Mokken scale analysis of the UPDRS: Dimensionality of the Motor Section revisited

Jan Štochl [1*], Anne Boomsma [2], Marijtje van Duijn [2], Hana Brožová [3] and Evžen Růžička [3]

1. Charles University, Department of Kinanthropology, Czech Republic,
2. University of Groningen, Department of Sociology/Statistics & Measurement Theory, The Netherlands,
3. Charles University, Department of Neurology, Movement Disorder Centre, Czech Republic

Correspondence to: Jan Štochl, Ph.D.
Department of Kinanthropology, Charles University
José Martího 31, CZ-162 52 Prague, Czech Republic
TEL./FAX: +420 777 574 778, E-MAIL: stochl@ftvs.cuni.cz

Submitted: 2007-11-22 *Accepted: 2007-12-12* *Published online: 2008-02-22*

Key words: **Motor Section of the UPDRS; dimensionality; reliability; Mokken scale analysis; Parkinson's disease**

Neuroendocrinol Lett 2008; **29**(1):151–158 **PMID:** 18283252 NEL290108A25 ©2008 Neuroendocrinology Letters • **www.nel.edu**

Abstract

The dimensionality and reliability of the Motor Section of the Unified Parkinson Disease Rating Scale (UPDRS III) was studied with non-parametric Mokken scale analysis. UPDRS measures were obtained on 147 patients with PD (96 men, 51 women, mean age 61, range 35–80 yrs). Mokken scale analysis revealed a four-dimensional structure of the UPDRS III. Left-sided bradykinesia and rigidity appeared to co-occur with axial signs, gait disturbance, and speech/hypomimia, whereas right-sided bradykinesia and rigidity formed a second scale. Two further small scales were found consisting of right- and left-sided tremor. Results from the scale analysis reveal that all four subscales are strong. The reliability of the two tremor scales is low because they only contain three and four items, respectively.

INTRODUCTION

The neurological syndromes such as the combination of hypokinesia, rigidity, resting tremor and postural abnormalities in Parkinson's disease (PD) represent hypothetical concepts, which can be statistically modeled as latent traits. Although usually combined into one scale measure, the identification of the dimensionality of these syndromes is important because knowledge about the co-occurrence of symptoms may help to define disease phenotypes and provide clues for differential diagnosis. This paper investigates the dimensionality and the reliability of the Motor Section of the Unified Parkinson Disease Rating Scale (UPDRS) within the framework of Mokken scale analysis.

In 1987, the Unified Parkinson Disease Rating Scale was introduced as an overall assessment scale, which quantifies motor and behavioral symptoms and impairments related with PD. Soon it became a golden standard reference scale. A recent review revealed that sixty-nine percent of the studies published in the years 1994–1998 used the UPDRS for rating Parkinson's disease (Goetz, 2003). Thus, the Motor Section of the UPDRS is the most widely applied evaluation scale to measure the severity of impairment and to monitor the progression of PD.

Responses to the 27 items of the Motor Section fall in one of five response categories of increasing severity, scored from zero to four. The wording of the response categories is formulated differently for each item. However, the ordering of categories is invariant across the items. From a statistical point of view, this implies that the higher the respondent's value of the corresponding latent trait, the higher his or her item response score. The latent traits express the hidden aspects (sometimes called dimensions) of Parkinson's disease, which are assumed to be measured through main motor symptoms such as rigidity, bradykinesia, tremor, et cetera. The number of dimensions is typically inferred by using (statistical) scaling techniques, usually either by Exploratory Factor Analysis (EFA) or Confirmatory Factor Analysis (CFA). These procedures rely on rather strong assumptions concerning the (continuous) measurement level, the probability distribution of the item scores, and a large number of observations; maximum likelihood estimation requires multivariate normality and at least a couple of hundreds of observations (Blahuš, 1996, Boomsma and Hoogland, 2001).

Given the distributional properties of the indicators in the Motor Section of the UPDRS, EFA nor CFA using maximum likelihood estimation are the most appropriate scaling techniques, where the ordinal measurement level is the most obvious violation of the assumptions of the underlying statistical model.

Instead, we used methods conforming to Nonparametric Item Response Theory (NIRT) modeling. The NIRT models represent a family of statistical measurement models based on a minimal set of assumptions necessary to obtain useful measurements with the ultimate aim to order persons and/or items with respect to their latent trait value (Sijtsma and Molenaar, 2002). Unlike parametric item response models, nonparametric models do not parametrically define the function that describes the relation between the probability of a response in an item response category and the value on the latent trait. Since NIRT models are designed for ordinal measurement, they are particularly well suited to access the dimensionality structure of the UPDRS.

We apply two NIRT models developed by Robert J. Mokken in 1971 for building one-dimensional scales as well as for ordering items in such scales. Although Mokken models were initially developed for scales with dichotomous items (e.g., answers can be "yes" and "no", or "true" and "false"), they have been extended to polytomous ordered items by Molenaar (1982), which makes them appropriate for the scale analysis of the UPDRS; see also Sijtsma and Molenaar (2002) for an outline of the Mokken model and its estimation procedures.

The Mokken models are based on three assumptions defining the *Monotone Homogeneity Model* (MHM), a fourth one defining the stronger *Double Monotonicity Model* (DMM) (Sijtsma and Molenaar, 2002):

A. *Unidimensionality*: all items from the item set measure the same latent trait.
B. *Local independence*: given the respondent's individual latent trait value, his response to item *i* is not influenced by his responses to any other item in the same set.
C. *Monotonicity* of the so-called Item Step Response Functions (ISRFs): an ISRF describes the probability of responding positively to an item response category as a function of the latent trait value. Mathematically, this assumption requires each ISRF to be a monotone non-decreasing function.
D. *Double monotonicity* additionally assumes non-intersecting of ISRFs across the latent trait. For each item, there are as many ISRFs as the number of response categories minus one (i.e., four in case of the UPDRS motor scale, where the lowest response category serves as a reference category). Although it is assumed that the ISRFs do not intersect each other, they may touch locally and even coincide completely in the extreme case. This assumption implies unambiguous ordering of the items and of response categories within each item, and its validity enables an improved estimate of the reliability within each dimension.

The Mokken models require no parametric assumptions about the distributions of item responses, thus formulating very general assumptions about the functional relationship between latent trait and item response behavior. The models are applicable also in the case of relatively small sample sizes (Sijtsma and Molenaar, 2002). Like factor analysis, Mokken scale analysis can be both confirmative (testing the homogeneity of a hypothesized scale by assessing the correlation between the items), and explorative (searching for one or more unidimensional scales from a set of items).

In classical test theory, the reliability of a test is defined as the ratio of so-called true-score variance and the observed variance of the test score. The true-score variance, however, is unknown, because of undetermined measurement errors, whose variance can only be estimated. In the context of NIRT, two methods are available for estimating a lower bound of the reliability of the test score: one proposed by Mokken (1971), the other by Sijtsma and Molenaar (1987). The reliability coefficient estimated by these methods is denoted as the Rho coefficient. Its estimator has a smaller bias with respect to the population correlation between repeated measurements than Cronbach's widely used alpha coefficient. In addition, Cronbach´s alpha has a larger sampling error than the Rho coefficient estimators. All measures of reliability increase with the number of items in the scale. It is important to note, however, that the reliability coefficient of the scale estimated by these methods is valid only for the DMM.

In a sample of UPDRS Motor Section data obtained in a group of PD patients, with non-normally distributed scale scores, the unidimensionality of the complete scale was tested, followed by an exploratory Mokken scale analysis to assess the dimensionality of the Motor Section of the UPDRS. Where feasible, lower boundaries of reliability were estimated and compared with Cronbach's alpha coefficient.

METHODS

Subjects

The study was approved by the local Ethics Committee of the General Faculty Hospital in Prague. All subjects gave informed consent before entering the study. One hundred forty-seven consecutive patients (96 men, 51 women, mean age 61, range 35–80 years) with PD diagnosed according to current clinical criteria (Hughes et al., 1992) were included in the sample at the Movement Disorder Centre, Charles University, Prague. Mean PD duration was 8.4 years (SD 6.0, range 1–27), median Hoehn and Yahr stage was 2.5 (mean 2.3, SD 0.72, 19% patients were in stages 1 or 1.5, 52% in stages 2 or 2.5, 27% in stage 3, and 3% in stages 4 or 5). 63% of the patients suffered from late motor complications of PD (motor fluctuations or dyskinesias). Most patients (86.3%) were on levodopa, mean duration of treatment 6.5 years (SD 5.3, range 1–24). 59.2% of the patients were taking dopamine agonists. UPDRS was tested as a part of clinical visits by one of a group of certified neurologists specialized in movement disorders. 60 patients were examined in 'on' and 52 patients in 'off' motor state. For 35 patients their motor state during evaluation was not specified.

Methods

Before analyzing the latent structure of the Motor Section of the UPDRS, an analysis of the basic statistical properties of each item was made in order to get an overview of data characteristics. For this purpose the NCSS program (Hintze, 1996) and the PRELIS program (Jöreskog and Sörbom, 2002) were used.

For Mokken scale analysis of the UPDRS data, the MSPWin 5.0 program (Boer, 2001) was used. After testing the complete scale, the (explorative) search option of the program was employed, seeking the most coherent unidimensional subscales that can be created from the set of all items under assumptions A, B and C of the Mokken model. To evaluate the homogeneity of a Mokken scale, Loevinger´s scalability or homogeneity coefficient H was used. The building stone of the H-coefficient is the item-pair scalability coefficient H_{ij}, measuring the association between two items by a standardized covariance taking into account the order of the items by their means. These pairwise coefficients can be combined into homogeneity coefficients for each item H_i, and one for the complete scale, denoted as H. If $H = 1$ there is no disordering or "inversion" of the item responses (absence of so-called Guttmann errors). If $H = 0$ there is no correlation among the test items (Sijtsma and Molenaar, 2002). Generally, scales with $H < 0.3$ are not considered to measure a unidimensional latent trait. Scales with $0.3 < H < 0.4$ are considered to be weak, if $0.4 < H < 0.5$ the scale is of medium strength, and if $H > 0.5$ the scale is interpreted as a strong one (Sijtsma and Molenaar, 2002).

To assess whether an item has sufficient association with other items in the set to be included into the scale, the corresponding item scalability coefficient H_i is used. As a rule of thumb, all H_is in a scale should be larger than 0.3.

In an exploratory Mokken scale analysis, scales are formed in a forward item selection process. The scale starts with the two items with the strongest association and ends when no further item fulfilling a pre-specified criterion of H_i can be added to the scale. In the exploratory Mokken scale analysis of the UPDRS data, the recommended initial cutoff criterion for Loevinger´s scalability coefficient $H_i > 0.30$ was chosen. Subsequently, this cutoff criterion was gradually increased to 0.45, following a suggestion of Sijtsma and Molenaar (2002), based on the study by Hemker, Sijtsma and Molenaar (1995) investigating the dimensionality of a set of items. On the basis of these analyses, the final scales were obtained by also taking the meaning of the items also into account, as recommended by Sijtsma and Molenaar (2002).

To obtain some validation of our findings with respect to behavior across groups and the association between the various subscales, we inspected the means and standard deviations of patients in the 'off' and 'on' states, and computed correlations of all subscales over all patients.

RESULTS

Data description

The Motor Section sum scores were ranging from 2 to 59 (mean 21.63, SD 12.23, skewness 1.07). The expected non-normal distribution of most item responses was empirically verified. In general, the sample distributions of the item responses were skewed to the right (skewness ranging from 0.26 to 4.34, mean skewness 1.16); see Table 1 for an overview.

Mokken Scale Analysis of the Motor Section of the UPDRS

The confirmative Mokken scale analysis of all 27 items of the Motor Section resulted in a scale H-coefficient equal to 0.36, characterizing a weak scale. Most items have H_i-coefficients larger than 0.3, except for the items related to speech and to tremor (face, lips, chin; at rest, and postural). Note that the tremor items are the least frequent symptoms. Results are exposed in Table 2 (Analysis 1).

Table 1. Basic statistical features of the data (N = 147)

Item		Mean	Range	St. deviation	Skewness	Kurtosis
Speech		0.86	0-4	0.84	1.09	4.66
Facial expression		1.32	0-4	0.70	0.78	4.18
Tremor – face, lips, chin		0.07	0-2	0.28	4.34	22.91
Tremor at rest	RUE	0.60	0-4	0.91	1.42	4.19
	LUE	0.53	0-4	0.84	1.55	4.88
	RLE	0.24	0-3	0.62	2.82	10.92
	LLE	0.24	0-3	0.64	2.78	10.24
Postural tremor	Right	0.54	0-3	0.72	1.48	5.38
	Left	0.57	0-3	0.69	1.05	3.77
Rigidity	H/N	0.79	0-3	0.82	0.85	3.16
	RUE	1.12	0-3	0.82	0.37	2.65
	LUE	0.99	0-3	0.87	0.26	1.91
	RLE	0.69	0-3	0.73	0.86	3.48
	LLE	0.73	0-3	0.76	0.76	3.01
Finger taps	Right	1.20	0-3	0.86	0.44	2.65
	Left	1.32	0-4	1.00	0.34	2.22
Hand movements	Right	0.90	0-3	0.78	0.52	2.72
	Left	0.98	0-3	0.91	0.64	2.60
Rapid alternating movements	Right	0.90	0-4	0.92	0.82	3.05
	Left	1.02	0-4	0.93	0.77	3.28
Leg agility	Right	0.90	0-4	0.92	0.89	3.27
	Left	1.02	0-4	0.97	0.94	3.54
Arise from chair		0.34	0-4	0.70	2.47	9.89
Posture		0.91	0-3	0.72	0.46	3.02
Gait		0.90	0-4	0.80	1.07	5.08
Postural stability		0.88	0-4	0.85	0.84	3.51
Body bradykinesia		1.10	0-4	0.82	0.57	3.34

Note: RUE – Right Upper Extremity; LUE – Left Upper Extremity; RLE – Right Lower Extremity; LLE – Left Lower Extremity, H/N – Head, Neck

Table 2. Exploratory results with cutoff criterion $H_i > 0.30$ (N = 147)

		Analysis 1	Analysis 2		
Item		H_i coefficients	H_i of subscale 1	H_i of subscale 2	H_i of subscale 3
Speech		0.26	0.35		
Facial expression		0.33	0.43		
Tremor – face, lips, chin		0.19			0.34
Tremor at rest	RUE	0.13		0.61	
	LUE	0.18			0.62
	RLE	0.13		0.53	
	LLE	0.19			0.52
Postural tremor	Right	0.17		0.46	
	Left	0.24			0.55
Rigidity	H/N	0.42	0.48		
	RUE	0.37	0.41		
	LUE	0.40	0.44		
	RLE	0.42	0.49		
	LLE	0.44	0.51		
Finger taps	Right	0.42	0.47		
	Left	0.44	0.52		
Hand movements	Right	0.40	0.46		
	Left	0.43	0.50		
Rapid alternating. movements	Right	0.41	0.47		
	Left	0.43	0.50		
Leg agility	Right	0.40	0.47		
	Left	0.43	0.53		
Arise from chair		0.42	0.52		
Posture		0.41	0.50		
Gait		0.41	0.49		
Postural stability		0.38	0.48		
Body bradykinesia		0.51	0.59		
Scale H		0.36	0.48	0.54	0.54
Reliability		–	–	0.74	0.77
Cronbach´s alpha		0.92	0.94	0.70	0.73

Note: RUE – Right Upper Extremity; LUE – Left Upper Extremity; RLE – Right Lower Extremity; LLE – Left Lower Extremity, H/N – Head, Neck

In the first exploratory analysis with the cutoff criterion $H_i > 0.3$, three subscales were found, reported in Table 2 (Analysis 2). The first subscale consists of all items related to rigidity, bradykinesia of the extremities and axial/gait bradykinesia; it is of medium strength (scale H equals 0.48). The item *Speech* has the lowest H_i coefficient (0.35), other H_i coefficients in this subscale range from 0.41 to 0.59. The scale does not conform to the DMM. Therefore, Cronbach´s alpha (0.94) is reported as an approximate measure of the scale's reliability.

The second and the third subscale correspond with the tremor-right concept (*Tremor at rest – RUE*, *Tremor at rest – RLE* and *Postural tremor – right hand*) and the tremor-left concept (*Tremor at rest – Face, Lips, Chin*, *Tremor at rest – LUE*, *Tremor at rest – LLE* and *Postural tremor – left hand*), respectively. Both subscales are strong, having high values of scale *H*s (0.54

both). For these subscales of limited size (they consist of three and four items only), there were neither violations against monotonicity nor non-intersection of the ISRFs. Hence, a lower bound for the reliability can be estimated by the Rho coefficient. The estimates, however, are not high (0.74 and 0.78, respectively) also because of the small number of items included, but higher than Cronbach´s alpha for these subscales (equal to 0.70 and 0.73, respectively).

After increasing the initial cutoff criterion for Loevinger´s scalability coefficient H_i to 0.45, subscale number one from the previous analysis was split into three subscales (see Table 3). The level of homogeneity of these three new subscales highly increased (strong scales with $H > 0.55$). Subscale 1 now included the items related to axial/gait bradykinesia (i.e., *Arise from chair*, *Posture*, *Gait*, *Postural stability* and *Body bradykinesia*), the item *Rigidity-head, neck* and all left-sided items measuring rigidity and bradykinesia of the extremities (*Rigidity-LUE*, *Rigidity-LLE*, *Finger taps-left hand*, *Hand movements – left hand*, *Rapid alternating movements – left hand* and *Leg agility – left leg*). Item *Leg agility – right leg* was also included in this subscale; it has the lowest value of H_i coefficients (0.45). The other H_i coefficients of this subscale ranged between 0.48 and 0.63. The second subscale contained right-sided items measuring rigidity and bradykinesia of the extremities except item *Leg agility – right leg* (*Rigidity-RUE*, *Rigidity-RLE*, *Finger taps-right hand*, *Hand movements – right hand* and *Rapid alternating movements – right hand*) with H_is ranging from 0.63 to 0.71. The first two subscales violated the double-monotonicity assumption; hence, their reliability could not be estimated by the Rho coefficient. Based on Cronbach's alpha, the reliability is satisfactory for both subscales (0.92 and 0.88, respectively). The items *Speech* and *Facial expression* generated another dimension, which is, however, very limited in size: it consists of two items only. The item *Tremor at rest – Face, Lips, Chin* was excluded by the program because of an excessively low H_i value, i.e., it does not fit to any of the four subscales. Increasing the cutoff *H*-value even more was considered useless, because it would fragmentize the scales even further.

Combining the results of the previous two exploratory analyses, where the three-scale solution seemed to be still insufficient, yet the 5-dimensional scale too fragmented, a four-dimensional structure was put to a confirmative test, with reasonable results as reported in Table 4. The four resulting scales are all strong. From subscale 1 in Table 2, a second scale was derived (with a scale $H > 0.70$) containing the right sided measurements of *Rigidity*, *Finger Taps*, *Hand Movements* and *Rapid Alternating Movements* and *Leg Agility*. The remaining items in the first scale concern the corresponding measurements of left-sided rigidity and bradykinesia of the extremities, as well as axial symptoms, gait, body bradykinesia and speech. The two other subscales measuring left and right-sided resting and postural

Table 3. Exploratory results with cutoff criterion $H_i > 0.45$ (N = 147)

Item		H_i of subscale 1	H_i of subscale 2	H_i of subscale 3	H_i of subscale 4	H_i of subscale 5
Speech						0.71
Facial expression						0.71
Tremor – face, lips, chin						
Tremor at rest	RUE			0.61		
	LUE				0.66	
	RLE			0.53		
	LLE				0.58	
Postural tremor	Right			0.46		
	Left				0.57	
Rigidity	H/N	0.48				
	RUE		0.63			
	LUE	0.51				
	RLE		0.66			
	LLE	0.58				
Finger taps	Right		0.71			
	Left	0.61				
Hand movements	Right		0.68			
	Left	0.59				
Rapid alternating movements	Right		0.68			
	Left	0.58				
Leg agility	Right	0.45				
	Left	0.60				
Arise from chair		0.57				
Posture		0.56				
Gait		0.52				
Postural stability		0.52				
Body bradykinesia		0.63				
Scale H		0.55	0.67	0.54	0.60	0.71
Reliability (Rho)		–	–	0.74	0.78	0.75
Cronbach´s alpha		0.92	0.88	0.70	0.73	0.71

Note: RUE – Right Upper Extremity; LUE – Left Upper Extremity; RLE – Right Lower Extremity; LLE – Left Lower Extremity, H/N – Head, Neck.

Table 4. Confirmatory results of the four-dimensional structure (N = 147)

Item		H_i of subscale 1	H_i of subscale 2	H_i of subscale 3	H_i of subscale 4
Speech		0.37			
Facial expression		0.46			
Tremor – face, lips, chin					0.34
Tremor at rest	RUE			0.61	
	LUE				0.62
	RLE			0.53	
	LLE				0.52
Postural tremor	Right			0.46	
	Left				0.55
Rigidity	H/N	0.48			
	RUE		0.60		
	LUE	0.50			
	RLE		0.62		
	LLE	0.58			
Finger taps	Right		0.69		
	Left	0.59			
Hand movements	Right		0.65		
	Left	0.56			
Rapid alternating movements	Right		0.66		
	Left	0.56			
Leg agility	Right		0.54		
	Left	0.57			
Arise from chair		0.54			
Posture		0.56			
Gait		0.51			
Postural stability		0.52			
Body bradykinesia		0.62			
Scale H		0.53	0.63	0.54	0.54
Reliability		–	–	0.74	0.77
Cronbach´s alpha		0.92	0.88	0.70	0.73

Note: RUE – Right Upper Extremity; LUE – Left Upper Extremity; RLE – Right Lower Extremity; LLE – Left Lower Extremity, H/N – Head, Neck

tremor were left unaltered. The means on these scales (correcting for the number of items) are considerably lower than the means on the first two scales, indicating far less severe symptoms of tremor in this sample.

All subscales in this confirmatory analysis (see Table 4) are unidimensional (assumption A) and follow the assumption of monotonicity of the ISRFs (assumption C). Moreover, subscales three and four also satisfy the assumption of double-monotonicity (assumption D). There are no reasons to presume that the assumption of local independence is violated.

Table 5. Spearman correlations among subscales (N =147)

	Subscale 1	Subscale 2	Subscale 3
Subscale 2	.50 *		
Subscale 3	–.01	.36 *	
Subscale 4	.31 *	.09	.32 *

* Correlation significant at the 0.01 level

Two further analyses were done to investigate the properties of the four scales in the current sample. First, Spearman correlations between the subscale scores revealed the highest correlation between the first (axial/left-sided) and second scale (right-sided), moderate correlations between the first and the fourth scale (left-sided tremor), the second and third (both right-sided) and between the third and fourth (both tremor) scales (see Table 5). Virtually no correlation was found between the first and third, and between the second and the fourth scale. These correlations show that the scales are measuring latent traits that are correlated to some extent but by no means identical.

Second, scale score means and standard deviations were computed for two groups of ‘off’ and ‘on’ patients in the sample (see Table 6). Although Student *t*-tests do not reveal significant differences between the groups, the differences in scale scores are in the expected direction, where the means of all items are higher in the ‘off’ state than in the ‘on’ state.

DISCUSSION

In the present study, Mokken scale analyses of the Motor Section of the Unified Parkinson Disease Rating Scale were performed. The suitability of using this nonparametric IRT model for this kind of scale follows, first of all, from the ordinal measurement level of the items, and was further enhanced by the tenability of assumptions of the Mokken model (Sijtsma and Molenaar, 2002).

After studying the distributional properties of the items, it was found that the items distributions were positively skewed, which reflects a tendency of choosing lower response categories. It can be argued that this was influenced by the choice of PD patients having relatively low levels of impairment. However, the sample

Table 6. Sample statistics of subscale scores for all patients, and for patients in 'off' and 'on' motor states

		all (*N=147*)	'off' (*N=52*)	'on' (*N=60*)
Subscale 1	Mean (SD)	13.16 (8.32)	15.50 (10.41)	11.73 (6.75)
	Mean; # items 15	0.94	1.11	0.84
	Median (range)	11 (0-42)	13 (3-39)	11 (0-30)
Subscale 2	Mean (SD)	5.73 (3.98)	7.02 (4.62)	4.92 (3.52)
	Mean; # items 6	0.96	1.17	0.82
	Median (range)	5 (0-19)	6.5 (0-19)	5 (0-15)
Subscale 3	Mean (SD)	1.39 (1.80)	1.69 (1.79)	1.17 (1.67)
	Mean; # items 3	0.46	0.56	0.39
	Median (range)	1 (0-9)	1 (0-7)	1 (0-7)
Subscale 4	Mean (SD)	1.41 (1.87)	1.94 (2.26)	1.20 (1.63)
	Mean; # items 3	0.35	0.49	0.30
	Median (range)	1 (0-10)	1 (0-10)	1 (0-7)

under study was taken in an outpatient clinic representing a current PD patient population where most often UPDRS testing is performed routinely. Another reason for the relatively high skewness of data might the wording of response categories, which may be more appropriate for testing severe impairments at a late stage of PD. The tendency for taking lower response categories was confirmed in a different sample (G.T. Stebbins, 2004, personal communication).

Several studies assessed the construct validity and the dimensionality of the Motor Section of the UPDRS using exploratory factor analysis or principle component analysis. Such studies found between three and six factors that accounted for a proportion of the total scale variance ranging from 59% to 78% in samples with no more than 300 patients (Stebbins et al., 1999, Stebbins and Goetz, 1998, Martinez-Martin et al., 1994, Cubo et al., 2000, Martignoni et al., 2003). Only Kroonenberg et al. (2006) performed confirmative factor analysis to establish a common factor structure in three groups of patients in a larger sample. They found the same expected difference between patients in the 'on' and 'off' states. These authors also noted the problem of non-normality of which their data suffered and an insufficient sample size to apply the most proper estimation method. Moreover, the measurement of UPDRS is obviously of ordinal type instead of continuous, which may also pose problems when using factor analysis. By contrast, Mokken scale analysis is designed for analyzing ordinal data, and because of its relatively weak assumptions it may be applied in small samples.

In the present study, Mokken scale analysis of the final four-dimensional structure revealed that side-sensitivity is an important characteristic. This is different from previous studies using (exploratory or confirmatory) factor analysis, where rigidity, bradykinesia of the extremities, axial/gait bradykinesia, and resting and postural tremor were always reported as separate dimensions (Stebbins et al., 1999, Stebbins and Goetz, 1998, Cubo et al., 2000). In our analysis, the items measuring rigidity and bradykinesia of the extremities were separated in two scales (left and right). Side-sensitivity was shown in a clinical cohort, indicating that initial PD symptoms start more frequently on the right-sided extremities than those on the left (Poewe and Wenning, 1998). This might account for the more independent behavior of right-sided items in group comparisons. Using EFA methods, possible side-sensitivity of bradykinesia of the extremities was mentioned before (Stebbins et al., 1999, Stebbins and Goetz, 1998) as well as that of postural tremor (Cubo et al., 2000). Side-sensitivity of rigidity and rest tremor, however, has never been reported so far.

On the other hand, the present results indicate that axial symptoms coincide with left-sided signs. Indeed, previous studies have suggested that the severity of signs and symptoms on the left side of the body correlate more strongly with disease progression than those on the right side of the body (Martinez-Martin et al., 1995, Martinez-Martin et al., 1994). However, the clinical and pathological relevance of these side-sensitive findings remain to be demonstrated. Furthermore, the relative independence of tremor from rigidity and bradykinesia can be viewed as an indication of the lack of relationship between tremor and PD disability, which is consistent with other reports (Henderson et al., 1991, Reynolds and Montgomery, 1987) and corresponds to common clinical experience.

Many studies on the dimensionality of the UPDRS Motor Section assessed its overall reliability by Cronbach´s alpha coefficient (Stebbins and Goetz, 1998, Cubo et al., 2000, Martignoni et al., 2003), and found its value to be very high, i.e., ≥ 0.9. In our study, we computed Cronbach's alphas for all the (one-dimensional) subscales separately. Due to a smaller number of items, they were not as high as the reliability measures reported previously. For the two small scales with only three items each pertaining to left and right-sided tremor, reliability remained low, even when based on the less biased Rho coefficient.

In the first scale, *Speech* was the item with the lowest H_i coefficient value. Speech may be harder to evaluate based on the item categories wording, which is also reflected in studies showing an unacceptable low interrater reliability of this item (Richards et al., 1994, Camicioli et al., 2001). From a clinical point of view, often the impairment of speech does not correspond to the patient's general motor status and may be incomparably worse (or, less frequently, better) than the rest of the examinations.

Although the four scales of the Mokken scale analysis, fully taking into account the ordinal nature of the data, shed a new light on the dimensionality of the UPDRS, our study has some limitations due to the composition and size of the sample. The data were insufficient to investigate the differences between patients in 'on' and 'off'-states further. Similar to Kroonenberg et al. (2006) it would be necessary to assess whether the 'final' four scales are equally valid for different groups of patients. Likewise, the validity of the scales for more severe patients of PD should be established in view of our sample predominantly consisting of mild to moderate PD cases. Moreover, the basically exploratory nature of the Mokken scale analyses performed requires cross-validation studies.

Acknowledgments

The authors are grateful to doctors Jan Roth, Petr Mecir, Robert Jech, Tereza Serranova, and Olga Ulmanova who performed UPDRS testing of patients. They are also obliged to professor K.L.Leenders (Groningen) and professor P. Blahuš (Prague) for their helpful comments to the manuscript. Jan Štochl was supported by Czech Science Foundation grant nr. 406/07/P513; and Foundation of Josef, Marie a Zdeňky Hlávkových. The study was partly supported by Czech Ministry of Education, Research Program MSM0021620849 and MSM0021620864; and Czech Ministry of Health, IGA MZ NR/9220-3.

REFERENCES

1 Blahuš P (1996) An example of COSAN-Model to investigate a non-linear hypothesis on factor loading. *Acta Universitatis Carolinae.* **32**: 57–69.
2 Boer P (2001) MSPWin. 5.0 ed. Groningen: ProGAMMA.
3 Boomsma A & Hoogland JJ (2001) The robustness of LISREL modeling revisited. IN Cudeck, R, Toit, S & Sörbom, D (Cudeck, R, Toit, S & Sörbom, D(Cudeck, R, Toit, S & Sörbom, Ds) *Structural equation modeling: Present and future. A festschrift in honour of Karl Jöreskog.* Chicago, IL: Scientific Software International.
4 Camicioli R, Grossmann SJ, Hudnell K & Anger KW (2001) Discriminating mild parkinsonism: Methods for epidemiological research. *Mov Disord.* **16**: 33–40.
5 Cubo E, Stebbins GT, Golbe LI, Nieves A, Leurgans S, Goetz CG, *et al.* (2000) Application of the Unified Parkinson's Disease Rating Scale in progressive supranuclear palsy: Factor analysis of the motor scale. *Mov Disord.* **15**: 276–279.
6 Goetz CG (2003) The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Mov Disord.* **18**: 738–750.
7 Hemker BT, Sijtsma K & Molenaar IW (1995) Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken's IRT model. *Appl Psych Meas.* **19**: 337–352.
8 Henderson L, Kennard C & Crawford T (1991) Scales for rating motor impairment in Parkinson's disease: studies of reliability and convergent validity. *J Neurol Neurosur Ps.* **54**: 18–24.
9 Hintze J (1996) NCSS. 6.0.21 ed. Kaysville, Utah.
10 Hughes AJ, Daniel SE, Kilford L & Lees AJ (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosur Ps.* **55**: 181–184.
11 Jöreskog KG & Sörbom D (2002) PRELIS. 2.54 ed. Lincolnwood, IL: Scientific Software International.
12 Kroonenberg PM, Oort FJ, Stebbins GT, Leurgans SE, Cubo E & Goetz CG (2006) Motor function in Parkinson's disease and supranuclear palsy: simultaneous factor analysis of a clinical scale in several populations. *BMC Med Res Method.* **26**.
13 Martignoni E, Franchignoni F, Pasetti C, Ferriero G & Picco D (2003) Psychometric properties of the Unified Parkinson's Disease Rating Scale and of the Short Parkinson's Evaluation Scale. *Neurol Sci.* **24**: 190–191.
14 Martinez-Martin P, Gil-Nagel A, Gracia LM, Gomez JB, Martinez-Sarries FJ, Bermejo F, *et al.* (1995) Intermediate Scale for Assessment of Parkinson's Disease. Characteristics and Structure *Parkinsonism Relat D.* **1**: 97–102.
15 Martinez-Martin P, Gil-Nagel A, Gracia LM, Gomez JB, Martinez-Sarries J & Bermejo F (1994) Unified Parkinson's Disease Rating Scale characteristics and structure. The cooperative multicentric group. *Mov Disord.* **9**: 76–83.
16 Mokken RJ (1971) *A theory and procedure of scale analysis,* The Hague: Mouton.
17 Molenaar IW (1982) Mokken scaling revisited. *Kwantitatieve Methoden.* **3**: 145–164.
18 Poewe WH & Wenning GK (1998) The natural history of Parkinson's disease. *Ann Neurol.* **44**: S1–9.
19 Reynolds N & Montgomery G (1987) Factor analysis of Parkinson's impairment. An evaluation of the final common pathway. *Arch Neur.* **44**: 1013–1016.
20 Richards M, Marder K, Cote L & Mayeux R (1994) Interrater reliability of the Unified Parkinson's Disease Rating Scale motor examination. *Mov Disord.* **9**: 89–91.
21 Sijtsma K & Molenaar IW (1987) Reliability of test scores in nonparametric item response theory. *Psychometrika.* **52**: 79–97.
22 Sijtsma K & Molenaar IW (2002) *Introduction to nonparametric item response theory,* London: Sage Publications.
23 Stebbins GT & Goetz CG (1998) Factor structure of the Unified Parkinson's Disease Rating Scale: Motor examination section. *Mov Disord.* **13**: 633–636.
24 Stebbins GT, Goetz CG, Lang AE & Cubo E (1999) Factor analysis of the motor section of the Unified Parkinson's Disease Rating Scale during the off-state. *Mov Disord.* **14**: 585–589.