Cognitive sequelae of subthalamic nucleus deep brain stimulation in Parkinson’s disease: a meta-analysis

Thomas D Parsons, Steven A Rogers, Alyssa J Braaten, Steven Paul Woods, Alexander I Tröster

Summary

Background Deep brain stimulation of the subthalamic nucleus (STN DBS) is an increasingly common treatment for Parkinson’s disease. Qualitative reviews have concluded that diminished verbal fluency is common after STN DBS, but that changes in global cognitive abilities, attention, executive functions, and memory are only inconsistently observed and, when present, often nominal or transient. We did a quantitative meta-analysis to improve understanding of the variability and clinical significance of cognitive dysfunction after STN DBS.

Methods We searched MedLine, PsycLIT, and ISI Web of Science electronic databases for articles published between 1990 and 2006, and extracted information about number of patients, exclusion criteria, confirmation of target by microelectrode recording, verification of electrode placement via radiographic means, stimulation parameters, assessment time points, assessment measures, whether patients were on levodopa or dopaminomimetics, and summary statistics needed for computation of effect sizes. We used the random-effects meta-analytical model to assess continuous outcomes before and after STN DBS.

Findings Of 40 neuropsychological studies identified, 28 cohort studies (including 612 patients) were eligible for inclusion in the meta-analysis. After adjusting for heterogeneity of variance in study effect sizes, the random effects meta-analysis revealed significant, albeit small, declines in executive functions and verbal learning and memory. Moderate declines were only reported in semantic (Cohen’s d 0.73) and phonemic verbal fluency (0.51). Changes in verbal fluency were not related to patient age, disease duration, stimulation parameters, or change in dopaminomimetic dose after surgery.

Interpretation STN DBS, in selected patients, seems relatively safe from a cognitive standpoint. However, difficulty in identification of factors underlying changes in verbal fluency draws attention to the need for uniform and detailed reporting of patient selection, demographic, disease, treatment, surgical, stimulation, and clinical outcome parameters.

Introduction

Deep-brain stimulation of the subthalamic nucleus (STN DBS) is an effective treatment for advanced Parkinson’s disease.1–3 The procedure alleviates tremor, rigidity, bradykinesia, and levodopa-induced dyskinesias—the latter probably as a consequence of the reduction in dopaminergic medication after surgery.4,5 STN DBS can also slow disease progression,4 although the procedure’s putative neuroprotective effect is uncertain.6–9

Although the effectiveness of the procedure in the treatment of motor symptoms is accepted, its effect on non-motor symptoms is less clear. The potential dissociability of the motor and neurobehavioural effects is implied by the observation that although health-related quality of life is improved by the procedure,10–13 these salutary effects can be limited to, or be more evident in, physical aspects of quality of life, such as mobility and bodily discomfort.14 Changes in quality of life are moderated by changes in depression,15 which might improve after surgery, occur de novo after deep-brain stimulation, or may be associated with recurrence or exacerbation of a pre-existing condition.16 Depression could be related to deep-brain stimulation, reduced dopaminergic medication after surgery, or psychosocial factors.17,18 Increased recognition that Parkinson’s disease, traditionally regarded as a movement disorder, is also associated with cognitive, behavioural, and emotional changes19,20 correlated with fronto-striatal circuit dysfunction,21,22 has generated interest in the non-motor effects of deep-brain stimulation. Several qualitative reviews of the psychiatric23–25 and cognitive aspects26–32 of deep-brain stimulation are an indication of this interest.

Cognitive and mood disturbances seem to be reported more frequently as side-effects of stimulation of the subthalamic nucleus than of the pallidum.13 This observation could implicate extraneous factors, such as the greater frequency with which the subthalamic nucleus is targeted or greater attention to and awareness of neurobehavioural issues in historically later studies, which tend to assess subthalamic nucleus rather than pallidal stimulation.26 However, direct comparisons between unilateral pallidotomy and bilateral STN DBS27,28 and bilateral pallidal and STN DBS29–31 consistently reveal an increased probability of cognitive and behavioural adverse effects after STN DBS. Whether this finding relates to the subthalamic nucleus being a smaller target with motor, associative, and limbic circuits lying in close proximity to each other, and thus there being a heightened...
risk of electrode misplacement and current spread to non-motor circuits remains speculative.

Empirical details of the effect of STN DBS on cognitive and neurobehavioural functions are only beginning to emerge. In a qualitative review of initial studies of the neuropsychological sequelae of STN DBS and subthalamotomy, Woods and colleagues\textsuperscript{32} concluded that the most consistently reported findings were reductions in verbal fluency and improvements in self-reported symptoms of depression. Indeed, mild to moderate declines in verbal fluency that are persistent and evident even 3 years after surgery are reported in 30\textendash;50\% of patients after the procedure.\textsuperscript{36} Reports of changes in global cognitive functioning, memory, attention, and executive functions are less common and severe cognitive impairments are seen in fewer than 1\textendash;2\% of patients.\textsuperscript{37} By contrast, mild to moderate declines of circumscribed scope are reported in 20\% of patients.\textsuperscript{36} A potential difficulty in the interpretation and reconciliation of discordant findings about the nature and extent of neurobehavioural changes ensuing from the procedure is that a host of factors other than stimulation per se can be associated with cognitive alterations, including, for example, selection criteria for and characteristics of patients, surgical experience, operative complications, comorbid disorders, and medication changes. Furthermore, most studies had small sample sizes and might have had inadequate power to identify effects other than very large postsurgical cognitive changes. Specifically, the median sample size of the studies of STN DBS reviewed by Woods and colleagues\textsuperscript{32} was ten (range one to 63; all single-group, pretest\textendash;post-test designs).

In view of the limitations in sample size and the number of inconsistencies in the published work for the extent and duration of possible changes in episodic memory, attention, executive functions, and verbal fluency after STN DBS, Woods and colleagues\textsuperscript{34} assessed the power of studies that looked at cognitive effects of the procedure. Their analysis showed that only two of 30 studies reviewed had adequate power (above 80\%) to detect large cognitive effects, and that none had sufficient power to detect cognitive changes associated with conventionally small or medium effect sizes. This noteworthy lack of statistical power and attendant high level of type 2 error risk could adversely affect clinical decision-making by potentially overestimating the neurobehavioural safety of the procedure and thus needs addressing.

Until large-scale studies on the cognitive effects of the procedure are published, statistical meta-analyses provide estimates of a population effect size across independent studies in the interim. Meta-analyses increase statistical power to detect true non-zero population effects by lowering the standard error and consequently narrowing the CIs associated with the population effect size estimate.\textsuperscript{34} Hence, a quantitative meta-analysis might enable a better understanding than a qualitative review of the variability and clinical significance of cognitive dysfunction subsequent to STN DBS for Parkinson’s disease. We sought to examine the extent of changes related to STN DBS in eight domains of neuropsychological functioning with a meta-analysis.

Methods
Study selection
The overall objective of study selection was to gather published journal articles that examined neuropsychological functioning before and after STN DBS for treatment of Parkinson’s disease. We did a preliminary article search using MedLine, PsycLIT, and ISI Web of Science electronic databases for the period 1990\textendash;2006. Standard searches were done with key words containing neuropsychological domains or cognition, as well as references to subthalamic nucleus deep brain stimulation or Parkinson’s disease. Key words used for the search included “deep brain stimulation”, “subthalamic”, “neuropsychologic”, “neuropsychological”, “cognition”, “cognitive”, “memory”, “attention”, “executive”, “perceptual organisation”, “verbal fluency”, and “processing speed”. Reference lists of selected articles were visually inspected to locate any cited journal articles that addressed neuropsychological performance before and after STN DBS for treatment of Parkinson’s disease.

Study eligibility criteria
Eligibility criteria for study inclusion consisted of: 1) report of interval or ratio data; 2) neuropsychological evaluation data presented before and after surgery; 3) use of at least one standardised neuropsychological instrument; and 4) sufficient report of study results—eg, mean and SD—to allow for effect size computation. All studies selected for inclusion were English-language publications. Unpublished sources were not considered in the present search.

Data coding
After an initial meeting, two researchers independently extracted the following information from the published articles and coded: 1) number of patients; 2) exclusion criteria; 3) confirmation of target by microelectrode recording; 4) verification of electrode placement via radiographic means; 5) stimulation parameters; 6) assessment time points; 7) assessment measures; 8) whether patients were on levodopa or dopaminomimetics; and 9) summary statistics needed for computation of effect sizes (table 1).\textsuperscript{35\textendash;38} Neurocognitive tests were categorised into the following eight neuropsychological domains: cognitive screening; attention and concentration; problem solving and executive functions; psychomotor speed; verbal functions; visual functions; phonemic fluency; and semantic fluency.\textsuperscript{39} Although most neuropsychological measures assess multiple cognitive functions, each test, consistent with accepted clinical practice, was assigned to the one
domain whose integrity the measure is thought to predominantly reflect (table 2). Such a convention also reduces potential overweighting effects associated with coding a neuropsychological test result across multiple domains. Once these independent classifications were achieved, the two investigators met to resolve coding disagreements. Phonemic and semantic verbal fluency tasks were regarded as indicators of separate domains for the following reasons. First, the most consistently reported findings in the published research of neurocognitive outcomes of STN DBS were reductions in verbal fluency. Second, a recent meta-analysis found that category and letter fluency can be differentially affected in Parkinson’s disease. Third, the mean effect sizes within a combined verbal fluency domain did not meet within-group homogeneity of variance requirement, which assumes that all studies are evaluating the same effect.

### Data analysis
We used the random-effects meta-analytical model. Analysis of continuous outcomes involved comparison of standardised differences in means before and after the

---

### Table 1: Summary of studies included in the meta-analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Exclusion criteria</th>
<th>Microelectrode confirmation</th>
<th>Placement verified</th>
<th>Stimulation parameters</th>
<th>Assessment time points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alegret et al, 2001</td>
<td>35</td>
<td>&gt;75 years, dementia, depression, abnormal scan</td>
<td>Yes</td>
<td>Yes</td>
<td>3.1 V, 60 μsec, 130 Hz</td>
<td>Base, 3 months</td>
</tr>
<tr>
<td>Ardouin et al, 1999</td>
<td>49</td>
<td>Dementia, depression, abnormal scan</td>
<td>Yes</td>
<td>Yes</td>
<td>2.4 V, 60 μsec, 137 Hz</td>
<td>Base, 3-6 months</td>
</tr>
<tr>
<td>Burchiel et al, 1999</td>
<td>5</td>
<td>Dementia, depression, abnormal scan, prior surgery, CNS disease, low IQ</td>
<td>Yes</td>
<td>Yes</td>
<td>2.8 V, 158 μsec, 185 Hz</td>
<td>Base, 3, 6, 12 months</td>
</tr>
<tr>
<td>Buna et al, 2001</td>
<td>3</td>
<td>NR</td>
<td>NR</td>
<td>120 V, 80–200 μsec, 180–200 Hz</td>
<td>Base, stimulation</td>
<td></td>
</tr>
<tr>
<td>Daniele et al, 2003</td>
<td>20</td>
<td>Brain injury, pacemaker, unstable medication, psychiatric history, dementia, cognitive impairment</td>
<td>Yes</td>
<td>Yes</td>
<td>3.6 V, 2.8 V, 2.9 V</td>
<td>Base, 3, 6, 12, 18 months</td>
</tr>
<tr>
<td>De Gaspari et al, 2006</td>
<td>26</td>
<td>Psychosis, cognitive impairment</td>
<td>Yes</td>
<td>Yes</td>
<td>3V, 92.1 μsec, 17/8 Hz</td>
<td>Base, 15-9-3-2 months</td>
</tr>
<tr>
<td>Dujardin et al, 2001</td>
<td>9</td>
<td>NR</td>
<td>Yes</td>
<td>Yes</td>
<td>NR</td>
<td>Base, 3, 12 months</td>
</tr>
<tr>
<td>Erola et al, 2005</td>
<td>19</td>
<td>NR</td>
<td>Yes</td>
<td>Yes</td>
<td>2.7 V, 77 μsec, 171 Hz</td>
<td>Base, 1, 12 months</td>
</tr>
<tr>
<td>Funkiewicz et al, 2003</td>
<td>50</td>
<td>NR</td>
<td>Yes</td>
<td>Yes</td>
<td>3.1 V, 63 μsec, 145 Hz</td>
<td>Base, 3, 12, 24, 36, 48 months</td>
</tr>
<tr>
<td>Funkiewicz et al, 2004</td>
<td>77</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Base, 1 year, 3 years</td>
<td></td>
</tr>
<tr>
<td>Gronell et al, 2003</td>
<td>16</td>
<td>Dementia, major depression, marked cerebral atrophy</td>
<td>Yes</td>
<td>Yes</td>
<td>40–80 μV, 500-1000 μsec, 130 Hz</td>
<td>Base, 6 months</td>
</tr>
<tr>
<td>Hällberg et al, 2004</td>
<td>12</td>
<td>Dementia, depression, low IQ</td>
<td>Yes</td>
<td>Yes</td>
<td>3.2 V, 725 μsec, 144 Hz</td>
<td>Base, stimulation</td>
</tr>
<tr>
<td>Hariz et al, 2000</td>
<td>1</td>
<td>NR</td>
<td>Yes</td>
<td>Yes</td>
<td>3.1 V, 60 μsec, 185 Hz</td>
<td>Base, 1, 2-12 months</td>
</tr>
<tr>
<td>Hilker et al, 2004</td>
<td>8</td>
<td>NR</td>
<td>Yes</td>
<td>Yes</td>
<td>2.7–4.5 V, 60–120 μsec, 130–145 Hz</td>
<td>Base, 2-7 months</td>
</tr>
<tr>
<td>Jahanshahi et al, 2000</td>
<td>7</td>
<td>NR</td>
<td>Yes</td>
<td>Yes</td>
<td>2.8 V, 60 μsec, 147 Hz</td>
<td>Base, stimulation</td>
</tr>
<tr>
<td>Limousin et al, 1998</td>
<td>24</td>
<td>Dementia, abnormal scan, &gt;70 years</td>
<td>Yes</td>
<td>Yes</td>
<td>2.8 V, 60 μsec, 130-185 Hz</td>
<td>Base, 12 months</td>
</tr>
<tr>
<td>Moretti et al, 2003</td>
<td>9</td>
<td>NR</td>
<td>Yes</td>
<td>Yes</td>
<td>2.7 V, 62 μsec, 157 Hz</td>
<td>Base, 1, 6, 12 months</td>
</tr>
<tr>
<td>Moro et al, 1999</td>
<td>7</td>
<td>Pacemaker, mild PD symptoms, unstable drug regimen, dementia, psychiatric disorder, prior surgery</td>
<td>Yes</td>
<td>Yes</td>
<td>2.9 V, 60 μsec, 185 Hz</td>
<td>Base, 9 months</td>
</tr>
<tr>
<td>Morrison et al, 2004</td>
<td>28</td>
<td>Dementia</td>
<td>Yes</td>
<td>Yes</td>
<td>NR</td>
<td>Base, 13.3 +9.7 weeks</td>
</tr>
<tr>
<td>Perozzo et al, 2001</td>
<td>20</td>
<td>Dementia, depression, psychosis, abnormal scan</td>
<td>Yes</td>
<td>Yes</td>
<td>2.9 V, 60 μsec, 144 Hz</td>
<td>Base, 6 months</td>
</tr>
<tr>
<td>Pillon et al, 2000</td>
<td>63</td>
<td>Neurological impairment, abnormal scan, dementia, mood impairment</td>
<td>Yes</td>
<td>Yes</td>
<td>2.4 V, 61 μsec, 137 Hz</td>
<td>Base, 3, 12 months</td>
</tr>
<tr>
<td>Saint-Cyr et al, 2000</td>
<td>11</td>
<td>Dementia, prior surgery, abnormal scan, psychiatric disorder, unstable medical status</td>
<td>Yes</td>
<td>Yes</td>
<td>2.46–3.21 mA</td>
<td>Base, 3-6, 9-12 months</td>
</tr>
<tr>
<td>Schipbach et al, 2005</td>
<td>37</td>
<td>Dementia, &gt;70 years, neurosurgical and neuroradiological complications</td>
<td>Yes</td>
<td>Yes</td>
<td>2.6–2.8 V, 61-65 μsec, 148-157 Hz</td>
<td>Base, 6, 24, 60 months</td>
</tr>
<tr>
<td>Smeding et al, 2005</td>
<td>20</td>
<td>Dementia, abnormal brain scan, unilateral symptoms, psychosis, depression</td>
<td>Yes</td>
<td>Yes</td>
<td>2.3 V, 60 μsec, 145 Hz</td>
<td>Base, 6, 12 months</td>
</tr>
<tr>
<td>Temel et al, 2006</td>
<td>39</td>
<td>Abnormal brain scan, psychosis, affective disorder, cognitive dysfunction</td>
<td>Yes</td>
<td>Yes</td>
<td>2.7–3.3 V, 119.9–131 Μsec, 170–174 Hz</td>
<td>Base, mean 13.6 months</td>
</tr>
<tr>
<td>Trépanier et al, 2000</td>
<td>9</td>
<td>Dementia, medical instability, prior surgery, abnormal scan, psychiatric disorder</td>
<td>Yes</td>
<td>NR</td>
<td>100–200 Hz, 50–100 ms</td>
<td>Base, 3-6 months</td>
</tr>
<tr>
<td>Whelan et al, 2003</td>
<td>5</td>
<td>Dementia, depression, anxiety, abnormal scan, CVA more than mild dysarthria</td>
<td>NR</td>
<td>NR</td>
<td>2.0–3.9 V, 60 μsec, 100-160 Hz</td>
<td>Base, 3 months</td>
</tr>
<tr>
<td>Witt et al, 2004</td>
<td>23</td>
<td>Dementia</td>
<td>NR</td>
<td>NR</td>
<td>139 Hz, 63 μsec, 2.2 V</td>
<td>Base, 6-12 months</td>
</tr>
</tbody>
</table>

IQ=intelligence quotient; NR=not reported; PD=Parkinson's disease; CVA=cerebrovascular accident.
Standardisation allowed the study results to be transformed to a common scale (SD units), which assisted pooling. Adjustments were made to correct for upward bias of effect-size estimation in small sample sizes. An unbiased estimation (Cohen’s d) was calculated for each study, in which the effect size is weighted by a sample-size based constant. Standardised mean differences were calculated and analysed for each study. In particular, \[ d = \frac{(M_h - M_c)}{S} \] where \( M_h \) and \( M_c \) are the mean scores on a neuropsychological test before and after DBS STN, respectively, and \( S \) is the SD for the pooled sample. In studies that did not provide means and SD, \( d \) values were computed from exact p values, t values, or F values. The variance for each \( d \) value was then calculated as \[ \text{variance} = \frac{(n_1 + n_2)}{(n_1n_2)} + \frac{d^2}{2(n_1+n_2)} \] where \( n_1 \) and \( n_2 \) represent the sample sizes before and after the procedure, respectively. The variance function was used to calculate a weighting factor for the unbiased effect size. We used the weighting factor to weight the unbiased effect-size estimate by its sampling error and then divided the result by the sum of the weighted factor for the unbiased effect size. The resulting weighted average composite unbiased effect-size estimate was established for each measure. Following the convention proposed by Cohen, an effect size of 0·20 was regarded as a small effect, 0·50 as a moderate effect, and 0·80 as a large effect.

To test whether the samples had a common underlying effect size, we calculated the homogeneity of the effect size using Cochran’s Q statistic. The appropriateness of a set of neuropsychological tests within a given domain was determined on the basis of the resulting p value from the Q-statistic, which is distributed approximately as \( \chi^2 \) with \( k-1 \) degrees of freedom (\( k \) = number of studies) and tests the null hypothesis that all studies are evaluating the same effect. In general, p values less than 0·10 for the Q-statistic are judged to indicate significant differences across studies.

Effect sizes were disaggregated to define groups sharing a common population effect size. Cochran’s Q statistic was computed by summing the squared deviations of each study’s estimate from the overall meta-analytic estimate, weighting each study’s contribution in the same manner as in the meta-analysis. Subsequently, the Q statistic was compared with a \( \chi^2 \) distribution with \( k-1 \) degrees of freedom. Consistent with a recent meta-analysis that reported category and letter fluency being differentially affected in Parkinson’s disease, results from the Cochran’s Q statistic showed the verbal fluency domain to have heterogeneity of variance. To maintain consistency with earlier findings and to deal with this heterogeneity of variance, verbal fluency was assessed in two stages: first, as a domain (semantic fluency and phonemic fluency analysed together); and then, verbal fluency effect sizes were disaggregated into semantic fluency and phonemic fluency to better explain the variance in terms of separate groups. After this adjustment, results from the Cochran’s Q statistic for both semantic and phonemic verbal fluency revealed homogeneity of variance. A random-effects meta-analytic model was used because it yields increased generalisability of parameter estimates.

We attempted to assess the potential effect on verbal fluency effect sizes of several potential moderators using categorical models. Moderators were selected on the basis of previous research identifying these variables as candidate moderators of cognitive changes. For example, research has suggested that advancing age, stimulation parameters, and reductions in dopaminergic medication after surgery might heighten risk of postoperative cognitive deterioration. Because microelectrode recording can increase risk of haemorrhage and this risk increases with the number of passes made with the electrode, it seems reasonable to

### Table 2: Tests included in each neuropsychological domain

<table>
<thead>
<tr>
<th>Test</th>
<th>n</th>
<th>%</th>
<th>K</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive screen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mini mental state exam</td>
<td>63</td>
<td>19·44</td>
<td>13</td>
<td>9·06*</td>
</tr>
<tr>
<td>Mattis dementia rating scale total</td>
<td>261</td>
<td>80·56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention and concentration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit span forward</td>
<td>73</td>
<td>20·88</td>
<td>21</td>
<td>26·91*</td>
</tr>
<tr>
<td>Digit span backward</td>
<td>64</td>
<td>18·13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual span forward</td>
<td>20</td>
<td>5·67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual span backward</td>
<td>20</td>
<td>5·67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASI Ill arithmetic</td>
<td>8</td>
<td>2·27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brief test of attention</td>
<td>17</td>
<td>4·82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dementia rating scale attention</td>
<td>142</td>
<td>40·23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial sequences</td>
<td>9</td>
<td>2·55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Executive functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop colour/word</td>
<td>78</td>
<td>16·02</td>
<td>43</td>
<td>64·91*</td>
</tr>
<tr>
<td>Trail making test part B</td>
<td>107</td>
<td>21·97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tower of Hanoi</td>
<td>15</td>
<td>3·64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontal assessment battery</td>
<td>76</td>
<td>16·61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin card sorting test</td>
<td>75</td>
<td>15·40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raven’s progressive matrices</td>
<td>57</td>
<td>11·70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baddley’s doors test</td>
<td>9</td>
<td>1·85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alphabetic span test</td>
<td>9</td>
<td>1·85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASAT</td>
<td>20</td>
<td>4·11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tower of London</td>
<td>48</td>
<td>9·86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychomotor speed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail making test part A</td>
<td>42</td>
<td>24·14</td>
<td>9</td>
<td>2·61*</td>
</tr>
<tr>
<td>Visual reaction time</td>
<td>8</td>
<td>4·60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditive reaction time</td>
<td>8</td>
<td>4·60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graphic series</td>
<td>49</td>
<td>28·16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor series</td>
<td>49</td>
<td>28·16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simple reaction time</td>
<td>9</td>
<td>5·17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choice reaction time</td>
<td>9</td>
<td>5·17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rey auditory verbal learning test</td>
<td>43</td>
<td>21·72</td>
<td>16</td>
<td>52·59*</td>
</tr>
<tr>
<td>WMS logical memory</td>
<td>25</td>
<td>12·63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paired associates</td>
<td>26</td>
<td>13·13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memo test</td>
<td>8</td>
<td>4·04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hopkins verbal learning test</td>
<td>17</td>
<td>8·59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selective reminding test</td>
<td>79</td>
<td>39·90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benton visual retention test</td>
<td>8</td>
<td>11·76</td>
<td>9</td>
<td>5·89*</td>
</tr>
<tr>
<td>Rey complex figure test</td>
<td>28</td>
<td>41·18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMS family pictures</td>
<td>17</td>
<td>25·00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judgment of line orientation</td>
<td>15</td>
<td>22·06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonemic fluency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter fluency</td>
<td>355</td>
<td>100·00</td>
<td>16</td>
<td>109·81*</td>
</tr>
<tr>
<td>Semantic fluency</td>
<td>337</td>
<td>100·00</td>
<td>16</td>
<td>37·38*</td>
</tr>
</tbody>
</table>

*p<0·10.*
test whether the number of microelectrode recording tracks is related to verbal fluency decrements. Similarly, because longer disease duration is associated with greater cognitive impairment, and because preoperative cognitive impairment may predispose to postoperative confusion, an analysis was done to determine the effect of disease duration on verbal fluency changes.

Role of the funding source

The funding sources had no role in study design, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The literature search identified 40 studies that had assessed neurobehavioural function before and after STN DBS in Parkinson’s disease. Among these studies, 28 articles met the eligibility criteria for inclusion in the meta-analysis. Table 1 lists sample size, exclusion criteria, stimulation parameters, assessment time points, and whether microelectrode confirmation of the anatomical target and verification of electrode placement was used.

### Table 3: Random effect sizes for the neuropsychological domains

<table>
<thead>
<tr>
<th>Functions assessed</th>
<th>Average random effect size</th>
<th>Effect size variance</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive screening</td>
<td>0.04</td>
<td>0.001</td>
<td>-0.05 to 0.12</td>
</tr>
<tr>
<td>Attention and concentration</td>
<td>0.02</td>
<td>0.001</td>
<td>-0.08 to 0.12</td>
</tr>
<tr>
<td>Executive functions</td>
<td>0.08*</td>
<td>0.001</td>
<td>-0.03 to 0.20</td>
</tr>
<tr>
<td>Psychomotor speed</td>
<td>0.22</td>
<td>0.020</td>
<td>-0.62 to 0.54</td>
</tr>
<tr>
<td>Verbal functions</td>
<td>0.21*</td>
<td>0.020</td>
<td>-0.04 to 0.46</td>
</tr>
<tr>
<td>Visual functions</td>
<td>0.06</td>
<td>0.010</td>
<td>-0.16 to 0.27</td>
</tr>
<tr>
<td>Verbal fluency</td>
<td>0.64*</td>
<td>0.030</td>
<td>0.32 to 0.96</td>
</tr>
<tr>
<td>Phonemic fluency</td>
<td>0.52*</td>
<td>0.080</td>
<td>-0.05 to 1.08</td>
</tr>
<tr>
<td>Semantic fluency</td>
<td>0.73*</td>
<td>0.030</td>
<td>-0.41 to 0.44</td>
</tr>
</tbody>
</table>

*Function was decreased after deep brain stimulation of the subthalamic nucleus for Parkinson’s disease.

### Table 4: Summary of neuropsychological changes reported by studies included in the meta-analysis

<table>
<thead>
<tr>
<th>Functions assessed</th>
<th>Improvements</th>
<th>Declines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burchiel et al, 1999</td>
<td>Motor, GC, mem, VF, VM, A/E, mood</td>
<td>Motor, mood</td>
</tr>
<tr>
<td>Brusa et al, 2001</td>
<td>Attn, A/E, mood</td>
<td>None</td>
</tr>
<tr>
<td>Daniele et al, 2002</td>
<td>Motor, GC, mem, VF, VM A/E, mood</td>
<td>Motor, GC, A/E, mood</td>
</tr>
<tr>
<td>De Gaspari et al, 2006</td>
<td>Motor, A/E, GC, mem, VF, VL, mood</td>
<td>Motor</td>
</tr>
<tr>
<td>Erola et al, 2006</td>
<td>Motor, A/E, GC, VF, PS</td>
<td>Motor</td>
</tr>
<tr>
<td>Funkiewiez et al, 2003</td>
<td>GC, mood, A/E</td>
<td>Mood</td>
</tr>
<tr>
<td>Funkiewiez et al, 2004</td>
<td>VF, mood, GC, attn, mem, A/E, VL</td>
<td>Mood</td>
</tr>
<tr>
<td>Gironell et al, 2003</td>
<td>Mem, attn, arith, problem solving, VS, A/E, lang</td>
<td>None</td>
</tr>
<tr>
<td>Halbig et al, 2004</td>
<td>Mem (declarative and non-declarative), RT, GC, attn, A/E, mood</td>
<td>RT</td>
</tr>
<tr>
<td>Haniz et al, 2000</td>
<td>GC</td>
<td>Motor</td>
</tr>
<tr>
<td>Hikker et al, 2004</td>
<td>GC, mood, mem, A/E, VF, VS, PS</td>
<td>Mem</td>
</tr>
<tr>
<td>Moretti et al, 2003</td>
<td>A/E, VF, attn, mem</td>
<td>None</td>
</tr>
<tr>
<td>Moro et al, 1999</td>
<td>Motor, GC, mem, VF</td>
<td>Motor, mem</td>
</tr>
<tr>
<td>Morrison et al, 2004</td>
<td>Mem, attn, VL, VF, lang, VS, A/E, mood</td>
<td>None</td>
</tr>
<tr>
<td>Peronzo et al, 2001</td>
<td>Motor, GC, mem, A/E</td>
<td>Motor</td>
</tr>
<tr>
<td>Pillon et al, 2000</td>
<td>Motor, GC, VF, mem, A/E</td>
<td>Motor, A/E, mood</td>
</tr>
<tr>
<td>Saint-Cyr et al, 2000</td>
<td>Motor, GC, mem, VF, A/E, mood</td>
<td>Motor, mood</td>
</tr>
<tr>
<td>Schüpbach et al, 2005</td>
<td>Motor, A/E, GC</td>
<td>Motor, A/E</td>
</tr>
<tr>
<td>Temel et al, 2006</td>
<td>Motor, PS</td>
<td>Motor, PS</td>
</tr>
<tr>
<td>Trépanier et al, 2000</td>
<td>Motor, GC, mem, VF, A/E, mood</td>
<td>Motor, Mem, VF, A/E</td>
</tr>
<tr>
<td>Whelan et al, 2003</td>
<td>Lang</td>
<td>Lang</td>
</tr>
<tr>
<td>Witt et al, 2004</td>
<td>Attn, VF, A/E, GC</td>
<td>None</td>
</tr>
</tbody>
</table>

A/E=attention and executive; mem=memory; VF=verbal fluency; VS=visuospatial; GC=global cognitive ability; VM=visuomotor; VL=verbal learning; RT=reaction time; PS=processing speed; attn=attention; arith=arithmetic; lang=language.
for all the included studies. Across studies the maximum combined sample size used for aggregated effect-size calculations was 612 patients.

Results of analyses of neurocognitive domain-specific homogeneity of effect sizes are presented in table 2. Although the subdivision of neurocognitive domains was not completely successful initially (eg, verbal fluency effect size was heterogeneous), removal of outliers (ie, three or more SD from the mean) and further domain subdivision (eg, verbal fluency subdivided into semantic and phonemic fluency) increased homogeneity to eight valid neurocognitive domains.

The average weighted effect sizes were calculated for each of the eight neuropsychological domains, with verbal fluency assessed as a full domain and as subcategories (semantic fluency and phonemic fluency analysed separately). This process involved combining the standardised effect sizes within each domain into a composite-mean weighted effect size, and examining each domain’s significance. Table 3 shows the average weighted effect sizes, standard error of the effect sizes, and confidence limits. Table 4 depicts the neurobehavioural functions that the studies assessed and which of these functions were reported as having improved or declined.

Table 5 shows the information gained from a review of possible moderators of verbal fluency effect sizes across studies. Moderator analysis of the effect of change in levodopa dose (or equivalent, as reported in each study) after STN DBS on fluency did not reveal significant results (phonemic p=0.10; semantic p=0.24). Moderator analysis of the effect of stimulation parameters on fluency did not reveal significant results (amplitude: phonemic p=0.24, semantic p=0.05; pulse width: phonemic p=0.46, semantic p=0.43; frequency: phonemic p=0.23, semantic p=0.28). The number of studies reporting baseline to post-surgical evaluation interval data (31% reported), number of tracks (14% reported), and unipolar versus bipolar stimulation information (58% reported) was too small to meaningfully interpret the correlation value.

There was very little variance in the distribution of age and disease duration. Disease duration values across studies were normally distributed in the case of semantic fluency (mean 13.74 years, SD 2.15, range 7.80–16.20) and phonemic fluency (mean 12.99 years, SD 2.57, range 7.80–16.20). Hence, disease duration is normally distributed and does not necessitate subdivision by cohorts. Similarly, examination of participants’ age within verbal fluency subdomains revealed normally distributed age values for semantic fluency (mean 58.11 years, SD 4.18, range 52.80–68.70) and for phonemic fluency (mean 59.41 years, SD 4.33, range 52.80–68.70).

A simplified moderator analysis controlling for the effects of age and disease duration did not change the results for either semantic (d=0.73) or phonemic (d=0.51) verbal fluency effect sizes after STN DBS. Hence, neither age nor disease duration significantly contributes to verbal fluency decline.

### Discussion

The results of this meta-analysis show that STN DBS in Parkinson’s disease has small effects on all cognitive domains assessed, apart from verbal fluency. However, only declines in the executive and verbal learning and memory domain were statistically significant. More noteworthy declines were identified in semantic (d=0.73) and phonemic verbal fluency (d=0.51) after STN DBS. These effects were closely similar to those described by Cohen as medium or moderate. Thus, STN DBS in Parkinson’s disease seems safe from a cognitive standpoint (bearing in mind that patients in most studies were selected after consideration of a variety of inclusion and exclusion criteria, meaning this conclusion may not be generalisable to unselected patients). Declines in verbal fluency (and thus word-finding facility), however, seem to be common and sizeable.
Findings of this meta-analysis must be interpreted with caution in view of the limitations of meta-analyses in general and the data available for this analysis in particular. Meta-analysis is limited by the quality of studies included, which we attempted to address this issue by using fairly strict study inclusion criteria. As in any review of studies in a given area, studies with non-significant results could be under-reported. The practice of publishing only studies with significant outcomes could create a distortion of the subject under investigation, especially if a meta-analysis is done. A further possible limitation is that a fixed effects meta-analytical model might have been used instead of a random-effects model (used in this analysis) because the latter makes undue distributional assumptions and needs more studies than does traditional meta-analysis. Fixed effects models, however, can be too restrictive. The deciding factor for the use of the random-effects model in the present analysis was that this model tends to yield more generalisable parameter estimates.

Our meta-analysis included studies that were repeated measures designs that used neuropsychological tests as dependent variables, many of which are susceptible to practice effects and other measurement artifacts such as regression to the mean. Although the use of a random-effects model allowed us to treat all of the observed difference between the studies as being due to chance, the small effects observed could indicate, in part, attenuation by significant practice effects. Practice effects are unlikely to have obscured clinically relevant declines in day-to-day function. However, the impact of practice effects should be assessed, and future studies could be enhanced by using alternate forms, assessing unoperated and normal control groups at similar time intervals, and statistically controlling for typical practice and regression to the mean effects.

Another limitation is that, for some variables, meta-analyses were based on relatively few patients. Whereas the effects found in semantic and phonemic fluency were based on samples of 337 and 355 patients, respectively, some other effects (e.g., visuo-perceptual functions) were based on fewer than 70 individuals. The small sample sizes of many studies could reveal the surgical experience at a given treatment centre, and because complications are greater among patients operated earlier than later, cognitive morbidity could be more common or pronounced in small patient series. Use of statistical methods to weight studies by sample size, as done in this study, may only partly reduce such bias. A further issue, by virtue of introducing possible bias to meta-analytical findings, is that when the same investigative team publishes multiple studies it is not always clear whether there is overlap in the samples of different studies. In situations where it appeared clear that two or more studies were drawn from the same or a highly overlapping sample, only the publication with the most complete data was used.

Most studies in our meta-analysis did not have control groups and were not randomised clinical trials, limiting the interpretation that cognitive and verbal fluency declines are directly related to or caused by stimulation. Even though we attempted to identify possible moderators declines in verbal fluency, this was often not possible because necessary information was not reported or was reported in insufficient detail. The lack of association between verbal fluency decline effect sizes and stimulation parameters, change in medication dose, age, and disease duration might be associated with a limited range of values in light of the selection criteria used by most studies. Thus, the findings of this study might not be generalisable to all patients with Parkinson’s disease. Similarly, a host of other factors that could not be directly analysed might moderate verbal fluency declines, including differences in beliefs about best practices for stimulator programming among different treatment centres, timing of adjustments, and dopaminergic medication reduction after STN DBS. Indeed, dopaminergic medication reductions after surgery have been linked to behavioural complications such as apathy and depression, and such complications could adversely affect verbal fluency and cognition.

Caution is also invited in interpreting the clinical significance of what are statistically moderate verbal fluency declines and small declines in other cognitive domains. Specifically, in Cohen’s effect-size classification system, somewhat arbitrary distinctions are made between magnitudes. Hence, a statistical consideration of data might describe 0.20 as a small but statistically notable effect size, statistical and clinical significance are not synonymous and an effect size is not fully informative for clinical interpretation. For example, while a small effect size might be extremely important in revealing life-threatening treatment side-effects, even large effect sizes might be of clinically modest significance when they refer to less important manifestations. Furthermore, the verbal fluency changes after STN DBS, even if only moderate, could represent declines from an already compromised state. In view of the fact that Henry and Crawford reported small-to-medium verbal fluency effect sizes in patients with Parkinson’s disease without dementia, a modest postsurgical decline could propel patients from borderline or mild impairment into the moderate-to-severe range of verbal fluency impairment.

Of course, the importance of adverse events is to some extent subjective. One facet of the clinical importance of an effect is the extent to which it affects quality of life. The mostly subtle declines in neurocognitive domains after STN DBS should, seemingly, not interfere with everyday functioning (assuming the decline does not propel a patient beyond the threshold associated with disability or handicap). However, moderate verbal fluency decrements can affect activities of daily living and quality of life. Although Drapier and co-workers reported only a
non-significant decline in communication on a quality of life measure after STN DBS, how this might relate to changes in verbal fluency (or their cause) is unclear, and studies are needed to assess whether reductions in verbal fluency significantly detract from quality of life and, specifically, patient satisfaction with communication.

Plausible neurocognitive mechanisms for verbal fluency declines after STN DBS might strengthen notions that these changes are related to the procedure rather than an extraneous or confounding variable. The neuroanatomical and cognitive mechanisms underlying a robust decline in verbal fluency after STN DBS are a matter of speculation and are not mutually exclusive. Recent intraoperative electrical stimulation studies during tumour resection suggest that the striatum might have dissociable roles in the motor and cognitive control of language. Consistent with this position, if motor speech mechanisms are postulated to underlie verbal fluency decrements, it would be sufficient to posit an effect of STN DBS on corticothalamostral motor circuits that are parallel to, but segregated from, associative (cognitive) and limbic (emotion and motivation) circuits according to dominant theoretical models. By contrast, if cognitive and more specifically semantic or executive mechanisms are postulated to be fundamental to verbal fluency changes induced by deep-brain stimulation, stimulation would need to spread beyond the motor circuit, active electrode contacts would need to be placed outside the putative motor area, different stimulation patterns would have to affect different basal ganglia structures and cortical regions, or basal ganglia circuits would be more open and interconnected than held by accepted models.

Evidence to date, albeit largely indirect, probably favours a non-motor speech explanation for verbal fluency changes. Arguing against a motor speech explanation is the fact that motor speech deterioration seems paradoxical in view of the motor improvements seen with deep-brain stimulation. Indeed, most studies reported improvement or no change in dysarthria with STN DBS and improvements in dysarthria are related to normalisation of cerebral metabolic patterns associated with speech activation, a finding paralleling that of normalisation of cortical metabolism in good motor responders but not non-responders to deep-brain stimulation. When negative effects on motor speech do occur, they might be related to misplacement of electrodes or stimulation at suboptimum parameters, dyskinesias related to medication and stimulation interactions, or an imbalance between right and left stimulation. Indeed, Törnvist and colleagues have shown that with use of typical stimulation settings there was no difference ‘on’ and ‘off’ stimulation in speech intelligibility, but that intelligibility declined with higher stimulation frequencies and amplitudes.

Further support for a cognitive rather than motor mechanism underlying verbal fluency changes comes from a study with seven patients undergoing PET while carrying out verbal fluency tasks with and without stimulation of the subthalamic nucleus. Whereas motor function improved with stimulation, verbal fluency performance declined by 15%. Additionally, verbal fluency differences between on and off stimulation were correlated with regional cerebral blood flow activation (verbal fluency-counting) decrements during on versus off stimulation in the left inferior frontal and temporal gyri. Other evidence supportive of cognitive and linguistic mechanisms underlying verbal fluency decrements after deep-brain stimulation are the findings that: STN DBS affects semantic processing; motor speech decrements affects performance on a range of expressive language tasks, yet verbal fluency decrements may be specific in that they can be accompanied by improvements on other language tasks, including visual confrontation naming and reductions in verbal fluency after pallidal surgery are associated with diminution of patients’ efficiency in switching between lexical-semantic categories during word search and retrieval, thus implicating specific cognitive mechanisms in verbal fluency deterioration.

Our study findings have several implications for future research concerning neurobehavioral effects of deep-brain stimulation. The average effect sizes determined in this study suggest that for studies to have adequate power (above 80%) to detect even only the most prominent neurocognitive effect of STN DBS (using the hitherto predominant single group, repeated measure design, and two-tailed tests with alpha set at 0.05), they would need a minimum sample size of 48 patients. Obviously, this is a minimum standard, and adequate assessment of neuropsychological effects, at least using instruments applied to Parkinson’s disease thus far, would ideally involve samples much larger than this. Thus, while future small-sample studies that identify significant effects would be of interest, studies with negative findings will probably be of interest only if they are adequately powered.

Another issue is that research groups need to reach consensus about critical variables that should be examined as possible risk factors for cognitive declines in multicentre studies. Attempts to identify via moderator analyses some factors that may play a part in cognitive decline were unsuccessful because mean values of potential moderator variables were too narrow in range to allow meaningful analyses or not adequately reported. Thus, future studies may seek to examine cognitive outcomes in subgroups of patients defined by different values of potential risk factors. Additionally, studies should seek uniformity in reporting in detail various patient, disease, treatment, and surgical procedural variables. For example, it may be critical to identify the exact location of the active contacts (although this itself is beset by methodological controversy), the percentages of patients showing changes on clinical outcome measures
of a given magnitude, and caseness (meaning the number of patients belonging to a diagnostic group, such as depression, before and after surgery), and the relationship of these outcomes to cognitive outcome. Such reporting is anticipated to help with the identification of factors that underlie cognitive morbidity.

Because cognitive changes, even if only mild to moderate in size, can occur after deep-brain stimulation, it is important to undertake preoperative and postoperative neuropsychological assessments. As noted in a recent consensus report,27 preoperative evaluation may be two tiered for efficiency. Cognitive screening instruments can be used to identify people with severe cognitive impairment for whom deep-brain stimulation might not be an appropriate treatment. Full neuropsychological assessment should be undertaken at the second step. Cognitive screening instruments are unlikely to be sufficiently sensitive to postoperative changes, as is evident from the effect sizes for screening measures found in this meta-analysis. Screening instruments tend to have very limited coverage of cognitive function and are subject to ceiling effects resulting in difficulty differentiating moderate from high functioning.

By conclusion, given the currently available data, STN DBS seems to be relatively safe from a cognitive standpoint in carefully selected patients. STN DBS can affect neurocognitive functioning, but the effects are statistically of small to moderate size and typically circumscribed. Whether the cognitive decrements are directly related to stimulation, the implantation procedure, or some other factor remains to be elucidated, as do the risk factors for such declines. The meta-analytical findings accord with those of qualitative reviews showing that the most pronounced effects are seen in semantic and phonemic fluency. Furthermore, this meta-analysis extends the existing published work through facilitation of a better understanding of the variability and clinical significance of cognitive dysfunction subsequent to STN DBS for Parkinson’s disease. Our findings draw attention to the fact that small declines in verbal and executive functioning can occur. There is a need for additional well-designed and adequately powered studies investigating the neurocognitive sequelae of STN DBS for Parkinson’s disease, more extensive and uniform reporting of data, and for meta-analysis of the effects of STN DBS on emotion, affect, and quality of life.

References


69 Lipsey MW, Wilson DB. The way in which intervention studies have “personality” and why it is important to meta-analysis. Eval Health Prof 2001; 24: 236–54.


