Prediction of Change in Overall Performance for Patients with Huntington's Disease using Multilevel Functional Principal Component Analysis (MFPCA)

Zhi Pan[1], Jeff Goldsmith[1], Yuanjia Wang[1]

[1]Zhi Pan, Center for Public Health Initiatives, University of Pennsylvania, 3620 Hamilton Walk, Philadelphia, PA 19104; Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY 10032

[1]Jeff Goldsmith, Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY 10032

[1]Yuanjia Wang, Department of Biostatistics, Mailman School of Public Health, Columbia University, 722 West 168th Street, New York, NY 10032

Abstract

Huntington's disease (HD) causes progressive cognitive & motor impairment along with behavioral & psychiatric disorders. Our purposes are to determine association among cognitive tests and to predict overall performance of patients using key signals extracted from cognitive tests. To address the objective, MFPCA is applied to assess association of five major cognitive tests: symbol digit modalities test (SDMT), three STROOP tests (color, word & interference) and frontal system behavior scale. Then PC scores of MFPCA are used to predict change in overall performance monitored through scoring total function capacity (TFC). Education year & baseline age are added to adjust prediction model. Our results show MPFCA integrates information from all of subjects and cognitive tests to capture two main modes of variation at subject & test levels. The 1st two PC scores at both levels are used as predictors for they are able to shrink high dimensional data while retaining most original cognitive information. Furthermore, MPFCA-based prediction model is a better methodology than benchmark analysis, and our finding shows both adjusted PC scores related to SDMT & STROOP color tests and adjusted subject-specific PC scores have significant impact on TFC change.

Key Words: functional data analysis, MFPCA, Huntington's disease, cognitive evaluation, overall performance prediction

1. Introduction

Huntington's disease causes progressive cognitive and motor impairment along with behavioral and psychiatric disorders, which finally leads to dementia [1]. To develop into clinically diagnosed HD, gene mutation carriers may progress over many years; patients are monitored through a standard collection of clinical assessments called the unified HD rating scale (UHDRS), which test particular cognitive, motor, and behavior functions [1&2].

Among various UHDRS tests, cognitive evaluation is important to predict overall performance of HD patients. However, it is challenging to explain variation among various cognitive-related tests for different patients. Therefore, two major purposes of our study are to determine and quantify association among different cognitive tests as well as to predict change in overall performance using key signals extracted from the cognitive tests.

To address the first objective, a multilevel functional principal component analysis (MFPCA) is applied to assess the association of five major cognitive related UHDRS tests: the symbol digit modalities test (SDMT), three STROOP tests (color, word and interference), and frontal systems behavior scale (FrSBe) at both subject-level and subject/test level.

The second objective is addressed by using subject-level and subject/test level information derived from the above MFPCA to predict change in overall performance of HD patients, which is monitored through scoring total function capacity (TFC) test of UHDRS, using a multiple linear regression (MLR) model. In addition to MFPCA-based predictors, year of education and age at baseline are added to adjust our MLR prediction models. A traditional benchmark analysis is used as reference to the novel MFPCA-based model. The details of data structure and MFPCA methodology are introduced in the following section.

2. Methods

2.1 Data Structure

R software was used to complete data processing and modeling through the entire study. Our raw data is collected by PREDICT-HD [3]. Among 941 mutation carriers at risk, 126 subjects clinically diagnosed as HD during the entire study from 2002 to 2011. Patients who completed at least 5 annual visits are included in our final analysis, and the observations of their first 5 visits were remained to create a balanced data set. The final cognitive data contains 57 HD patients with M*N*J total observations, where M = 57 subjects, N = 5 visits per subject, J = 5 standardized tests per visit. The data structure has three features. First, the basic observation unit is a function in that each test is observed for each subject over time, and thus M*J subject/test-specific curves are observed over a grid of time points. Second, the sampling design is sparse and irregular. Specifically, the number of time points is five for each subject, and the collection of time points is a subset of the 11 possible values of our time variable, year to diagnosis, $t \in \{-6, 4\}$. Minus and positive signs indicate before and after HD diagnosis. Finally, data is hierarchical with multiple levels (level 1 for subject, level 2 for test).

2.2 MFPCA

The MFPCA model is chosen to address our multilevel sparse functional data [4&5]. Let Xij(t) denote the observed function measured over a time variable *t* for the j^{th} test within the i^{th} subject. Briefly, MFPCA combines functional PCA with functional ANOVA at multiple levels. The unique feature is to introduce level1 Zi(t) and level2 Wij(t) functions as nested ***random effects*** into functional ANOVA (1.1). Then the core step is to conduct functional PCA to decompose Zi(t) and Wij(t) using the KL expansion to get a two-level MFPCA model (1.2):

$$(1.1)\quad X_{ij}(t) = \mu(t) + \eta_j(t) + \boldsymbol{Z_i(t)} + \boldsymbol{W_{ij}(t)} + \epsilon_{ij}(t)$$

$$(1.2)\ X_{ij}(t) = \mu(t) + \eta_j(t) + \sum_k \boldsymbol{\xi_{ik}}\, \phi_k{}^{(1)}(t) + \sum_l \boldsymbol{\xi_{ijl}}\, \phi_l{}^{(2)}(t) + \epsilon_{ij}(t)$$

Where i = 1,…,M; j = 1,…, J; $t \in \{t_{ijs} = t_{is}: s = 1, \dots, N\}$ is from the set of 11 integer grid points {-6, 4} over which functions are estimated. Fixed functions μ(t) and $\eta_j(t)$ denote the overall mean and its test-specific shifts, respectively. $\varepsilon_{ij}(t)$ is a random measurement error with (0, σ^2). Fixed eigenfunctions $\phi_k{}^{(1)}(t)$, $\phi_l{}^{(2)}(t)$ are an

orthonormal basis of $L^2[0,1]$ for level 1 and level 2 respectively, and their eigenvalues ${\lambda_k}^{(1)}$ & ${\lambda_l}^{(2)}$ are variances of respective PC scores $\boldsymbol{\xi_{ik}}$ & $\boldsymbol{\xi_{ijl}}$.

The sparse MFPCA algorithm is used to estimate μ(t), $\eta_j(t)$, ${\phi_k}^{(1)}$& ${\phi_l}^{(2)}$, ${\lambda_k}^{(1)}$& ${\lambda_l}^{(2)}$ and σ^2 [4&5]. Briefly, discrete observations are converted into smooth functions of time using penalized spline smoothing, and eigenvalues and eigenfunctions of PCs are estimated using eigenanalysis. The MFPCA model can be rewritten as a linear mixed model conditional on these quantities:

$$(1.3)\quad X_{ij}(t_{ijs}) = \mu(t_{is}) + \eta_j(t_{is}) + \sum_{k=1}^{N1} \boldsymbol{\xi_{ik}}\, {\phi_k}^{(1)}(t_{is}) + \sum_{l=1}^{N2} \boldsymbol{\xi_{ijl}}\, {\phi_l}^{(2)}(t_{is}) + \epsilon_{ij}(t_{is})$$

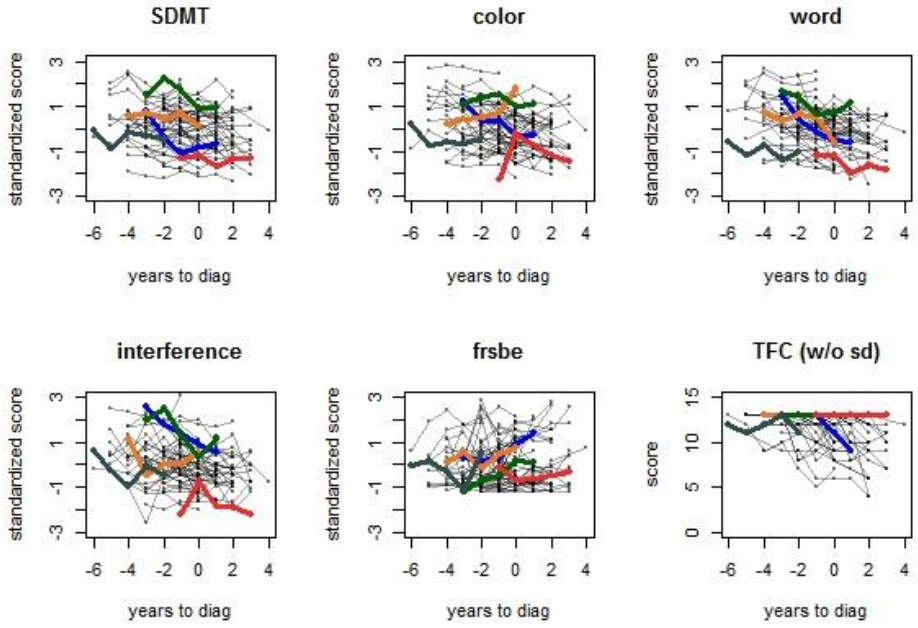
$$\text{With } \boldsymbol{\xi_{ik}} \sim \left(0, {\lambda_k}^{(1)}\right), \boldsymbol{\xi_{ijl}} \sim \left(0, {\lambda_l}^{(2)}\right) \text{ and } \epsilon_{ij}(t_{is}) \sim (0, \sigma^2)$$

The PC scores $\boldsymbol{\xi_{ik}}$ & $\boldsymbol{\xi_{ijl}}$ are estimated by the best linear unbiased predictions (BLUPs), and thus the linear functions *Zi(t)* and *Wij(t)* can be estimated using BLUPs as well.

3. Results & Discussion

A first look at the data reveals large within and between-subject heterogeneity, for measures differ in subjects and tests quite dramatically over time as exhibited in spaghetti plots **Figure1**. The sparse multilevel functional feature of these cognitive measures can also be observed in the plot. As described in method, a sparse MFPCA was applied in order to reveal the dominant mode of variation for this high dimensional functional data.

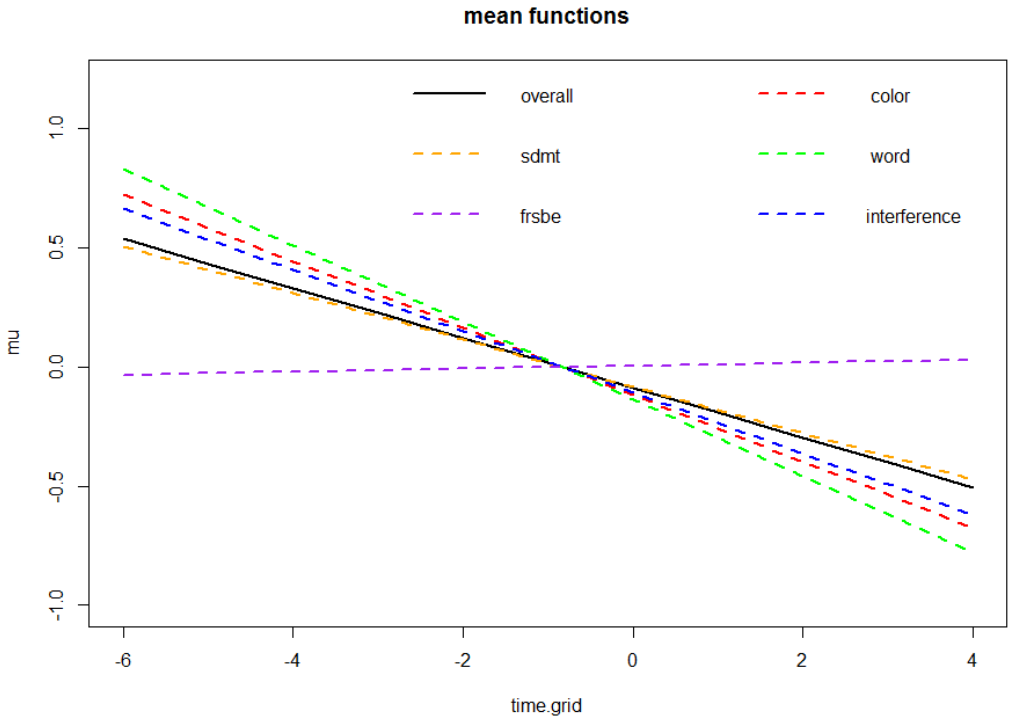
Figure1. Spaghetti plot of various UHDRS tests of 57 HD patients. Scores of five cognition-related tests are standardized due to different scales. Outcome of interest, TFC, remains its original scale as for prediction. Colorful lines correspond to the first five patients in the standardized HD data.



3.1 MFPCA on cognitive evaluation

After modeling, we first examine the fixed part of our MFPCA. From **Figure2,** the estimated mean functions display a linear trend, suggesting that cognition deteriorates over time. Test-specific mean functions $\hat{\mu}(t) + \hat{\eta}_j(t)$ show that FrSBe shifts quite differently from the overall mean, consistent with the fact that higher FrSBe score represents worse behavior, while lower scores of other tests indicate decline in cognition.

Figure2. Plot the estimated overall mean function $\hat{\mu}(t)$ (black line) and test-specific mean functions $\hat{\mu}(t) + \hat{\eta}_j(t)$, $j = 1, \ldots, 5$ (colored lines)



Then the variation explained by random level 1 and level 2 basis functions is exhibited in **Table1**. First, total functional variability is attributed to subject level 1 and subject/test level 2 around 52% and 48% respectively, and almost 100% of variation can be explained by their first four PCs. More specifically, most of the information in the subject level is contained in two uncorrelated dimensions, for its 1st two PCs characterize about 93% of variation at this level, splitting into 55% and 38 % in the direction of the respective 1st and 2nd eigenfunctions. Similarly, information contained in subject/test level 2 varies in two uncorrelated directions, for its 1st two PCs capture in total 91% of variation at this level with about 62% by the 1st PC and 29% by the 2nd PC.

Together, these results indicate the presence of clear differences across subjects as well as across tests within subjects, and illustrate that a few patterns suffice to describe most of these differences. Furthermore, numbers of PCs kept for level 1 & 2 are $N_1 = N_2 = 2$, as $N_1 = \min\left\{k: \rho_k{}^{(1)} \geq 0.9, \lambda_k{}^{(1)} < \frac{1}{10} = 0.1\right\}$ & $N_2 = \min\{k: \rho_k{}^{(2)} \geq 0.9, \lambda_k{}^{(2)} < 0.1\}$.

Therefore, PC score vectors, $\xi_i = (\xi_{i1}, \xi_{i2})$ at subject level 1 and $\xi_{ij} = (\xi_{ij1}, \xi_{ij2})$ at subject/test level 2, are key signals extracted from cognitive variables to predict TFC change in HD patients.

Table1. Variance explained by PCs of our sparse MFPCA model.

Model	**Level 1 (subject)**				**Level 2 (test)**			
PCs	1	2	3	4	1	2	3	4
Eigenvalues(λ)	4.90	3.41	0.64	0.02	5.05	2.39	0.59	0.09
per var %	54.61	38.02	7.09	0.21	61.97	29.29	7.29	1.08
cum per var % (ρ_k)	54.61	92.63	99.72	99.93	61.97	91.26	98.55	99.63
rho(ρ) %	52.41				47.59			

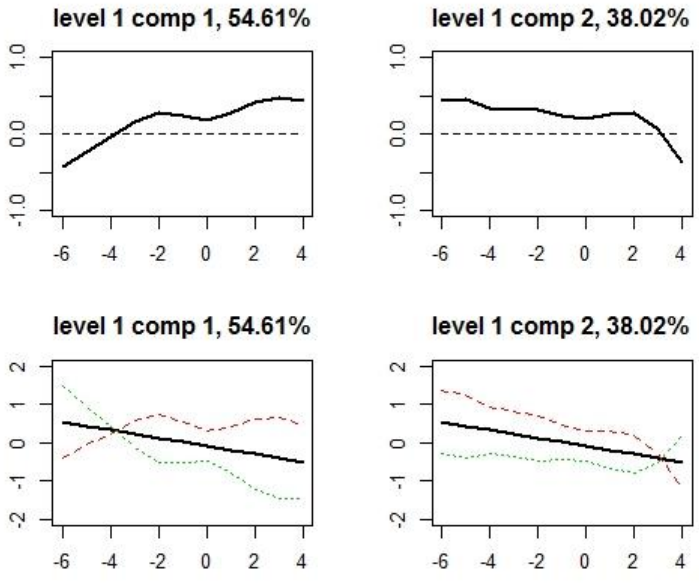
To better understand variability in major dimensions, we further visualize shape and magnitude of the first two eigenfunctions (top panel) and their effects on overall mean (bottom panel) at both levels in Figure3.

In the top panel of **Figure 3A**, subject-specific $\phi_1{}^{(1)}(t)$ gradually increases and turns positive from about 4 years before diagnosis. As a result, subjects with positive PC_1 scores will tend to have *smaller* cognitive test scores during 6 to 4 years before HD diagnosis and *greater* scores from then than the population average. In contrast, $\hat{\phi}_2{}^{(1)}(t)$ *is* positive until 3 years after diagnosis, and subjects with positive PC_2 scores will tend to have greater test scores than population average during that time. Effects of loading respective PCs is shown in bottom panel by adding or subtracting estimated $\sqrt{\lambda} \times \phi(t)$ from the overall mean. Although overall trend of cognitive decline over time is obscured by loading PC1, it is remained by loading PC2.

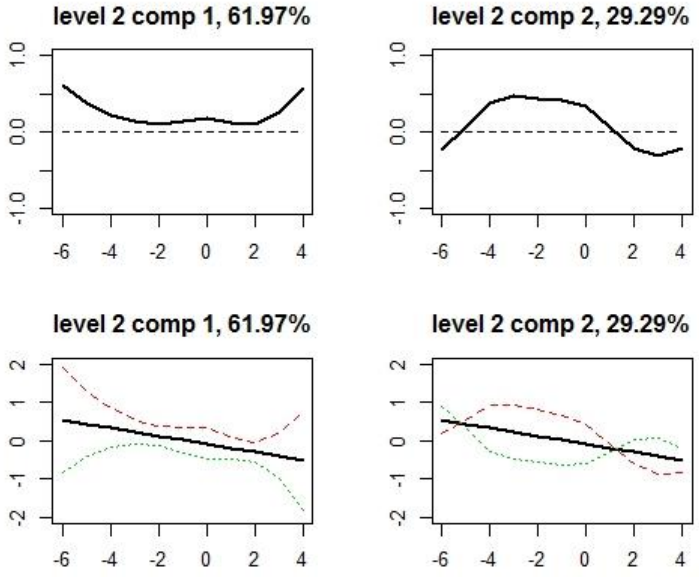
For subject/test specific level 2, **Figure 3B** shows that its 1st eigenfunction is positive throughout the entire time period, suggesting that subject tests with a positive PC_1 score correspond to a larger relative test score than subject average or interpreted as more prone to a shift in the related cognitive test. A further examination reveals that the weight placed in the middle is quite stable and is about five times less than that on the edges of the period. This result indicates that variation in the middle period is much smaller and consistent, which is also confirmed by loading PC1 on the overall mean. For the 2nd PC, its eigenfunction is positive and quite stable from 4 years before diagnosis to the time of diagnosis, and a subject with positive 2nd PC score thus tend to have a shift in related cognitive test. Loading the 2nd PC shifts the amplitude of overall mean quite consistently during the middle period.

Figure3. Plot estimated eigenfunctions $\phi(t)$ vs years to diagnosis (t) for the first two PCs in top panel for each level. Display $\mu \; and \; bands \; \mu \pm \sqrt{\lambda} \times \phi(t)$ in bottom panel with red dashed line (+) and green dashed line (-).

A. Level 1 (subject)

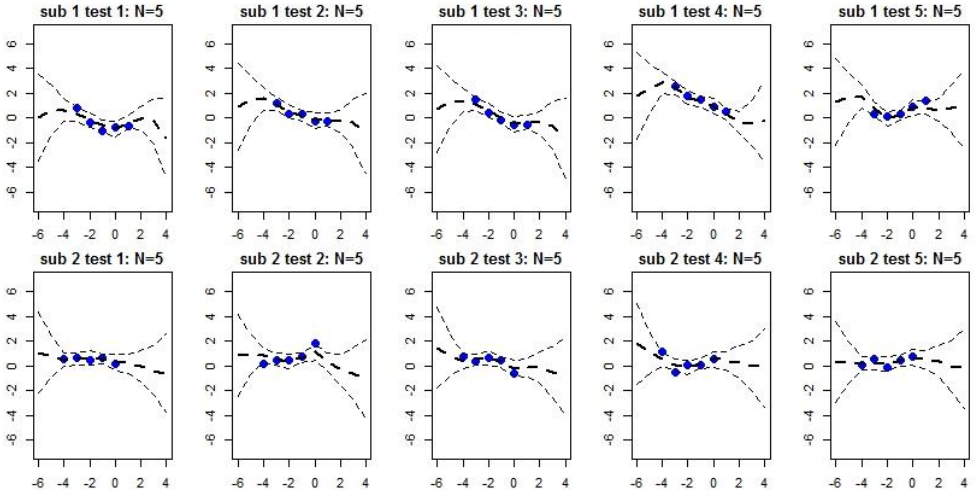


B. Level 2 (subject/test)



Next, we evaluate the performance of MFPCA model on predicting/fitting observed functions of our sparse data. Results of the first two patients are illustrated in **Figure 4**. Since observed sparse functions are fitted very well by predicted functions, our MFPCA model provides accurate estimates for sparse HD data. The 95% point-wise confidence bands (thin dashed lines) are much wider at both ends of the time interval due to fewer observations during these periods (data of EDA is not shown).

Figure4**.** Plot predicted functions and their 95% confidence bands for the 1st two HD patients. Each row corresponds to a patient, and column 1- 5 stands for tests (1, SDMT; 2, stroop color; 3, stroop word; 4, stroop interference; 5, FrSBe). For each subject, observations are pinpointed by blue dots {Xij(t)}. The predicted functions are labeled by thick dashed lines, i.e. $\hat{X}_{ij}(t) = \hat{\mu}(t) + \hat{\eta}_j(t) + \hat{Z}_i(t) + \widehat{W}_{ij}(t)$



3.2 MFPCA-based Prediction of HD performance

Finally, we construct a multivariate linear regression (MLR) model to predict change in TFC using the 1st two PCs of both levels from MFPCA (2.1). When evaluating effectiveness of a novel model, it's important to compare it against a benchmark. Hence, we also choose a standard model (2.2) for comparison. For both models, outcome, $Y_i = TFC_i(t_5) - TFC_i(t_1)$, is used to present progress of impairment in overall performance, and $Y_i \in \{-9,1\}$ for our data. Vi contains 2 covariates, age and education years at baseline, to adjust models. Major distinction is that PC scores from MFPCA model represent cognition progress in model 2.1, while baseline and change in original cognitive test are used in model 2.2

$$(2.1)\ E(Y_i) = \beta_0 + \sum_{k=1}^{N1} \beta_k \boldsymbol{\xi_{ik}} + \sum_{l=1}^{N2} \beta_{jl} \boldsymbol{\xi_{ijl}} + V_i{}^T \gamma$$

$$(2.2)\ E(Y_i) = \beta_0 + \sum_{j}^{5} \beta_j X_{ij}(t_1) + \sum_{j}^{5} \theta_j [X_{ij}(t_5) - X_{ij}(t_1)] + V_i{}^T \gamma$$

We use best-subset method for model section, and our result of final model (5V) in **Table2** show that adjusted PC1 score from subject/stroop color test is the most significant predictor (estimated β= 0.53, p< 0.001), indicating higher values of this PC score are associated with smaller TFC deterioration over time. In addition, both adjusted PC1 and PC2 scores from subject/SDMT test are significant but in two opposite ways, and subject-specific PCs and baseline education year are also useful covariates.

Table2. Coefficient estimates for the best MLR models using PC scores of MFPCA as predictors (5V) vs Benchmark analysis (5B) selected by subset method. Int = Intercept, C1 = the 1st PC score, C2 = the 2nd PC score, S = SDMT, C = color test, W = word test, I= interference test, F= FrSBe, E = education at baseline, and A = age at baseline; ('***' 0.001, '**' 0.01, '*' 0.05, and'.' 0.1)

MFPCA-based model (5V)

Level 1			**Level2 C1**					**Level2 C2**					**Adjust**	
Int	C1	C2	S	C	W	I	F	S	C	W	I	F	E	A
-5.57	0.55	-0.78	0.38	0.53				-0.70					0.25	
**	(.)	**	**	***				*					*	

Benchmark analysis (5B)

	Baseline					**Change**					**Adjust**	
Int	S	C	W	I	F	S	C	W	I	F	E	A
-4.16		1.04	-0.46	-0.36		1.29	0.78				0.23	
*		*				**	(.)				(.)	

The MFPCA-based model (5V) is much better than the benchmark analysis (5B) in terms of three aspects. First, the MPFCA-based model has higher adjusted R2 than the benchmark model (adjusted R^2 35.6% vs 29.0%). Second, the MPFCA-based model provides much more in-depth information. For example the benchmark model only shows that SDMT change is positive for prediction; however, the MPFCA-based model indicates that SDMT actually exerts two opposite impacts derived from two uncorrelated PCs at level 2, although the positive effect is more significant. Finally, the MPFCA-based model can quantify subject and subject/test specific effects separately while benchmark model cannot.

4. Conclusion

In summary, MPFCA can integrate information from all of subjects and cognitive tests to capture two major modes of variation at both levels of data. The first two PC scores at both levels are used as predictors of overall performance change in HD patients, for they are able to shrink high dimensional data while retaining most original cognitive information. Our finding reveals that MPFCA is a better methodology to identify key cognitive predictors for overall HD performance than standard benchmark analysis, and adjusted scores of subject specific PC and subject/test specific PC related to SDMT and STROOP color tests have significant impact on total functional capacity change.

References

1. Nance M. et al. 2011. A physician's guide to the management of Huntington's disease. *Huntington's disease society of America,* the 3rd ed
2. Klempíra J. et al. 2006. Unified Huntington's Disease Rating Scale: clinical practice and a critical approach. *Functional Neurology* 21(4): 217-221
3. Paulsen J. S .et al. 2008. Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *Journal of Neurology, Neurosurgery & Psychiatry* 79 (8): 874-880.
4. Di Chong-Zhi et al 2009. Multilevel functional principle component analysis. *The annuals of applied statistics* 3 (1): 458-488.
5. Di Chong-Zhi, Crainiceanu, C. M, Jank WS. 2014. Multilevel sparse functional principal component analysis. *Stat* 3(1): 126-143.