Statistical Analysis and Mapping of the Unified Parkinson's Disease Rating Scale to Hoehn and Yahr Staging

Authors: Athanasios Tsanas[1,2,*], Max A. Little[2,3], Patrick E. McSharry[1,2,4], Blake K. Scanlon[5,6], Spyridon Papapetropoulos[7]

* *Asterisk denotes corresponding author.*

Affiliations: [1]Systems Analysis, Modelling and Prediction (SAMP), Department of Engineering Science, University of Oxford, Oxford, UK

[2]Oxford Centre for Industrial and Applied Mathematics (OCIAM), University of Oxford, Oxford, UK

[3]Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

[4]Smith School of Enterprise and the Environment, University of Oxford, Oxford, UK

[5]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA

[6]Sierra-Pacific Mental Illness Research, Education, and Clinical Center, VA Palo Alto Health Care System, Palo Alto, CA, USA

[7]Department of Neurology, Miller School of Medicine, University of Miami, USA

Contact Information:

Athanasios Tsanas, OCIAM, Mathematical Institute, 24-29 St. Giles, Oxford, UK, OX1 3LB
Tel. 0044 1865280623, Fax 0044 1865270515.

Emails: (A. Tsanas) tsanas@maths.ox.ac.uk , tsanasthanasis@gmail.com

(M. Little) littlem@physics.ox.ac.uk

(P. McSharry) patrick@mcsharry.net

(B. Scanlon) bscanlon@stanford.edu

(S. Papapetropoulos) SPapapetropoulos@med.miami.edu

Key Words: Unified Parkinson's Disease Rating Scale (UPDRS), Hoehn and Yahr (H&Y) stage, statistical analysis, statistical machine learning, Parkinson's Disease Data and Organizing Center (PD-DOC), genetic algorithms

Short title: Statistical mapping of UPDRS to H&Y

Disclosure/Conflict of Interest: We have no conflict of interest. A. Tsanas is funded, in part, by Intel Corporation and in part by the Engineering and Physical Sciences Research Council (EPSRC). M.A. Little acknowledges the financial support of the Wellcome Trust, grant number WT090651MF. P.E. McSharry acknowledges the financial support of the European Commission through the SafeWind project (ENK7-CT2008-213740). B. Scanlon is currently supported by the Office of Academic Affiliations, Advanced Fellowship Program in Mental Illness Research and Treatment, Department of Veteran Affairs, by the Medical Research Service of the Veterans Affairs Palo Alto Health Care System, and by the Department of Veterans Affairs Sierra-Pacific Mental Illness Research, Education, and Clinical Center (MIRECC). S. Papapetropoulos is an employee and holds stock at Allergan Inc (Irvine, USA).

The data source for this work was the Parkinson's Disease Data and Organizing Center (PD-DOC; U01NS050095, PI: Roger Kurlan), currently managed by the National Institute of Neurological Disorders and Stroke (NINDS). Data has been contributed to the PD-DOC by the NINDS Udall Centers of Excellence for Parkinson's Disease Research, the NINDS Neuroprotective Exploratory Trials in Parkinson's Disease Program and the Parkinson Study Group. We used data from two studies available in the PD-DOC repository: (1) the study

from three Udall Centers (PIs: (i) T.M. Dawson, (ii) M.K. Lee, (iii) V.L. Dawson), and (2) the PostCEPT study (PIs: B. Ravina and A.E. Lang).

A. Tsanas had full access to the data and organized the mathematical analysis. All authors jointly take responsibility for the decision to submit for publication.

Article type: Letter to the editor

Parkinson's disease (PD) symptom severity is typically quantified using clinical metrics, where a medical rater assesses the subject's condition and ability to perform a range of tasks. Two of the most commonly used PD metrics are the Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn and Yahr (H&Y) scale [1]. UPDRS quantifies PD symptom severity, and H&Y quantifies disease stage; there have been few attempts to correlate these metrics. Recently, some mathematical formulas have been proposed expressing H&Y as a function of UPDRS using intuitive rules based upon H&Y evaluation guidelines [2], [3]. However, these formulas have not been tested against PD databases to assess how accurately H&Y might be predicted in general. In this letter, we (a) systematically study the statistical relationship between the two metrics, (b) optimize the previously proposed formulas by refining their parameters, and (c) test these formulas using a large, publicly available, PD database. The present study's findings provide (a) an optimized formula exhibiting 9% improvement in H&Y estimation compared to the formulas in [2], [3], (b) insight into the statistical association between the two most commonly used clinical metrics in PD, (c) a technique for substituting missing H&Y data in prospective clinical studies, and (d) an approach to harmonizing data in retrospective clinical studies.

The database source of this study is the Parkinson's Disease Data and Organizing Center (PD-DOC) [4], which was developed to facilitate the planning, study design, and statistical analysis of PD-related data. There are 566 individual subjects (and 1623 samples because some subjects had more than one evaluation), collected by several independent clinical centers, organized by the University of Rochester. Ethics approval was obtained by the IRBs of each of the participating centres. We used data from 525 (340 male) subjects, giving 1486 (979 male) samples with no missing values (the dataset comprises between 1 and 4 samples

per subject). Ages were (mean ± standard deviation) 62.7±9.3 years, with 4.7±1.9 years since PD diagnosis, 32.7±15.6 total-UPDRS, and 2.0±0.5 H&Y. Subjects who have not been diagnosed with very high probability (>90%) with idiopathic PD were discarded since the focus of this study is PD patients; similarly, samples collected when PD subjects fluctuated during the actual clinical assessments as a result of treatment (i.e. PD subjects were in the 'wearing off' phase) were discarded to avoid mismatch between UPDRS and H&Y values. Samples with missing values in UPDRS entries used in the formulas of Scanlon et al. [2], [3] or missing H&Y values were discarded to avoid ambiguities in the UPDRS to H&Y mapping. PD symptoms may be different in males and females [5]; therefore, we partitioned the data according to gender: the male subset contained 979 samples and the female subset contained 507 samples.

To quantify the strength of statistical association of the UPDRS entries with H&Y, we used the Spearman rank correlation coefficient (which quantifies *monotonic* relationships), and the *normalized mutual information* (MI) [5], which can detect general statistical relationships. We also computed statistical significance of the null hypothesis that UPDRS entries have no correlation with H&Y. The results are presented in Table 1: in broad agreement with clinical expectations, postural stability, gait, arising from chair and bradykinesia are, statistically, the UPDRS entries most strongly correlated with H&Y. Interestingly, the results are similar for males and females, with a few subtle differences.

Next, Scanlon's formulas [2], [3] (we refer to the latter reference as the *modified* Scanlon formula), are used to convert UPDRS to H&Y. The proposed model *structure* (chosen UPDRS entries, and logical operators) in those formulas is sensible and clinically interpretable, but we hypothesized that the *parameters* (the values used as thresholds) might not be optimally tuned. Since exhaustive brute force search through the *parameter space* (the range of all possible combinations of all the parameters) would be practically infeasible (with

25 and 27 parameters, where each parameter is an integer 0…4, to fully explore the parameter space we would need $5^{25} \approx 3 \times 10^{17}$ evaluations for Scanlon's formula and $5^{27} \approx 7.5 \times 10^{18}$ evaluations for the modified Scanlon formula), we used a genetic algorithm to identify whether a different parameter set would provide more accurate results (where accuracy is defined by the *mean absolute error*, MAE). Genetic algorithms (inspired by, but not to be confused with genetics in biology) explore the parameter space by identifying promising parameter sets and combining them; parameter sets that lead to large errors are penalized and the genetic algorithm focuses on those good-performing parameter sets which are most likely to lead to lower prediction error in future iterations. We used 200 chromosomes (with four elite chromosomes) and 5000 iterations to explore the parameter space. The initial population was randomly generated.

To avoid the risk of *overfitting*, the genetic algorithm optimization was performed in a 10-fold cross validation setting: the dataset consisting of N samples (N=979 for the male subset, and N=507 for the female subset) was randomly permuted, and we used 90% of the data to determine the model parameters. The process was repeated 10 times, each time randomly permuting the initial dataset and using 90% of the samples. In total, we obtained 10 optimized combinations of parameter values (chromosomes) as a result of the 10 repetitions, and each model parameter was assigned to the most frequently-occurring, corresponding value of the parameter in the optimized chromosomes. The updated Scanlon formulas appear in Table 2: most of the parameters selected based on clinical experience would be optimal in a statistical sense. The presence of more than one sample from some PD subjects might have suggested some selection bias, but no individual is contributing disproportionally to the database to make a significant difference.

Scanlon's formulas estimate H&Y with MAE around 0.176 points; optimizing the formulas' parameters using the genetic algorithm leads to statistically significant (p<0.001)

improvement (MAE drops to 0.161, a ~9% improvement). Overall, the formulas mapping UPDRS to H&Y are intuitively attractive, clinically relevant, easily interpretable, and statistically accurate. We conclude this study by endorsing the formulas with the refined parameters (Table 2), to infer H&Y stage from UPDRS.

Acknowledgments: We want to thank A. Kramer for constructive feedback and comments. A. Tsanas gratefully acknowledges the financial support of Intel Corporation and the Engineering and Physical Sciences Research Council (EPSRC). M. Little acknowledges the Wellcome Trust and Alex 'Sandy' Pentland at MIT for financial support. P.E. McSharry acknowledges the financial support of the European Commission through the SafeWind project (ENK7-CT2008-213740). B. Scanlon is currently supported by the Office of Academic Affiliations, Advanced Fellowship Program in Mental Illness Research and Treatment, Department of Veteran Affairs, by the Medical Research Service of the Veterans Affairs Palo Alto Health Care System, and by the Department of Veterans Affairs Sierra-Pacific Mental Illness Research, Education, and Clinical Center (MIRECC). S. Papapetropoulos is an employee and holds stock at Allergan Inc (Irvine, USA). The data source for this work was the Parkinson's Disease Data and Organizing Center (PD-DOC; U01NS050095, PI: Roger Kurlan), currently managed by the National Institute of Neurological Disorders and Stroke (NINDS). Data has been contributed to the PD-DOC by the NINDS Udall Centers of Excellence for Parkinson's Disease Research, the NINDS Neuroprotective Exploratory Trials in Parkinson's Disease Program and the Parkinson Study Group. We used data from two studies available in the original PD-DOC repository: (1) the study from three Udall Centers (PIs: (i) T.M. Dawson, (ii) M.K. Lee, (iii) V.L. Dawson), and (2) the PostCEPT study (PIs: B. Ravina and A.E. Lang).

References

[1] Ramaker, C, Marinus, J, Stiggelbout AM, van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Movement Disorders 2002; 17:867-876

[2] Scanlon BK, Katzen HL, Levin BE, Singer C, Papapetropoulos S. A formula for the conversion of UPDRS-III scores to Hoehn and Yahr stage. Parkinsonism and Related Disorders 2008; 14:379–80.

[3] Scanlon BK, Katzen HL, Levin BE, Singer C, Papapetropoulos S. A revised formula for the conversion of UPDRS-III scores to Hoehn and Yahr stage. Parkinsonism & Related Disorders 2010, 16(2), 151-152.

[4] Kurlan R, Murphy D. Parkinson's disease data and organizing center. Movement Disorders 2007; 22:904–905.

[5] Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. Journal of the Royal Society Interface 2011; 8:842-855

Tables

Table 1: Statistical strength association between UPDRS sections and H&Y, using normalized mutual information and rank correlations (see text).

Male subset (N = 979 samples)			Female subset (N = 507 samples)		
UPDRS section	MI	Spearman correlation	UPDRS section	MI	Spearman correlation
(43) Postural stability	**0.283**	0.627	(43) Postural stability	**0.330**	0.664
(42) Gait	**0.119**	0.393	(42) Gait	**0.164**	0.425
(40) Arise from chair	**0.118**	0.398	(40) Arise from chair	**0.111**	0.365
(44) Bradykinesia	**0.106**	0.390	(44) Bradykinesia	**0.106**	0.396
(41) Posture	**0.096**	0.378	(15) Walking	**0.105**	0.324
(15) Walking	**0.089**	0.330	(41) Posture	**0.091**	0.372
(39) Left leg agility	**0.086**	0.363	(11) Hygiene	**0.088**	0.244
(18) Speech	**0.081**	0.366	(39) Left leg agility	**0.082**	0.346
(31) Rigid LLE	**0.078**	0.323	(37) Left hand P/S	**0.077**	0.349
(35) Left hand grip	**0.077**	0.362	(10) Dressing	**0.076**	0.293
(27) Rigid neck	**0.074**	0.330	(27) Rigid neck	**0.075**	0.305
(14) Freezing	**0.074**	0.314	(33) Finger tap LH	**0.074**	0.343

All UPDRS sections were, with statistical significance, rank correlated with the modified H&Y stages ($p<0.001$). For clarity, only the 12 UPDRS sections most strongly associated with H&Y are presented here. We report the normalized mutual information (MI, range 0-1). The ordering is determined using MI. P/S stands for 'pronation-supination', LH for 'left hand', and LLE for 'lower left extremity'.

Table 2: Scanlon's formulae with parameters optimized using a genetic algorithm.

Refined Scanlon's formula

if ((item18 = 0 AND item19 = 0) AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L = 0) OR (item20R + item21R + item22R + item23R + item24R + item25R + item26R = 0))), **HY = 1**;

if ((item18 > 0 OR item19 > 0) AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L = 0) OR (item20R + item21R + item22R + item23R + item24R + item25R + item26R = 0))), **HY = 1.5;**

if (item30 = 0 AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L > 0) AND (item20R + item21R + item22R + item23R + item24R + item25R + item26R > 0))), **HY = 2**;

if ((item30 = 1) AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L > 0) AND (item20R + item21R + item22R + item23R + item24R + item25R + item26R > 0))), **HY = 2.5**;

if ((item30 > 1 AND item30 < 4) AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L > 0) AND (item20R + item21R + item22R + item23R + item24R + item25R + item26R > 0))), **HY = 3**;

if (((item29 > $\underline{\mathbf{X_1}}$ AND item29 < 4) AND (item27 < 4) AND (item31 > $\underline{\mathbf{X_2}}$ AND item31 <= 4))), **HY = 4**;

if (item29 = 4) OR (item30 = 4), **HY = 5**;

Refined, modified Scanlon's formula

if ((item18 = 0 AND item19 = 0 AND item22_neck = 0) AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L = 0) OR (item20R + item21R + item22R + item23R + item24R + item25R + item26R = 0))), **HY = 1;**

if ((item18 > 0 OR item19 > 0 OR item22_neck >0) AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L = 0) OR (item20R + item21R + item22R + item23R + item24R + item25R + item26R = 0))), **HY = 1.5;**

if (item30 = 0 AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L > 0) AND (item20R + item21R + item22R + item23R + item24R + item25R + item26R > 0))), **HY = 2;**

if ((item30 = 1) AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L > 0) AND (item20R + item21R + item22R + item23R + item24R + item25R + item26R > 0))), **HY = 2.5;**

if ((item30 > 1 AND item30 < 4) AND ((item20L + item21L + item22L + item23L + item24L + item25L + item26L > 0) AND (item20R + item21R + item22R + item23R + item24R + item25R + item26R > 0))), **HY = 3;**

if (((item29 > $\underline{\mathbf{X_3}}$ AND item29 < 4) AND (item27 < 4) AND (item31 > $\underline{\mathbf{X_4}}$ AND item31 <= 4))), **HY = 4;**

if (((item29 = 4) OR (item30 = 4))), **HY = 5;**

$\mathbf{X_1}$ = 1(S), 2(M), 2(F), $\mathbf{X_2}$ = 2(S), 2(M), 1(F), $\mathbf{X_3}$ = 1(S), 1(M), 2(F), $\mathbf{X_4}$ = 2(S), 3(M), 1(F)

The parameters in Scanlon's formulae which have been refined using the genetic algorithm appear in the form $\mathbf{X_k}$, as explained in the last row of this Table. 'S' stands for Scanlon's original formula, and the genetic algorithm values are denoted 'M' for the male subset, and F for the female subset.