Levodopa and Subthalamic Deep Brain Stimulation Responses are not Congruent

Adam Zaidel, MSc,1,2* Hagai Bergman, MD, PhD,1,2 Ya’acov Ritov, PhD,3 and Zvi Israel, MD 4

1The Interdisciplinary Center for Neural Computation, The Hebrew University of Jerusalem, Israel
2Department of Medical Neurobiology, Canada-Israel Institute for Medical Research, The Hebrew University Hadassah Medical School, Jerusalem, Israel
3Department of Statistics, The Hebrew University of Jerusalem, Israel
4Department of Neurosurgery, Hadassah University Hospital, Jerusalem, Israel

Abstract: There is a consensus that in Parkinson’s disease, the extent of preoperative levodopa responsiveness predicts the efficacy of subthalamic nucleus deep brain stimulation (STN DBS). However, this may be the result of statistical methods and primary assumptions. We were able to reproduce previously published correlation results on our data (N = 49 patients). Yet, these same results were demonstrated even after random shuffling of our data. Notably, we did not observe a correlation between STN DBS efficacy and preoperative levodopa responsiveness when using their respective baselines and fractional scores of motor improvement. Furthermore, postoperative responses were not limited by preoperative scores, with tremor demonstrating the greatest discrepancy. We conclude that preoperative levodopa responsiveness does not predict or limit the outcome of STN DBS. These results imply different therapeutic mechanisms for levodopa and STN DBS and therefore question the validity of using substantial preoperative levodopa responsiveness as a selection criterion for STN DBS.

INTRODUCTION

Classical models of the basal ganglia1,2 suggest that, both dopamine replacement therapy and subthalamic nucleus (STN) inactivation ameliorate Parkinson’s disease (PD) symptoms by reducing the pathologically increased discharge rate of the basal ganglia output nuclei. In line with this, it has become widely accepted that in PD, the extent of preoperative levodopa (L-dopa) responsiveness predicts the extent of responsiveness to STN deep brain stimulation (DBS).3–6 Thus, a substantial response to L-dopa has become an accepted selection criterion for STN DBS. In contrast, recent publications have shown that preoperative L-dopa responsiveness does not predict the outcome of long term STN DBS,7,8 and there are occasional reports of responsiveness to STN DBS despite preoperative unresponsiveness to L-dopa.9,10

The different clinical effects of STN DBS vs. L-dopa11–14 also seem to suggest that their therapeutic mechanisms may not be fully congruent. Therefore, although L-dopa responsiveness may be a good indication that the patient has idiopathic PD, the extent of the response may not predict the outcome of STN DBS. In this manuscript, we reassess the notion of a strong correlation between preoperative L-dopa motor responsiveness and efficacy of STN DBS,3–6 by investigating the statistical methods used to date and analysis of data from 49 PD patients who underwent STN DBS at our center. We have not attempted to analyze all the predictive factors of STN DBS outcome; rather we focus on correlation analysis between preoperative L-dopa responsiveness and STN DBS efficacy – the same tool by which a predictive capacity has been attributed to L-dopa responsiveness.3–6
METHODS

Patients and Surgery

Fifty-one PD patients undergoing bilateral STN DBS implantation were included in this study. Two patients with a prior pallidotomy were excluded due to the possibility that their response to treatment could be biased. Clinical details of the remaining 49 patients are presented in Supporting Information Table S1. All patients met accepted selection criteria for STN DBS and signed informed consent for surgery with microelectrode recording. This study was authorized and approved by the Institutional Review Board of Hadassah University Hospital in accordance with the Declaration of Helsinki.

Surgery was performed using the CRW stereotactic frame (Radionics, Burlington, MA). STN target coordinates were chosen as a composite of indirect anterior commissure—posterior commissure atlas based location and direct T2 magnetic resonance imaging (MRI), using Framelink 4 or Framelink 5 software (Medtronic, Minneapolis, MN). The positions of the implanted STN DBS electrodes were verified by intraoperative physiological recording and postoperative computerized tomography (CT) fused with the preoperative MRI (also using Framelink software).

Clinical Testing

The third subsection (motor score) of the unified PD rating scale (UPDRS-III) was assessed preoperatively both off and on dopaminergic medication (PREoff and PREmed respectively). Postoperatively, UPDRS-III was assessed in four scenarios: (i) off-medication; off-stimulation (ii) off-medication; on-stimulation (iii) on-medication; off-stimulation, and (iv) on-medication; on-stimulation (POSToff, POSTstim, POSTmed, and POSTboth, respectively).

Patients were off-medications for >12 hour for PREoff, POSToff, and POSTstim testing, and off-stimulation for > 1/2 hour for POSToff and POSTmed testing. PREmed testing was performed on a suprathreshold (25% above the regular) dose of antiparkinsonian medication, whereas POSTmed and POSTboth were tested on the regular postoperative dose (generally 30–90 minute after oral intake, depending on the individual patient). Postoperative testing was 17 ± 9 months (median ± median absolute deviation) after surgery. Of the 49 patients, 4 did not have POSToff scores, 1 did not have POSTboth scores, and 16 did not have POSTmed scores. POSTmed scores were not used much in this manuscript, hence the number of patients for any given comparison was N= 49, 48, 45, or 44, depending on whether neither scores, POSTboth but not POSToff, POSToff but not POSTboth or both scores were required respectively.

To compare specific symptom improvement, the UPDRS-III (motor) section was broken up into 4 composite symptom scores as follows:

i. Tremor – sections 20–21 (max 28),
ii. Rigidity – section 22 (max 20),
iii. Limb bradykinesia – sections 23–26 (max 32), and

Statistical Methods

To assess different analysis methods, we compared correlation results from our data to those after random shuffling of the data. In shuffling, the preoperative scores were randomly relinked to the postoperative scores. This was repeated 100 times to generate a distribution of correlation results for shuffled data. Additional analysis of simulated random data is presented as Supporting Material. Simulation and analysis were performed with custom developed software, using MATLAB7.1 (The Mathworks, Natick, MA).

RESULTS

A Common Baseline in Calculating Preoperative and Postoperative Responsiveness can Yield a False Correlation Between the Two

To the best of our knowledge, all the studies that have found a correlation between preoperative responsiveness to L-dopa and responsiveness to STN DBS have used the PREoff score as the baseline also for measuring responsiveness to STN DBS. Interestingly, when reported the mean POSToff UPDRS-III was lower than PREoff (also seen in our data, P < 0.05; Supporting Information Table S1), possibly due to insertion, microlesion effects, or residual effects of DBS. In that case, on average, a higher postoperative motor improvement may be calculated when using PREoff vs. POSToff as a baseline.

In itself, using PREoff as a baseline for assessment of STN DBS efficacy is legitimate. However, a problem arises when correlating this result with the preoperative motor improvement from L-dopa, as presented in Figure 1A. This is because both measures use the identical score (PREoff) as a baseline. The resulting comparison:
PREoff – PREmed vs. PREoff – POSTboth \( (1) \) will tend to yield a positive correlation, as it correlates an entity (PREoff) with itself. For the comparison represented by Eq. (1), even randomly shuffled UPDRS-III data (see methods section) yields a high correlation with \( P < 0.0001 \) (Fig. 1B). The distribution of correlation coefficients for randomly shuffled data indicates that this statistical artifact is strong and consistent (Fig. 1C).

In addition, using PREoff scores, which were acquired at variable periods before postoperative testing, as a baseline (despite microlesion effects, disease progression, or other changes over time) would seem less optimal in quantifying STN DBS treatment responsiveness than using POSToff scores, which were assessed in the same session as the postoperative “on” scores. We therefore propose using PREoff scores as a baseline in calculating preoperative L-dopa responsiveness and POSToff scores as a baseline in calculating postoperative DBS responsiveness.

The Absolute, but not Relative, Improvement in UPDRS-III Score Correlated with the Baseline

A straightforward approach to quantifying improvement due to treatment may be to simply calculate the absolute difference between the “off-treatment” and “on treatment” UPDRS-III motor section scores \( [\text{as in Eq. (1)}] \). We term this the absolute or “additive” approach, because it indicates that only the magnitude of improvement is relevant, whereas the baseline (off-treatment) score is in itself irrelevant, for example, improvement from a baseline score of 90 down to 70 is equivalent to improvement from 20 down to 0, as both improved by 20 points. In contrast, a multiplicative approach would calculate improvement relative to the
baseline score, for example, improvement from a score of 90 to 45 is equivalent to improvement from a score of 20 to 10, as both present an improvement of 50%.

In general, a good measure of improvement should correlate as little as possible with the baseline score. In our data, there was a strong correlation between absolute UPDRS-III improvement (the additive “approach”) and the respective baseline score (Fig. 2, left column; \( P < 0.0001 \)). Although the “multiplicative” approach showed a slight negative correlation between preoperative improvement and the baseline score (\( P < 0.05 \)), no correlation (\( P > 0.1 \)) was seen for postoperative improvement (Fig. 2, right column). We therefore propose using relative improvement for comparing preoperative vs. postoperative improvement (each using their respective baseline scores, as explained in the previous section) as follows:

\[
\begin{align*}
\text{Absolute} & : \frac{\text{PREoff} - \text{PREmed}}{\text{PREoff}} \\
\text{Relative} & : \frac{\text{POSToff} - \text{POSTboth}}{\text{POSToff}}
\end{align*}
\]

(2)

It should be noted, however, that even for Eq. (2), correlation could result from the expected correlation
between PREoff and POSToff or between PREmed and POSTboth, hence a correlative result should be treated with caution. However, a negative result (i.e. of no correlation) is an indication that, there is no linear relation between the preoperative improvement because of L-dopa and the postoperative improvement due to STN DBS.

No Correlation Between Preoperative Levodopa vs. DBS Responsiveness

Based on the data from the 49 patients included in this study, we could reproduce the problematic correlation as described by Eq. (1) for both the real and shuffled data (Fig. 1A,B). However, the relative responsiveness to preoperative L-dopa did not correlate with
the relative responsiveness to STN DBS [Eq. (2)] for either the real or the shuffled data (Fig. 1D,E). The correlation coefficients of the real data clearly lay well within the distribution for shuffled data (Fig. 1A,D).

The results presented in Figure 1F is based on POSTboth values, that is, assessing postoperative improvement due to medication and DBS. Additional correlations using POSTboth can be seen in Figure 3 (right column). Similar results were observed when using POSTstim, that is, improvement from stimulation alone (left column). When using PREoff as a baseline for both preoperative and postoperative improvement [as in Eq. (1)], but comparing relative (as opposed to absolute) UPDRS-III reduction, no correlation was seen in our data (Fig. 3C,D). In addition, no significant correlation was found when comparing absolute UPDRS-III reduction, but using POSToff as the postoperative baseline (Fig. 3E,F). These results oppose the notion that UPDRS-III reduction, because of STN DBS correlates with preoperative UPDRS-III reduction from L-dopa. When simulating random data (Supporting Information) similar results were seen (Supporting Information Fig. S1).

To test whether our observation of no correlation could result from a slightly longer period between surgery and postoperative testing (17 ± 6 months; median ± MAD), we divided the patients into two groups: (i) those tested within one year of surgery and (ii) those tested after one year. Both groups still showed no correlation between preoperative L-dopa responsiveness and DBS responsiveness (R = −0.23, P = 0.41, N = 15 for the first group; R = 0.19, P = 0.32, N = 29 for the second group). We conclude that the lack of correlation is probably due to statistical methods and not duration since surgery.

Responsiveness was Greater to DBS in Conjunction with Levodopa vs. Preoperative Levodopa

Despite possible disease progression, the mean POSTboth score (i.e. on-medication and -stimulation) was significantly lower vs. PREmed (P = 0.002; Supporting Information Table S1), and the majority of patients had better POSTboth vs. PREmed scores (POSTboth ≤ PREmed for 71% of the patients, supporting information Figure S2A, bottom plot; additional preoperative vs. postoperative “on” score analysis is presented in the Supporting Information). In addition, we found that the relative responsiveness to DBS and medication was better than preoperative L-dopa responsiveness (Fig. 4A). This was despite observations that the POSToff baseline was lower than the PREoff baseline (as explained above, a lower baseline can reduce the measure of improvement) and the postoperative L-dopa equivalent daily dose (LEDD) was significantly lower than the preoperative LEDD (P < 0.0001; Supporting Information S1). When assessing symptoms separately, it was seen that tremor demonstrated the greatest discrepancy between postoperative vs. preoperative responsiveness, followed by rigidity.
and limb bradykinesia (Fig. 4B). These findings further support the notion that individual responsiveness to STN DBS is not limited by the degree of preoperative responsiveness to L-dopa.

Finally, responsiveness from stimulation alone (off-medication) of specific symptoms did not correlate with the symptom’s preoperative L-dopa responsiveness (Table 1). However, with the addition of medication (i.e., both stimulation and medication), significant correlations were seen for rigidity and limb bradykinesia, with a similar trend observed for tremor. This effect seemed to remain even after partial correlation analysis (given the response to DBS alone, Table 1). The finding of significant correlations for particular symptoms on DBS and medication, but not DBS alone, further supports the claim that the mechanisms of these treatments are not congruent.

**DISCUSSION**

We believe that the prevailing notion of a strong correlation between preoperative L-dopa responsiveness and benefit from STN DBS may be the result of methods used in statistical analysis and therefore requires reassessment. According to our data, the preoperative L-dopa responsiveness of PD patients does not predict the outcome of STN DBS. However, due to the limitations of our data, such as possible residual effects of DBS on POSToff (which was used as a baseline) and differences between the preoperative and postoperative L-dopa challenge, further studies are required to validate this finding. Such studies should take into account specific parkinsonian symptoms which, as we have shown (Fig. 4 and Table 1, in line with previous studies), respond differently to DBS and L-dopa.

Piboolnurak et al. found that STN DBS benefit at 3 and at 5 year did not correlate with preoperative L-dopa responsiveness. They explained the discrepancy between their results and other studies which did find a correlation as probably being due to their long term vs. the other’s short-term postoperative testing. However, as they noted in their discussion, unlike previous studies they used POSToff as a baseline. They also compared relative improvement (essentially, their comparison was identical to ours, i.e., Eq. (2)). Hence, we believe that their finding of no correlation was probably due to their unbiased statistical methods. We did not find a difference when comparing short vs. long term postoperative testing in our data.

Although specific symptoms did respond postoperatively in accordance with their preoperative responsiveness (rigidity and limb bradykinesia), this was only on DBS in conjunction with medication. On DBS alone there was no correlation. It would therefore seem logical to conclude that the correlation was due to L-dopa, however DBS itself seems to employ other or additional mechanisms. This could be due to STN DBS not having an effect on the “direct pathway” of the basal ganglia or activation of STN-neighboring axonal pathways. It is possible that the abovementioned correlations have led to the clinical impression that responsiveness to DBS correlates with preoperative responsiveness to L-dopa in the first place.

Correlation between POSTstim and PREmed motor scores seems to be a consistent finding and is also seen in our data (Supporting Information). This correlation, however, should not be confused with a correlation between responsiveness to DBS and preoperative L-dopa responsiveness. The former correlation compares the state of the patient (UPDRS-III score on-treatment), whereas the latter compares improvement (i.e., reduction in the baseline UPDRS-III score as a result of treatment). It should not be surprising to see a correlation between the preoperative and postoperative patient’s “on” state (no matter what the treatment), as the patient (and disease severity) is a unifying factor.

**TABLE 1.** Symptom specific L-dopa responsiveness does not predict STN DBS responsiveness, unless in conjunction with medication

<table>
<thead>
<tr>
<th>(A) DBS alone no-medication</th>
<th>(B) DBS and medication</th>
<th>(C) Partial correlation of (B) given (A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>P</td>
</tr>
<tr>
<td>Tremor</td>
<td>0.15</td>
<td>0.403</td>
</tr>
<tr>
<td>Rigidity</td>
<td>0.24</td>
<td>0.149</td>
</tr>
<tr>
<td>Limb bradykinesia</td>
<td>0.20</td>
<td>0.225</td>
</tr>
<tr>
<td>Axial bradykinesia</td>
<td>−0.08</td>
<td>0.614</td>
</tr>
</tbody>
</table>

(A) On DBS alone (without L-dopa), the relative preoperative responsiveness of particular symptoms did not correlate with their relative response to STN DBS.

(B) With the addition of L-dopa medication, rigidity and limb bradykinesia demonstrated significant correlations ($P < 0.05$, marked by shaded region) and tremor demonstrated a trend ($P = 0.08$). The lack of significance in linear correlation for tremor was probably due to ceiling effects, as on DBS and medication tremor demonstrated 100% improvement in the majority of cases.

(C) For the significant scores in B (rigidity and limb bradykinesia), partial correlations were calculated between postoperative responsiveness on both DBS and medication, given the responsiveness of DBS alone, vs. preoperative responsiveness. For both symptoms, correlations remained significant. R and P are the correlation coefficient and p-statistic, respectively, of the correlations between preoperative and postoperative responsiveness. N/A – not applicable.
Therefore, to compare different treatments, a fractional comparison of responsiveness is required. Finally, our findings that POST both scores were better than PREmed scores, and that postoperative responsiveness to DBS and medication was better than preoperative L-dopa responsiveness are in line with previous publications. These consensus observations strongly indicate that individual responsiveness to DBS is not limited by the degree of preoperative responsiveness to L-dopa, and that DBS and L-dopa mechanisms are not congruent. Thus, our results question the validity of L-dopa, and that DBS and L-dopa mechanisms are not limited by the degree of preoperative responsiveness to L-dopa. The latter three also provided guidance, critique, and review during surgery and Zvi Israel the neurosurgeon. Ya’acov Ritov provided advice and critique of the statistical methods.

Acknowledgments: We would like to thank Dr. E. Linet-sky and E. Serri for UPDRS testing and T. Moller for help in collecting patient data. Financial Disclosures: This research was supported in part by the “Fighting against Parkinson” and the Max Vorst Family Foundations of the Hebrew University Netherlands Association (HUNA) to H. Bergman and the Parkinson’s at the Hadassah (PATH) committee of London to ZI.

Author Roles: Adam Zaidel did the analysis and wrote the manuscript. Hagai Bergman was the neurophysiologist during surgery and Zvi Israel the neurosurgeon. Ya’acov Ritov provided advice and critique of the statistical methods. This research was supported in part by the “Fighting against Parkinson” and the Max Vorst Family Foundations of the Hebrew University Netherlands Association (HUNA) to H. Bergman and the Parkinson’s at the Hadassah (PATH) committee of London to ZI.

REFERENCES