**Supplementary material of manuscript ‘A machine learning approach to predict motor response to subthalamic nucleus stimulation in Parkinson disease patients’ by Habets et al.**

**Legend**

1. Literature on clinically relevant UPDRS differences
2. Imputation methods
3. References
4. **Literature on clinically relevant UPDRS differences**

|  |  |  |
| --- | --- | --- |
| **Reference** | **Compared (MDS-) UPDRS score** | **Sign difference** |
| Kostoglu1 | III Med-OFF vs. Med-OFF + STIM | 38% |
| Horvath2 | III | - 3.25+ 4.63 |
| Shulman3 | IIITotal | +/- 2.5 (minimal)+/- 5.2 (moderate)+/- 10.8 (large)+/- 4.3 (minimal)+/- 9.1 (moderate)+/- 17.1 (large) |
| Schrag4 | IIIII (H&Y > 2) | +/- 53 |
| Makkos5 | II + IIII + II + IIITotal | -4.9+ 4.2-6.7+5.2-7.1+6.3 |

**Table 1: overview clinically significant changes in UPDRS scores reported in literature.**

We searched PubMed using search terms: (Parkinson\*) AND (UPDRS) AND ((clinical\* relevant) OR (clinical\* significant)) AND (improve\* OR change OR difference). We selected papers which aimed to define minimal clinically relevant UPDRS changes and included predictive papers which used absolute thresholds for UPDRS changes in defining outcome.

Table 1 shows the reported UPDRS differences suggested to be clinically relevant.

For the UPDRS III score, the relevant differences range between 2.5 and 5 for the smallest relevant change. In one predictive analysis, a cut off change of 38% was used for UPDRS III off-medication. This translates to an absolute change of >10 in many patients. We averaged these findings and set the cut off for clinically relevant improvement after STN DBS at 5 points.

Since the UPDRS II and IV scores consist of fewer points, it is logical that the cut off values are lower. Based on the mentioned cut offs in the literature, we set the cut off for the UPDRS II and IV scores at 3 points.

1. **Imputation methods**

We imputed missing values in the presurgical parameters using Random Forrest Regressors or Classifiers for each individual parameter. The models were trained on all participants that had the data available while missing values were imputed based on the available presurgical parameters for the remaining participants. For continuous parameters, we used Random Forrest Regressors while categorical parameters were imputed using Random Forrest Classifiers. Only presurgical parameters were imputed to ensure that our results were not based on imputed values.

1. **References**

1. Kostoglou K, Michmizos KP, Stathis P, Sakas D, Nikita KS, Mitsis GD. Classification and Prediction of Clinical Improvement in Deep Brain Stimulation From Intraoperative Microelectrode Recordings. IEEE transactions on bio-medical engineering 2017;64(5):1123-1130.

2. Horvath K, Aschermann Z, Acs P, et al. Minimal clinically important difference on the Motor Examination part of MDS-UPDRS. Parkinsonism & related disorders 2015;21(12):1421-1426.

3. Shulman LM, Gruber-Baldini AL, Anderson KE, Fishman PS, Reich SG, Weiner WJ. The clinically important difference on the unified Parkinson's disease rating scale. Archives of neurology 2010;67(1):64-70.

4. Schrag A, Sampaio C, Counsell N, Poewe W. Minimal clinically important change on the unified Parkinson's disease rating scale. Movement disorders : official journal of the Movement Disorder Society 2006;21(8):1200-1207.

5. Makkos A, Kovacs M, Aschermann Z, et al. Are the MDS-UPDRS-Based Composite Scores Clinically Applicable? Movement disorders : official journal of the Movement Disorder Society 2018;33(5):835-839.