Evidence of possible bias in pre-surgical assessment of Parkinson's disease using levodopa challenges

Authors: G. Morinan, J. O'Keeffe
Affiliation: Machine Medicine Technologies

Introduction: Parkinson's disease (PD) patients who no longer reliably respond to medication may be treated with deep brain stimulation (DBS). Whether DBS is appropriate for a given patient is assessed, in part, by a levodopa challenge, in which a patient will undergo the MDS-Unified PD Rating Scale (MDS-UPDRS) part-3 assessments before and after taking levodopa. However, a significant fraction of patients do not benefit sufficiently from DBS to justify the attendant surgical risk and cost of the procedure. This suggests the process of selecting patients for DBS intervention could be improved.

Because motor requires subjective ratings of disease severity, by unblinded clinicians, a rater bias could result in a patient receiving, or being denied, treatment inappropriately. We investigated whether there is an identifiable bias in levodopa challenges when conducted for the purpose of surgical pre-assessment, as compared to levodopa challenges conducted during clinical trials that do not involve patients being considered for surgical treatment.

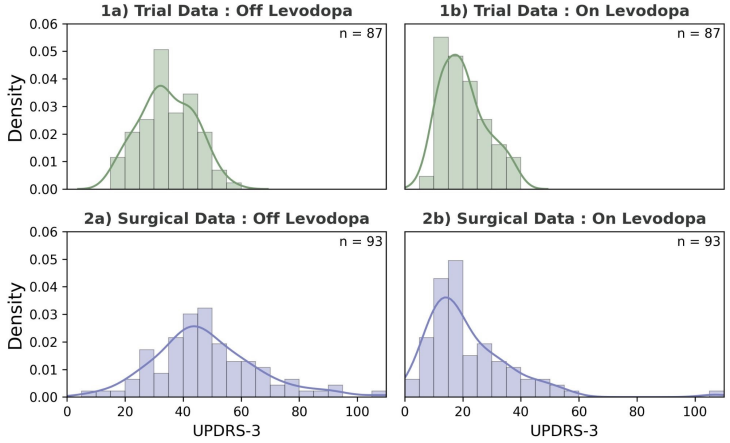

Figure 1: Distribution of UPDRS-3 scores for levodopa challenges, whereby patients are assessed once in the off-state (left) and once in the on-state (right). The dataset includes levodopa challenges conducted as part of clinical trials (upper, n=87), as well as those for pre-surgical assessment (lower, n=93).

Method: The dataset consists of 180 levodopa challenges on 146 different patients, conducted at six different sites and by a variety of assessors, collected using the KELVIN-PD™ platform. Of these, 93 were conducted for pre-surgical assessment for DBS, while the remaining 87 were conducted as part of a pharmaceutical clinical trial (see Figure 1).

The null hypothesis was that these two groups would see a similar level of improvement, as measured by the improvement in UPDRS part-3 score between the 'off levodopa' and 'on levodopa' assessments. We tested this hypothesis by two methods.

Firstly, we considered whether a patient was undergoing pre-surgical assessment as a boolean variable (1 meaning they are, 0 otherwise), alongside the 'off levodopa' rating, within a linear regression model to predict the 'on levodopa' rating. The coefficient of this boolean being significantly different from zero would be consistent with the hypothesis that there exists a systematic bias when PD patients are assessed for DBS surgery.

Secondly, we computed the distribution of percentage improvements for both groups, and performed a Mann-Whitney test for difference in distributions. The Mann-Whitney test was chosen because as a non-parametric method for testing whether one distribution's mean is significantly greater than another.

For this second analysis we controlled for baseline disease severity (off levodopa UPDRS part-3), because we observed that percentage improvement had a weakly significant correlation with baseline rating (Pearson's r = -0.149, p-value = 0.046). We controlled for this by conducting the analysis on a subset of 80 levodopa challenges, selected such that the distribution of baseline rating among patients was identical between the surgical and pharmaceutical trial groups (see Figure 2).

Result: The average improvement in UPDRS part-3 score was 40.9% among the pharmaceutical trial group, and 55.1% among the surgical assessment group. Our first test found the linear regression model coefficient of the surgical assessment boolean to be significantly different from zero ($F_{1,178}$ = 10.2 , p < 0.01), with a mean estimated value of -5.3 (see Table 1). Our second test found the difference between the distributions of these percentage improvements to be highly significant (Mann-Whitney's U = 552, p-value < 0.01).

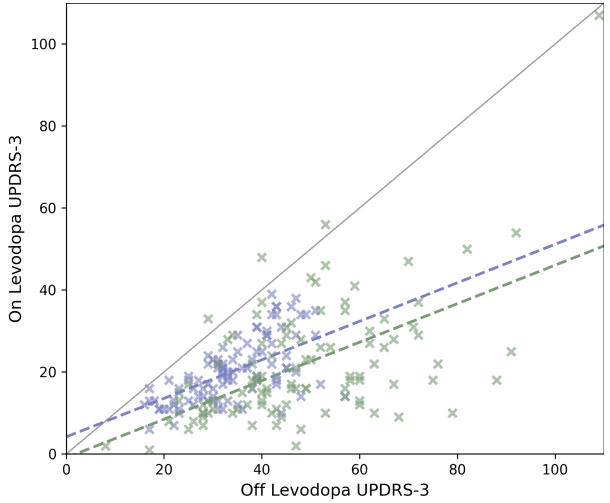

Figure 2: Scatterplot of off-state and on-state UPDRS-3 scores for the trial group (green, n=87) and the surgical group (blue, n=93). The estimates of the regression model for each off-state score are plotted for the situation where the surgical Boolean is false (green line) and where it is true (blue line). The distance between these lines being the size of the surgical Boolean.

Variable	Coefficient	Std. Error	t	P > |t|	Lower bound (2.5%)	Upper bound (97.5%)
Off-medication UPDRS part-3	0.5617	0.027	20.833	0.000	0.508	0.615
Surgical Boolean	-5.3061	1.661	-3.194	0.002	-8.585	-2.027

Table 1: Summary of the regression model to estimate on-state UPDRS-3, using two inputs; off-state UPDRS-3 and a Boolean variable indicating whether the patient is undergoing a pre-surgical assessment. The coefficient of this surgical Boolean was found to be significantly different from zero ($F_{1,178}$ = 10.2 , p < 0.01), with an a mean estimated value of -5.3, indicating an effect size equal to roughly 5.3 points on the UPDRS-3.

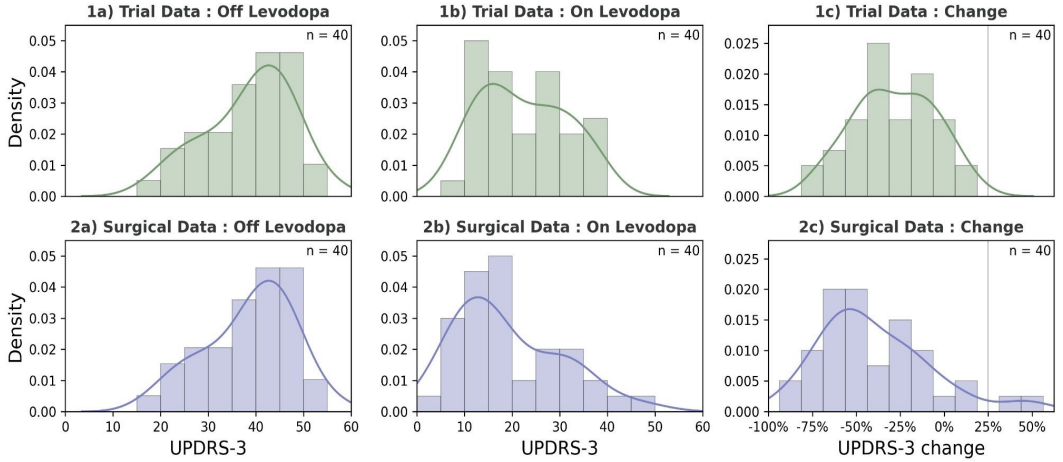

Figure 3: Distribution of UPDRS-3 scores during the off-state (left), the on-state (middle) and change in UPDRS-3 between off and on state (right). This subset of the data was chosen such that the distribution of off-state UPDRS-3 for the trial group (upper, n=40) matched precisely the surgical group (lower, n=40). Comparing the distribution of UPDRS-3 change for trial patients (mean 40.9%) with that of the surgical group (mean 55.1%) found a highly significant difference (Mann-Whitney's U = 552, p-value < 0.01).

Conclusion:

While further work is required to exclude the possibility of systematic differences between pre-surgical and other patients, these results are consistent with there being a clinically significant bias in levodopa assessments when conducted on patients being considered for DBS, and that this bias equates to approximately a difference of 5.3 on the UPDRS part-3.

Should further work bear this result out, it would suggest that extra care should be given to objectively assess patients who are being considered for DBS.

Blinded rating might go some way towards removing any bias, although this may be logistically challenging. Alternatively, true objectivity may be achieved through algorithmic measurement of motor function.

Contact: jonathan@machinemedicine.com