Examination of the UPDRS Bradykinesia Subscale: Equivalence, Reliability and Validity

Philip O. Buck[a,*], Ronald E. Wilson[b], Lauren C. Seeberger[c], Jill B. Conner[a] and Jane Castelli-Haley[a]
[a]*Teva Neuroscience, Inc., Kansas City, MO, USA*
[b]*Brentwood Neurology, Brentwood, TN, USA*
[c]*Idaho Elks Rehabilitation Hospital, Boise, ID, USA*

Abstract. Administering items or subscales separately from the measure for which they were designed to be a part may have unintended consequences for research and practice in Parkinson's disease (PD). The current study tested the equivalence of the bradykinesia subscale when administered alone versus as a component of the full 14-item Unified Parkinson's Disease Rating Scale (UPDRS) motor examination, as well as examined the reliability and validity of the bradykinesia subscale. The study sample consisted of 112 patients with PD. Patients were randomly assigned to either the bradykinesia subscale alone group ($n = 56$), who were administered the bradykinesia subscale separately from the rest of the UPDRS motor examination, or the full scale group ($n = 56$), who were administered the UPDRS motor examination in its standard format. The two one-sided *t*-test (TOST) procedure was used to test for mean equivalency between the two administration groups. Additionally, reliability and validity analyses were performed. The bradykinesia subscale mean scores from the full scale group and the subscale alone group were not statistically equivalent. However, in both groups, the bradykinesia subscale had exceptional reliability and was strongly and similarly related to age, activities of daily living, disability, and other assessments of motor symptom severity. The bradykinesia subscale is a valid and reliable assessment when administered separately from the rest of the UPDRS motor examination; however, caution should be taken when comparing mean scores across studies or occasions when different administrations are used.

Keywords: Parkinson's disease, bradykinesia, UPDRS, equivalence, psychometrics

INTRODUCTION

Parkinson's disease (PD) is a chronic and progressive neurodegenerative disorder of the central nervous system associated with a decrease in dopamine production and characterized by impaired motor functioning. The prevalence of PD in the US is estimated to be about 0.3% for the entire population and approximately 1% for people over 60 years of age [1, 2]. The incidence of PD increases sharply after the age of 60 years [3–5].

Though PD patients routinely experience both motor and nonmotor symptoms [6], it is the motor symptoms that classically define PD. These include bradykinesia (slowness of movement), rigidity, tremor of resting muscles, postural instability or impaired balance, and gait disturbances [4, 5]. Recent research suggests that bradykinesia may be related to many important outcomes such as poorer performance on activities of daily living (ADL), as well as lower health-related quality of life, even after controlling for other motor symptoms (axial impairment), mood, and comorbidity [7]. In two

*Correspondence to: Philip O. Buck, PhD, MPH, Health Economics and Outcomes Research, Teva Neuroscience, Inc. 901 E. 104th Street, Suite 900, Kansas City, MO 64131, USA. Tel.: +1 215 293 7247; Fax: +1 816 508 5013; E-mail: Philip.Buck@tevapharm.com.

different comprehensive reviews, higher bradykinesia scores were prognostic of subsequent functional disability [8, 9]. In fact, greater bradykinesia scores have been associated with more severe symptoms of depression, specifically, difficulty sleeping and suicidal thoughts [10]. Given the significant association between bradykinesia and important health outcomes, frequent assessment of bradykinesia by researchers and clinicians is important in understanding health-related quality of life, well-being, and the progression of disability in patients with PD.

The Unified Parkinson's Disease Rating Scale (UPDRS) version 3.0 motor examination is a widely used assessments of motor impairment severity for PD patients [11]. The administration of the full UPDRS motor examination can be time intensive and requires extensive training to ensure inter-rater reliability, particularly when used in multi-center studies [12]. However, relative to the other items, the five items comprising the bradykinesia subscale of the UPDRS motor examination may be easier to score since they are based on patient movements that are already performed as part of a clinical assessment. Consequently, clinicians can assess a patient's bradykinesia severity during a routine examination by using just the bradykinesia subscale items. Moreover, depending on the goals and breadth of a study, researchers may elect to only include the bradykinesia items [13]. This practical assessment would allow clinicians and researchers to closely monitor and quickly track the progression of bradykinesia severity.

To date, research has not examined the reliability or validity of the bradykinesia subscale when it is administered separately from the rest of the UPRDS motor examination. The bradykinesia items are embedded within the UPDRS motor examination and are administered after items assessing speech, facial expression, tremor, rigidity, posture, and gait. Consequently, bradykinesia scores may be impacted when the subscale is administered alone. This impact may be explained by question-order effects which refer to differences in item responses due to changes in assessment format [14]. The consequences of administering items or subscales separately from the measure of which they were designed to be a part have generally been overlooked by clinical research in PD.

The current study examined the UPDRS bradykinesia subscale in two groups of patients: those who were administered the subscale separately from the rest of the UPDRS motor examination and those who were administered the subscale as part of the 14-item UPDRS motor examination. Three specific aims

were examined. The first aim was to determine if the means of the bradykinesia subscale were statistically equivalent between the two administration groups. The second aim was to determine if the bradykinesia subscale had similar reliabilities in both administration groups. The final aim was to examine the construct validity of the bradykinesia subscale by examining the correlations between the subscale and related constructs within each of the administration groups.

MATERIALS AND METHODS

Subjects

The study sample consisted of 112 PD patients with bradykinesia and at least one additional cardinal sign (i.e., resting tremor, rigidity) at the time of assessment, recruited from two centers in the US. The sample included male and female patients who were at least 30 years of age, had a diagnosis of idiopathic PD, were in the ON state (only for fluctuating patients), and had not received an experimental drug within the last 30 days. Patients that met eligibility criteria were stratified by center and previous Hoehn and Yahr stage [15], determined from patient charts and assessed between 3 and 180 days prior to study visit, into one of two predefined PD severity classifications: mild (stage 2.0 or less) or moderate/severe (greater than stage 2.0). Following stratification, patients were randomly assigned to one of two administration groups and were tested during a single clinic visit. No drugs were supplied or administered as part of this study. In the bradykinesia subscale alone group ($n = 56$), the bradykinesia subscale was completed separately from the remaining items of the UPDRS motor examination. In the full scale group ($n = 56$), the bradykinesia subscale was administered as part of the standard administration of the UPDRS motor examination. The study was approved by a central Institutional Review Board (Schulman Associates IRB, Inc.), and all patients provided informed consent prior to their participation.

Assessments

For each patient, information regarding age, gender, ethnicity, and PD duration (based on date of diagnosis) was collected. The bradykinesia subscale of the UPDRS motor examination consisted of the following five items: finger taps (left and right), hand movement (left and right), rapid alternate movements of hands (right and left), leg agility (right and left), and body bradykinesia and hypokinesia. Each sub-item

was scored from 0 to 4 (half points were allowed), where higher scores signified greater severity. The range of possible scores on the bradykinesia subscale was between 0 and 36. At each center, the ratings were performed by a single investigator who had completed certification training and who was required to review a UPDRS teaching tape prior to the start of the study [16].

Other outcomes included patients' current Hoehn and Yahr stage (half stages were allowed for stages 1 and 2), where higher stages represented greater progression of symptoms. The Schwab and England scale [17] was included to provide both a patient and clinician assessment of the patient's ability to perform activities of daily living (ADL). Patient- and clinician-reported ADL were scored in 5% increments on a 0 to 100% scale, where higher scores represented increased perceived competency with these activities. Finally, the first eight items of the Short Parkinson's Evaluation Scale/Scales for Outcomes in Parkinson's Disease (SPES/SCOPA) motor evaluation [18] was used to provide an alternative, shorter clinical assessment of motor impairment. The SPES/SCOPA was scored on a 4-point scale, where higher scores represented greater impairment; the total score could range from 0 to 42. The SPES/SCOPA bradykinesia item, rapid alternating movements of hands (right and left), was used to represent another specific assessment of bradykinesia; the subscale score could range from 0 to 6.

Statistical analysis

To insure randomization was successful, independent samples *t*-tests were employed to examine mean differences between the two groups; chi-square tests were used to examine categorical differences. The Mann-Whitney test was used to examine differences in current Hoehn and Yahr stage.

To test for equivalency of the bradykinesia subscale administered alone as opposed to administered within the full scale, the two one-sided *t*-test (TOST) procedure was used [19]. The TOST procedure is one of the most common statistical approaches for establishing equivalence. As outlined by Lauzon and Caffo [20], the TOST procedure first includes the a priori specification of the maximum expected lower and upper bound mean difference between the two groups. Next, using a 90% confidence interval [16] around the difference between the two means and an alpha level of 0.05, separate one-sided *t*-tests evaluated the null hypothesis that the difference between the means is either lower than the expected lower bound difference

or greater than the upper bound difference. That is, the null hypothesis states that the groups differ by more than the specified amount, while the alternative hypothesis states that the groups do not differ by more than the specific amount. Consequently, failing to reject the null hypothesis (e.g., at least one nonsignificant *t*-test) indicates that the means of the two groups are not statistically equivalent, while rejecting the null hypothesis (e.g., both *t*-tests are significant) indicates the means are equivalent. It is important to note that the more familiar independent sample *t*-test determines if two means are significantly different from one another, while the TOST procedure tests whether or not two means are statistically the same. Therefore, two means may not be significantly different from one another, while at the same time not be statistically equivalent. For the current study, the expected maximum difference between the two groups was specified to be half a standard deviation, which would reflect a medium effect size.

Chronbach alpha was used to assess the internal reliability of the bradykinesia subscale within each of the groups. Pearson and Spearman correlation coefficients between the bradykinesia subscale and the demographic/outcome variables were estimated separately within each group to assess construct validity. Fisher's *z*-test [21] was used to determine whether these correlations significantly differed between the two groups.

RESULTS

The sample was predominately Caucasian and consisted of approximately 60% males. Half of the patients (51%) had a previous Hoehn and Yahr stage of 2.0 or less (Table 1). The sample consisted of 79% of the patients being between 65 and 90 years old, with the remaining 21% ranging in age from 40 to 64. No statistically significant differences for the baseline variables were found between administration groups, providing evidence that the assumptions of random assignment were met.

Based on the independent sample *t*-test, the two administration groups did not significantly differ on the bradykinesia subscale ($p = 0.08$; Table 1). However, based on the TOST procedure, the difference between the two groups on the bradykinesia subscale was 2.13 (95% CI = 0.12, 4.13), with a maximum expected difference of -3.23 (lower bound) and 3.23 (upper bound). The lower bound *t*-test was significant ($t(110) = 4.43$, $p < 0.05$) with the lower bound CI

Table 1
Group differences for study variables

|	Subscale alone Group ($n = 56$) Mean (SD)	Full scale Group ($n = 56$) Mean (SD)	$t/\chi^2/z$	p-value
Demographics				
Age (Years)	72.45 (8.49)	70.14 (9.74)	1.33	0.19
Gender (Male)	57.10%	60.70%	0.15	0.70
Ethnicity (Caucasian)	96.40%	98.20%	0.34	0.56
Months since PD diagnosis	76.61 (58.61)	77.45 (56.80)	−0.08	0.94
UPDRS Bradykinesia Subscale	15.29 (6.27)	13.16 (6.53)	1.76	0.08
Construct validity outcomes				
Activities of daily living (Patient)	73.75 (17.95)	77.50 (21.76)	−1.00	0.32
Activities of daily living (Clinician)	70.00 (21.74)	74.38 (23.47)	−1.02	0.31
Current Hoehn and Yahr Stage	2.54 (0.75)	2.57 (0.69)	−0.43	0.66
SPES/SCOPA Motor Clinical Exam	9.80 (4.75)	9.25 (5.63)	0.56	0.58
SPES/SCOPA Bradykinesia Subscale	1.27 (0.60)	1.15 (0.75)	0.91	0.37
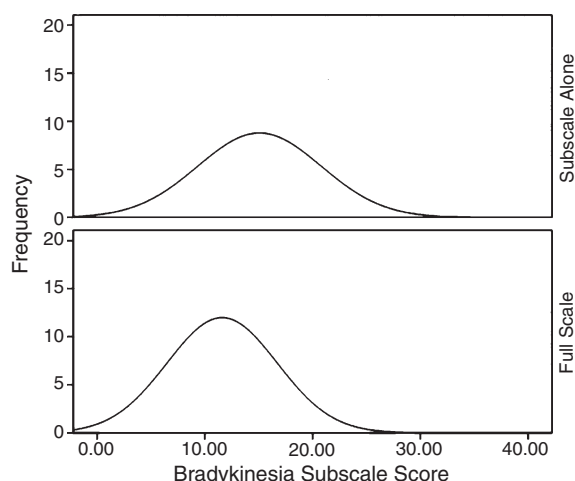
Fig. 1. Distribution curve for the bradykinesia subscale by administration group.

(0.12) falling well within the range of the maximum expected difference. However, the *t*-test for the upper bound was nonsignificant ($t(110) = -0.91$, $p > 0.05$), suggesting that the upper bound confidence interval (4.13) for the difference between the group means was beyond the range of the maximum expected difference (3.23). These results from the TOST procedure suggest that means for the bradykinesia subscale were not statistically equivalent between the subscale alone and full scale groups. However as can be seen in Figure 1, the distribution of the bradykinesia subscale was similar for the two administration groups. Kolmogrov-Smirnov tests indicated that the distribution in the subscale alone ($z = 0.61$, $p > 0.05$) and the full scale ($z = 1.19$, $p > 0.05$) groups were not statistically different from normal.

Evidence of strong reliability was found for the bradykinesia subscale in both the subscale alone ($\alpha = 0.93$) and full scale ($\alpha = 0.94$) groups. Table 2 contains the correlations between the UPDRS bradykinesia subscale and relevant demographics as well as each of the construct validity outcomes for both the subscale alone and full scale groups. Across the two groups, the bradykinesia subscale score was significantly related to each of these variables in the expected direction, with the exception of time since PD diagnosis, where it was only significant in the full scale group. Specifically, higher scores on the UPDRS bradykinesia subscale were related to greater chronological age, worse ADL competency, increased disability, and more severe motor symptoms. These correlations did not significantly differ between the two administration groups as evidenced by the nonsignificant *z*-tests for each pair of correlations (Table 2).

DISCUSSION

Clinical research on PD has largely ignored the issue of question-order effects, where scores on items or subscales do not behave the same when the method or order of scale administration is changed. Consequently, the current study was designed to determine whether the UPDRS bradykinesia subscale behaved similarly when administered separately to one group of PD patients relative to a group of patients who received the UPDRS motor examination in its original order.

Results from the TOST equivalency test revealed that the means of the two groups on the bradykinesia subscale were not statistically equivalent. That is, the bradykinesia subscale was not statistically the same when it was administered apart from the rest of the

Table 2
Construct validity correlations

Variables	UPDRS Bradykinesia Subscale			
	Subscale alone Group ($n = 56$)	Full scale Group ($n = 56$)	Fisher's z	p-value
Age	0.36*	0.51*	−0.96	0.34
Months since PD diagnosis	0.17	0.38*	−1.18	0.23
Activities of daily living (Patient)	−0.55*	−0.73*	1.60	0.11
Activities of daily living (Clinician)	−0.70*	−0.75*	0.54	0.59
Current Hoehn and Yahr Stage	0.55*	0.71*	−1.38	0.17
SPES/SCOPA Motor Clinical Exam	0.71*	0.75*	−0.44	0.66
SPES/SCOPA Bradykinesia Subscale	0.62*	0.79*	−1.78	0.08

Note. Pearson coefficients were used for all correlations except current Hoehn and Yahr stage, for which spearman coefficients were used. *$p < 0.05$.

UPDRS motor examination. However, it is important to note that while not statistically the same, the means of the two administration groups were also not statistically different from one another. These findings may seem at odds, but failing to reject the null hypothesis for traditional hypothesis testing merely indicates that the means are not different; it does not in any way suggest that they are the same. In the current study, the means of the two groups on the bradykinesia subscale were not the same, and they were not different from one another.

Though not statistically different, the mean bradykinesia score from the subscale alone group was higher than the full scale group. One possibility for this finding could be the presence of context effects, which is a type of question-order effect where the scores/answers on items change as a function of preceding items [22–24]. When the full UPDRS motor examination is administered, the scoring of the bradykinesia items may be influenced by the five preceding questions including rigidity and tremor. These early items might provide a context of symptom evaluation where the clinician is differentially primed for evaluating symptoms. In turn, this priming impacts the administration and scoring of the bradykinesia items. Alternatively, these early items could be used as points of reference when completing the bradykinesia items and the lower scores could be partially the product of comparative judgments. A similar finding was reported by Bowling and Windsor [25] who found that self-rated health status was worse when asked before rather than after a set of health-related questions. In the current study, bradykinesia was rated as more severe if the bradykinesia items were administered without questions assessing other symptoms.

Results indicated that the internal consistency of the bradykinesia subscale was extremely high regardless of administration approach. Moreover, there was strong support for the construct validity of the bradykinesia subscale when it was administered separately from the rest of the UPDRS motor examination. Specifically, it was significantly correlated with age as well as clinical measures assessing functional disability, motor symptoms, and other bradykinesia symptoms. More importantly, these relationships were not statistically different across groups. This means that when the bradykinesia subscale was administered alone, its relationship with age and clinically meaningful variables was the same as when it was administered in the context of the UPDRS motor examination. The only exception was time since PD diagnosis, which, while not significant in the subscale alone group, was also not significantly different between the two groups. Consequently, clinicians and researchers can feel comfortable administering the bradykineisa subscale separately from the UPDRS motor examination without any loss of predictive validity. This is consistent with other research indicating that significant context effects often do not impact the ability of a measure to correlate with background characteristics or indices of other behaviors [24].

The findings of the current study suggest that using the UPDRS bradykinesia subscale alone is acceptable. However, because the mean bradykinesia score from the subscale alone group was higher than the full scale group, clinicians and researchers should use caution when comparing mean scores between or within studies that used different administration approaches of the subscale. In other words, given the lack of equivalence found here, researchers and clinicians should not compare the mean of the bradykinesia subscale found in one study that administered the scale alone to that of a study that employed the standard administration. When examining change in bradykinesia severity, switching from the full UPDRS motor examination to the bradykinesia subscale alone may overestimate

change in severity of symptoms, while an underestimation might occur from the opposite ordering. However, given the reliability and validity reported here, there should be no cause for concern if the bradykinesia subscale is used alone throughout the course of the study.

Researchers and clinicians wishing to administer just the UPDRS bradykinesia subscale can be assured that it is a reliable and valid assessment of bradykinesia symptom severity. This is particularly important given the influence of bradykinesia in accounting for individual differences in health-related quality of life and depression, as well as being predictive of future functional disability. In conclusion, these results have implications beyond the bradykinesia subscale of the UPDRS: it is important that PD researchers and clinicians are aware of order effects and how varying the administration of items or scales may impact the resulting data.

ACKNOWLEDGMENTS

REFERENCES

[1]   de Lau LM & Breteler MM (2006) Epidemiology of Parkinson's disease. *Lancet Neurol*, **5**, 525-535.

[2]   Nussbaum RL & Ellis CE (2003) Alzheimer's disease and Parkinson's disease. *N Engl J Med*, **348**, 1356-1364.

[3]   de Lau LM, Koudstaal PJ, Hofman A & Breteler MM (2006) Subjective complaints precede Parkinson disease: The rotterdam study. *Arch Neurol*, **63**, 362-365.

[4]   Lang AE & Lozano AM (1998) Parkinson's disease. First of two parts. *N Engl J Med*, **339**, 1044-1053.

[5]   Olanow CW, Stern MB & Sethi K (2009) The scientific and clinical basis for the treatment of Parkinson disease. *Neurology*, **72**, S1-S136.

[6]   Buck PO, Trautman H & Clark J (2010) Scales for assessing nonmotor symptom severity changes in Parkinson's disease patients with symptom fluctuations. *Int J Neurosci*, **120**, 523-530.

[7]   Muslimovic D, Post B, Speelman JD, Schmand B & de Haan RJ (2008) Determinants of disability and quality of life in mild to moderate Parkinson disease. *Neurology*, **70**, 2241-2247.

[8]   Marras C, Rochon P & Lang AE (2002) Predicting motor decline and disability in Parkinson disease: A systematic review. *Arch Neurol*, **59**, 1724-1728.

[9]   Post B, Merkus MP, de Haan RJ & Speelman JD (2007) Prognostic factors for the progression of Parkinson's disease: A systematic review. *Mov Disord*, **22**, 1839-1851.

[10]   Koertsa J, Leenders KL, Koning M, Bouma A & van Beilen M (2008) The assessment of depression in Parkinson's disease. *Eur J Neurol*, **15**, 487-492.

[11]   Fahn S, Elton RL., members of the UPDRS Development Committee (1987) Unified Parkinson's Disease Rating Scale. In: *Recent Development in Parkinson's Disease*, S. Fahn, C.O. Marsden, D.B. Calne, M. Goldstein, eds., Macmillan Health Care Information, Florham Park, NJ, pp. 153-164.

[12]   Goetz CG, Stebbins GT, Chmura TA, Fahn S, Klawans HL & Marsden CD (1995) Teaching tape for the motor section of the unified Parkinson's disease rating scale. *Mov Disord*, **10**, 263-266.

[13]   Wilson RE, Seeberger LC, Silver D, Griffith A, Conner JB & Salzman PM, Rasagiline: Time to onset of antiparkinson effect is similar when used as a monotherapy or adjunct treatment. *The Neurologist*, in press.

[14]   Schuman H, Presser S (1981) Questions and answers in attitude surveys: Experiments on question form, wording, and context, Academic Press, New York.

[15]   Hoehn MM & Yahr MD (1967) Parkinsonism: Onset, progression and mortality. *Neurology*, **17**, 427-442.

[16]   Berger R & Hsu J (1996) Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, **11**, 283-319.

[17]   Schwab RS & England AC Jr (1969) Projection technique for evaluating surgery in Parkinson's disease. In: *Third Symposium on Parkinson's Disease*, F. Gillingham, I. Donaldson, eds., E & S Livingstone, Edinburgh pp. 152-157.

[18]   Marinus J, Visser M, Stiggelbout AM, Rabey JM, Martinez-Martin P, Bonuccelli U, Kraus PH & van Hilten JJ (2004) A short scale for the assessment of motor impairments and disabilities in Parkinson's disease: The SPES/SCOPA. *J Neurol Neurosurg Psychiatry*, **75**, 388-395.

[19]   Schuirmann DJ (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J Pharmacokinet Biopharm*, **15**, 657-680.

[20]   Lauzon C & Caffo B (2009) Easy multiplicity control in equivalence testing using two one-sided tests. *Am Stat*, **63**, 147-154.

[21]   Fisher RA (1912) On an absolute criterion for fitting frequency curves. *Messenger of Mathematics* 155-160.

[22]   Schwarz N & Bless H (1992) Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment In. *The construction of social judgments*, L.L. Martin, A. Tesser, eds., Lawrence Erlbaum Associates, Inc, Hillsdale, NJ England pp. 217-245.

[23]   Tourangeau R & Rasinski KA (1988) Cognitive processes underlying context effects in attitude measurement. *Psychol Bull*, **103**, 299-314.

[24]   Tourangeau R, Singer E & Presser S (2003) Context effects in attitude surveys: Effects on remote items and impact on predictive validity. *Sociological Methods & Research*, **31**, 486-513.

[25]   Bowling A & Windsor J (2008) The effects of question order and response-choice on self-rated health status in the English Longitudinal Study of Ageing (ELSA). *J Epidemiol Community Health*, **62**, 81-85.