations performed. The postoperative BDI II scores were themselves highly correlated ($p < .001$). Although the BDI II completed during the same session as the FrSBE resulted in a higher score than that completed at the preoperative or the postoperative neuropsychological assessments (Table 2), these differences did not reach significance ($p > .05$, corrected for two comparisons). Examination of items on the BDI II that correspond to the BDI items assessing loss of energy and changes in interest in sex. It is also notable that of the five BDI II items with the highest mean endorsement, three (tiredness, loss of interest in sex, and loss of energy) are considered to assess somatic symptoms of depression. Relative to the preoperative BDI II scores, the items that showed the largest increases in mean endorsement assessed loss of interest in sex, pessimism about the future, degree of past failure, and worthlessness. Thus, the majority of these items are neither somatic nor do they overlap with items from the Apathy subscale of the FrSBE.

3.2.2. Relationships between motor and cognitive variables

Not surprisingly, higher preoperative levodopa equivalent doses were related to poorer preoperative performance of ADLs (UPDRS II scores) when OFF medication ($r_s(14) = .75$, $p < .01$). As well, the change in UPDRS II scores (post–preoperative score) off treatment was related to the number of years of education patients had completed, though this correlation fell just above the cut-off for significance ($r_s(14) = −.62$, $p = .01$). From this, it appears the lower the educational level of the patient, the more deterioration they showed in ADLs off of treatment during the period from pre- to post-surgery; that is, their disease appeared to progress more rapidly.

In terms of neuropsychological measures, the higher the BDI II scores (increased preoperative depression) the greater the decline in phonemic fluency score from pre- to post-surgery ($r_s(14) = −.86$, $p < .001$). Preoperative performance on a semantic fluency test was correlated with the change in OFF UPDRS motor scores ($r_s(14) = .75$, $p < .01$). That is, better semantic fluency performance prior to surgery was associated with better maintenance of motor func-

$r_s(14) = .90$ (family; $p < .01$). However, Apathy and Disinhibition subscales were not significantly correlated ($r_s(14) = .32$ for self-ratings and $r_s(14) = .55$ for family ratings). As well, there were no significant correlations between self- and family FrSBE scores (maximum $r_s = .4$; average $r_s = .3$). Notably, for family ratings no significant correlations were seen with any variable outside of correlations among the family-rated FrSBE subscales. However, there were significant relationships between FrSBE self-ratings and two non-FrSBE variables (Fig. 2). The percent reduction in levodopa dose from the preoperative to the 1-year postoperative assessment was significantly correlated with self-rated apathy ($p < .01$), though in a negative fashion. That is, the larger the decrease in medications, the lower the self-rated apathy scores. In addition, the BDI II that was completed during the same session as the FrSBE ($M = 14.53$, $SD = 6.95$) was positively correlated with self-rated apathy ($p < .01$). The BDI II score completed during the follow-up neuropsychological assessment showed a similar relationship ($r_s(14) = .52$, $p < .05$) but did not meet the more stringent alpha level imposed ($p < .01$) due to the number of correlations performed. The postoperative BDI II scores were themselves highly correlated ($p < .001$). Although the BDI II completed during the same session as the FrSBE resulted in a higher score than that completed at the preoperative or the postoperative neuropsychological assessments (Table 2), these differences did not reach significance ($p > .05$, corrected for two comparisons). Examination of items on the BDI II that correspond to the BDI items assessing loss of energy and changes in interest in sex. It is also notable that of the five BDI II items with the highest mean endorsement, three (tiredness, loss of interest in sex, and loss of energy) are considered to assess somatic symptoms of depression. Relative to the preoperative BDI II scores, the items that showed the largest increases in mean endorsement assessed loss of interest in sex, pessimism about the future, degree of past failure, and worthlessness. Thus, the majority of these items are neither somatic nor do they overlap with items from the Apathy subscale of the FrSBE.

$r_s(14) = .90$ (family; $p < .01$). However, Apathy and Disinhibition subscales were not significantly correlated ($r_s(14) = .32$ for self-ratings and $r_s(14) = .55$ for family ratings). As well, there were no significant correlations between self- and family FrSBE scores (maximum $r_s = .4$; average $r_s = .3$). Notably, for family ratings no significant correlations were seen with any variable outside of correlations among the family-rated FrSBE subscales. However, there were significant relationships between FrSBE self-ratings and two non-FrSBE variables (Fig. 2). The percent reduction in levodopa dose from the preoperative to the 1-year postoperative assessment was significantly correlated with self-rated apathy ($p < .01$), though in a negative fashion. That is, the larger the decrease in medications, the lower the self-rated apathy scores. In addition, the BDI II that was completed during the same session as the FrSBE ($M = 14.53$, $SD = 6.95$) was positively correlated with self-rated apathy ($p < .01$). The BDI II score completed during the follow-up neuropsychological assessment showed a similar relationship ($r_s(14) = .52$, $p < .05$) but did not meet the more stringent alpha level imposed ($p < .01$) due to the number of correlations performed. The postoperative BDI II scores were themselves highly correlated ($p < .001$). Although the BDI II completed during the same session as the FrSBE resulted in a higher score than that completed at the preoperative or the postoperative neuropsychological assessments (Table 2), these differences did not reach significance ($p > .05$, corrected for two comparisons). Examination of items on the BDI II that correspond to the BDI items assessing loss of energy and changes in interest in sex. It is also notable that of the five BDI II items with the highest mean endorsement, three (tiredness, loss of interest in sex, and loss of energy) are considered to assess somatic symptoms of depression. Relative to the preoperative BDI II scores, the items that showed the largest increases in mean endorsement assessed loss of interest in sex, pessimism about the future, degree of past failure, and worthlessness. Thus, the majority of these items are neither somatic nor do they overlap with items from the Apathy subscale of the FrSBE.

$r_s(14) = .90$ (family; $p < .01$). However, Apathy and Disinhibition subscales were not significantly correlated ($r_s(14) = .32$ for self-ratings and $r_s(14) = .55$ for family ratings). As well, there were no significant correlations between self- and family FrSBE scores (maximum $r_s = .4$; average $r_s = .3$). Notably, for family ratings no significant correlations were seen with any variable outside of correlations among the family-rated FrSBE subscales. However, there were significant relationships between FrSBE self-ratings and two non-FrSBE variables (Fig. 2). The percent reduction in levodopa dose from the preoperative to the 1-year postoperative assessment was significantly correlated with self-rated apathy ($p < .01$), though in a negative fashion. That is, the larger the decrease in medications, the lower the self-rated apathy scores. In addition, the BDI II that was completed during the same session as the FrSBE ($M = 14.53$, $SD = 6.95$) was positively correlated with self-rated apathy ($p < .01$). The BDI II score completed during the follow-up neuropsychological assessment showed a similar relationship ($r_s(14) = .52$, $p < .05$) but did not meet the more stringent alpha level imposed ($p < .01$) due to the number of correlations performed. The postoperative BDI II scores were themselves highly correlated ($p < .001$). Although the BDI II completed during the same session as the FrSBE resulted in a higher score than that completed at the preoperative or the postoperative neuropsychological assessments (Table 2), these differences did not reach significance ($p > .05$, corrected for two comparisons). Examination of items on the BDI II that correspond to the BDI items assessing loss of energy and changes in interest in sex. It is also notable that of the five BDI II items with the highest mean endorsement, three
tion when OFF. Finally, post-surgical semantic fluency performance was negatively correlated with the percent change in levodopa equivalent dose ($r_{14} = -.65, p < .01$). In other words, better semantic fluency performance was seen in individuals with smaller decreases in their antiparkinsonian medications following surgery.

4. Discussion

Currently, few reports measuring neurobehavioral aspects of STN DBS outcomes using standardized validated instruments are available. Yet behavior is purported to be significantly affected by STN DBS (for review see Voon et al., 2006). We focused on the FrSBe to quantify behavior following surgery, because it is particularly appropriate for issues anticipated in a PD population (Zgaljardic et al., 2003) and allows for comparison with normative data for both self- and family/collateral ratings. The key findings from the FrSBe were that both self- and family ratings of behavior increased significantly when pre- and postoperative scores were compared, with no significant differences identified between self- and family ratings. However, exact item agreement was higher for current ratings as compared to the retrospective preoperative ratings. For family ratings no significant correlations were seen with any variable outside of correlations among the family-rated FrSBe subscales. However, there were significant relationships between FrSBe self-ratings of apathy and both depression and the percent decrease in dopaminergic medications.

In our sample, significant increases were seen across the three subscales of the FrSBe (Apathy, Disinhibition, and Executive Dysfunction) when retrospective pre-surgical ratings were compared with the current postoperative ratings. This was true for self-reports and for family reports. The finding of increased postoperative FrSBe family ratings is consistent with one study (Saint-Cyr et al., 2000) that reported post-STN DBS data for 11 patients using a previous version of the FrSBe, known as the Frontal Lobe Personality Scale (FLOPS; Grace, Stout, & Malloy, 1999). In contrast to the change in family ratings, the patients’ self-ratings did not show change on any measure of the FLOPS. This is quite unlike the present study where changes in patient self-ratings were generally equivalent to family ratings. Notably, the patients in the current study were 12 years younger on average and had better pre-surgical motor function as per their UPDRS scores when compared to the patients described by Saint Cyr et al. Thus, it is possible that our sample had less advanced PD, and as a result had better insight or fewer behavioral issues to report.

Assessing both patient and family scores is typically used as a tool for identifying cases where patients have limited insight into behavioral issues. In the current sample, three patients endorsed lower frequencies of current behavioral issues on the majority of items when compared to ratings made by family members. However, the opposite pattern was also evident; three patients almost never (<10% of items) endorsed lower frequencies of current behavioral issues when compared to ratings made by family members. In addition one FrSBe item, related to food having no taste or smell, reliably resulted in higher patient endorsement than that indicated by family members. Although decreased taste and smell are common in PD patients (Shah et al., 2009), this symptom may be less noticeable and less problematic to caregivers than other PD symptoms.

Aside from the results reported by Saint-Cyr et al. (2000) and the present study, the only other study that reported self- and proxy ratings from a behavioral measure specific to executive function in an STN DBS population was that of Smeding et al. (2006) using the DEX. They found no significant increase in DEX scores 6 months after bilateral STN DBS in either self- or proxy ratings despite a much larger sample size ($n = 99$) than either the current study or that of Saint Cyr et al. Although the average patient characteristics in Smeding’s sample were similar to those of the current sample in terms of age and pre-surgical motor functioning, the relatively short follow-up period may explain their null results. In addition, the sensitivity of the 20-item DEX to issues specific to PD and those issues that arise after STN DBS is uncertain. Mathias (2003) found no difference between DEX scores for PD patients and age-matched controls for either self- or proxy ratings. The lack of a normative sample or established cutoffs for clinically significant scores further undermine the utility of the DEX as a measure of STN DBS outcome.

In the present study, correlations between FrSBe subscale scores and two cognitive measures of executive function did not approach significance. Notably, significant changes were seen in FrSBe subscale scores and phonemic verbal fluency, and the direction of these changes would typically be interpreted as worsened executive function. The lack of correspondence between FrSBe scores and either verbal fluency or the WCST supports the suggestion that both performance based measures and questionnaires should be used to assess executive function (Miyake, Emerson, & Friedman, 2000). Indeed, the nature of a structured cognitive assessment controls aspects of functioning relevant to performance based measures. Examinees are prompted, provided with response rules, and usually required to focus on a single task for a relatively brief period of time. Thus, the addition of a standardized questionnaire taps different aspects of executive functioning across different periods of time.

As well, cognitive measures are subject to practice effects. The lack of significant improvement in our sample on the WCST brings into question whether this may reflect executive dysfunction. However, a study examining reliable change interval in a non-surgical PD population over an average 17-month test–retest demonstrated that greater than 80% of patients tested maintained stable scores on the WCST, with fewer than 10% of patients showing improvement (Tröster, Woods, & Morgan, 2007). Similarly, when test–retest performance of surgical and non-surgical PD patients was directly compared (with the surgical group tested 6 months after the completion of surgery) both groups showed similar patterns of change on the WCST (York et al., 2008). The vast majority of patients evidenced no reliable change on the WCST, regardless of whether or not they had had surgery. Taken together, these findings suggest that the lack of a significant practice effect in WCST performance in the present study provides no evidence of executive function decline related to surgery or stimulation.

No significant relationship between FrSBe scores and cognitive scores, only two significant relationships were identified between FrSBe scores and non-neuropsychological measures. First, self-rated Apathy and the percent decrease in levodopa equivalent dose were negatively correlated. Higher levels of apathy were associated with a smaller relative decrease in antiparkinsonian medications following STN DBS. This finding contrasts with the idea that the increase in behaviors characteristic of apathy following STN DBS is related to the reduction of dopaminergic treatments allowed by the surgery (Krack et al., 1998). However, subsequent studies using quantitative measures of apathy have not consistently supported this relationship. Czernicki et al. (2005) found no difference between the doses of dopaminergic agonists between patients whose apathy improved, and those whose apathy did not improve with STN stimulation. Indeed, the link between levodopa and apathy appears to show significant individual variability. In the recent report (Czernicki et al., 2008) on 20 patients who were able to stop dopaminergic medication completely after successful STN stimulation, of those patients willing to be treated and available for follow-up, six patients did not develop apathy. Seven of eight patients that developed apathy responded well to the dopamine agonist ropinirole. So although improvement in apathy with dopaminergic medication would be expected in an apathetic...
population, in a mixed apathetic and non-apathetic population, the correlation between apathy and medication would not necessarily follow the same pattern.

The second significant correlation was seen between postoperative FrSBe self-rated Apathy scores and BDI II scores generated during the same session as the FrSBe scores. This finding, along with the correlation between the FrSBe subscales of Executive Dysfunction and Apathy, is consistent with previous findings using these same measures in a non-surgical PD population (Zgaljardic et al., 2007). Indeed, the relationship between apathy and depression in PD has generated much interest, though recent findings indicate that apathy can be dissociated from depression in a non-surgical PD population (Kirsch-Darrow, Fernandez, Marsiske, Okun, & Bowers, 2006) and in a PD population following STN DBS (Draper et al., 2006). However, in the present study, two of the BDI II items with the highest postoperative mean endorsement (loss of energy and changes in interest in sex) happen to correspond quite closely to items on the Apathy subscale of the FrSBe.

It is also notable that of the five BDI II items with the highest mean endorsement, three (tiredness, loss of interest in sex, and loss of energy) assess somatic symptoms of depression. Previous research suggests that depression in PD is not a somatic artifact and that the BDI is a reliable and valid measure of depression in PD (Levin, Llabre, & Weiner, 1988), though adjustment to the cut-points of the BDI are recommended (Visser, Leentjens, Marinus, Stiggelbout, & van Hilten, 2006). As well, BDI scores do not correlate with UPDRS motor scores (Fernandez et al., 2009; Tumas, Rodrigues, Farias, & Crippa, 2008), and it is the patients’ perception of the impact of the disease on their lives that is more important than the severity and degree of symptomatic impairment in terms of level of depression in PD (Schrag, Jahanshahi, & Quinn, 2001). Overall, the majority of patients did not evidence BDI scores diagnostic of depression using proposed cut-points for PD on this scale (e.g., Leentjens, Verhey, Luijckx, & Troost, 2000; Tumas et al., 2008), suggesting that depression does not account for the changes seen on the FrSBe.

4.1. Limitations

Key limitations of the current study were the subjective nature of the FrSBe ratings and the retrospective pre-surgical data collection. Subjective ratings are undoubtedly impacted by preconceived expectations of the outcome of STN DBS. Montel and Bungener (2009) compared quality of life in a group of Parkinson's patients who underwent STN DBS with patients who remained on their usual pharmacotherapy. Despite substantial improvements in motor signs in the stimulated patients, the quality of life was the same for the two groups one year after surgery. The authors proposed that the subjective nature of quality of life evaluations and patients unrealistic expectations of surgical outcome led to this result. In terms of the retrospective nature of the data collection, the finding that correspondence between patient and family ratings was better for current (postoperative) as opposed to the retrospective (preoperative) ratings points to differences in retrospective recall. Although, the retrospective ratings in our sample were generally consistent with scores reported in the literature for similar non-surgical PD populations (Stout, Ready, Grace, Malloy, & Paulsen, 2003; Zgaljardic et al., 2007), literature from other patient populations raises concerns about changes or decline in function rated retrospectively following surgery. For example, after surgical treatment of epilepsy, only the minority of patients who experienced significant declines were sensitive to the actual changes in memory relative to their pre-surgical functioning (Lineweaver, Naugle, Cafaro, Bingaman, & Lüders, 2004). The link between accuracy in patients’ retrospective ratings and degree of change may also hold true for non-surgical treatments. In a study examining the relationship between serial assessment outcomes and patients’ retrospective views during arthritis treatment, correlations improved as the degree of change increased (Fischer et al., 1999). Applied to our study, ratings may be most accurate for those patients who showed a significant degree of behavioral change relative to their pre-surgical function. For other patients, the degree of change may well be overestimated. Prospective investigations should be undertaken to determine how patients and family members rate themselves on the FrSBe prior to surgery, and whether scores can assist in predicting behavioral change following bilateral STN DBS. Additionally, although the latency between surgery and FrSBe scores were not significantly correlated in the present study, post-surgical ratings of current behavior carried out serially over a number of years would allow for specific analysis of the effect of time since surgery on behavioral aspects of executive dysfunction.

A further limitation of the present study was the absence of a control group outside of family ratings of behavior. Therefore the possibility that some of the changes in self-reported behavior across time are secondary to PD, rather than surgery or ongoing stimulation, cannot be ruled out without comparison to a pharmacologically treated group of patients of equivalent disease stage. The use of family ratings as a control raises issues in terms of changes in patient/caregiver roles resulting from motor improvement, and patient/caregiver expectations of surgery. The ideal control group would be PD patients with similar symptoms undergoing either an alternative surgical treatment or a sham surgery. Unfortunately, this study design is not possible at the present time. Finally, the results of the present study should be considered preliminary as our sample may have been younger and more educated that previously reported samples.

In conclusion, results from a standardized measure demonstrate an overall increase in behaviors related to executive dysfunction in patients who have undergone bilateral STN DBS, despite an absence of change on standard neuropsychological tests. We propose that patients being screened for STN surgery undergo a standardized assessment of behavioral aspects of executive dysfunction. Negative behavioral outcomes can have a significant impact on functioning, and can be dissociated from motor outcomes and cognitive aspects of executive function. Prospective measurement of behavioral aspects of executive dysfunction may allow for prediction of which patients are likely to experience clinically significant levels of maladaptive behaviors postoperatively, and specific counseling may be performed preoperatively to minimize these impacts on the patient and their families.

References
