ARTICLES

Psychometric Properties of Activity, Self-Efficacy, and Quality-of-Life Measures in Individuals with Parkinson Disease

Vanina Dal Bello-Haas, Laura Klassen, M. Suzanne Sheppard, Amy Metcalfe

ABSTRACT

Purpose: To examine the psychometric properties of six outcome measures in people with Parkinson disease (PD).
Method: Twenty-four participants completed the following twice within 2 weeks: the timed up-and-go test (TUG), Northwestern University Disability Scale (NUDS), Schwab & England ADL Scale (S&E), Activities-specific Balance Confidence (ABC) Scale, PD Questionnaire—Short Form (PDQ 8), and Stanford Self-Efficacy for Managing Chronic Disease 6-Item Scale (SSE). Internal consistency, test–retest reliability (ICC[3,1]), and minimal detectable change (MDC) scores were calculated. Convergent and discriminant validity of the ABC were examined.
Results: Cronbach's alpha scores for the NUDS, ABC, PDQ-8, and SSE were 0.47, 0.92, 0.72, and 0.91 respectively. The intra-class correlation coefficient (ICC[3,1]) for the TUG was 0.69 and could be improved by averaging two trials. ICCs for the NUDS, S&E, ABC, PDQ-8, and SSE were 0.56, 0.70, 0.79, 0.82, and 0.72 respectively. The ABC correlated with the TUG ($r = -0.44$, $p = 0.03$) and with PDQ-8 ($r_s = 0.51$, $p = 0.01$) and NUDS ($r_s = 0.48$, $p = 0.02$) walking items. The ABC was able to discriminate between stages 1 and 3 of disease progression but not between stages 1 and 2, which suggests that the ABC can distinguish large differences in disease progression but cannot detect more subtle differences.
Conclusions: Homogeneity of the ABC, PDQ-8, and SSE is good to excellent. Test–retest reliability scores of all measures except the NUDS are moderate to good. The ABC is a valid measure for use in PD. The MDC statistic may be useful for interpreting group score changes.

Key Words: outcome measures, Parkinson disease, reliability, validity

Dal Bello-Haas V, Klassen L, Sheppard MS, Metcalfe A. Psychometric properties of activity, self-efficacy, and quality-of-life measures in individuals with Parkinson disease. Physiother Can. 2010;preprint. doi:10.3138/ptc.2009-08

RÉSUMÉ

Objectif : Analyser les propriétés psychométriques de six mesures de résultats chez des personnes atteintes de la maladie de Parkinson (MP).
Méthode : Vingt-quatre participants ont pris part aux tests suivants durant une période de deux semaines : test chronométré « timed up-and-go » (TUG), échelle d'incapacité de l'Université Northwestern (NUDS), échelle de Schwab & England (S&E), échelle de confiance en position d'équilibre lors d'activités précises (échelle ABC), questionnaire court sur la MP, formule abrégée (PDQ-8) et échelle d'auto-efficacité de Stanford en 6 points pour la gestion des maladies chroniques (SSE). La cohérence interne, la fiabilité test-retest (CCI [3,1]) et les changements minimaux détectables ont été calculés. La validité convergente et la validité discriminante de l'échelle ABC ont également été analysées.
Résultats : Les coefficients alpha de Cronbach pour la NUDS, l'ABC, la PDQ-8 et la SSE se sont chiffrés respectivement à 0,47, 0,92, 0,72 et 0,91. Le coefficient de corrélation intraclasse (CCI [3,1]) pour le test TUG a été de 0,69 et pourrait être amélioré en faisant la moyenne de deux tests. Le CCI pour la NUDS, la S&E, l'ABC, la PDQ-8 et la SSE se sont chiffrés respectivement à 0,56, 0,70, 0,79, 0,82 et 0,72. L'ABC a été corrélé avec le test TUG ($r = -0{,}44$, $p = 0{,}03$) et avec les points concernant la marche dans la PDQ-8 ($r_s = 0{,}51$, $p = 0{,}01$) et la NUDS ($r_s = 0{,}48$, $p = 0{,}02$). L'ABC a été en mesure de séparer les étapes 1 et 3 de la progression de la maladie, mais n'a pu le faire entre les étapes 1 et 2, ce qui porte à croire que l'ABC peut différencier de plus grandes étapes dans la progression de la maladie, mais ne peut détecter les différences plus subtiles.
Conclusions : L'homogénéité de l'ABC, de la PDQ-8 et de la SSE varie de bonne à excellente. Les pointages au chapitre de la fiabilité test-retest de toutes les mesures, sauf celles de la NUDS, varient de modérés à bons. L'ABC constitue une mesure valide pour une utilisation pour la maladie de Parkinson. La statistique MDC peut être utile pour l'interprétation des changements dans les pointages du groupe.

Mots clés : fiabilité, maladie de Parkinson, mesure de résultats, validité

This work was supported in part by a grant from the Royal University Hospital Foundation, Saskatoon, Saskatchewan. A portion of this work was presented at the World Congress of Physical Therapy, Vancouver, BC, in June 2007.

The authors have no conflicts of interest to declare.

We wish to acknowledge and thank those who participated in this study. We also thank Mirna Verbancic, Saskatoon Health Region, for her assistance with the cognitive and depression screening tools, and Tasha Thornhill, School of Physical Therapy, University of Saskatchewan, for her assistance with this manuscript.

Vanina Dal Bello-Haas, PhD, MEd, BScPT: School of Physical Therapy, University of Saskatchewan, Saskatoon, Saskatchewan.

Laura Klassen, MSc, BPT: School of Physical Therapy, University of Saskatchewan, Saskatoon, Saskatchewan.

M. Suzanne Sheppard, PhD, BScPT, ACSM ES: Saskatoon Health Region, Royal University Hospital, Saskatoon, Saskatchewan.

Amy Metcalfe, MSc, BSc: Department of Community Health and Epidemiology, University of Saskatchewan, Saskatoon, Saskatchewan (at time of study).

Address correspondence to *Vanina Dal Bello-Haas*, School of Physical Therapy, University of Saskatchewan, 1121 College Drive, Saskatoon, SK S7N 0W3 Canada; Tel.: 306-966-6570; Fax: 306-966-6575;
E-mail: vanina.dalbello-haas@usask.ca.

DOI:10.3138/ptc.2009-08

INTRODUCTION

Parkinson disease (PD) is a chronic neurodegenerative disease associated with selective loss of the dopaminergic neurons in the pars compacta of the substantia nigra.[1] PD causes a variety of impairments that gradually worsen with time, and as impairments worsen, activity limitations, participation restrictions, and psychosocial problems develop. In people with PD, physical function and performance may fluctuate both during a given day and between days, for a variety of reasons, the most common being the influence of long-term pharmacological therapy.[2,3] Levodopa (L-DOPA) is currently the most effective treatment for the cardinal symptoms of PD—rigidity, bradykinesia (slowness of movement), and resting tremor[4]—but its long-term use is associated with the development of motor complications.[4] The effectiveness of L-DOPA begins to wear off at the end of a dose, causing ''wearing-off'' or ''end-of-dose'' deterioration in motor performance and worsening of symptoms. These random unpredictable fluctuations in motor performance (''on-off'' phenomenon) occur in about 50% of individuals who are treated with L-DOPA for more than 2 years. In addition, ''peak-dose'' dyskinesias, involuntary movements that occur when L-DOPA is at its peak clinical effect (about 30–60 minutes post-administration), are common.[5] Although motor fluctuations are the most common reasons for physical performance fluctuations in the later stages of PD, fluctuations in non-motor symptoms are also very prevalent,[6,7] including fatigue,[8] mood disorders,[7] slowness in thinking,[8] lack of motivation,[9] and daytime sleepiness.[10] Fluctuations resulting from these non-motor symptoms may be problematic for individuals in the earlier stages of PD. For example, Ewing Garber and Friedman found a moderate association between fatigue and performance on the timed up-and-go (TUG) test, with performance time increasing with increasing levels of fatigue, in 40 people with PD in Hoehn and Yahr Stages 0.5 to 3.0.[10]

The usefulness of clinical measures depends on their reliability and validity. *Reliability,* the extent to which a measurement is consistent and free from error, can be conceptualized as reproducibility or dependability. Although measurements are rarely perfectly infallible, a reliable clinical measure is one that provides consistent responses under given conditions.[11] Reliability is a major criterion to consider when choosing an outcome measure, because a measure's validity is inherently linked to its reliability.[12] Use of reliable outcome measures provides assurance that any significant changes in measurement score over time are due to real changes in ability rather than to measurement errors resulting from rater, instrument, or patient variability.[11] Because it cannot be assumed that an instrument is equally reliable for different client groups, reliability must be demonstrated within the population of interest. Common approaches to reliability testing include test–retest reliability, inter- and intrarater reliability, internal consistency, and alternate-forms reliability.

Validity, on the other hand, is the extent to which an instrument measures what it is intended to measure.[11] Validity is concerned with the extent to which inferences can be drawn from data and with how the results of a test can be used.[11] Like reliability, validity is not inherent but must be evaluated in the context of a clinical measure's intended use and within a specific population.[11] Determining the validity of a clinical measure is never a complete process, and numerous tests are typically required to substantiate validity. Determining the face validity, content validity, criterion-related validity, and construct validity (convergent or divergent) of a measure can provide evidence of its validity. *Convergent validity* is a function of how well a measure relates to other tests with the same constructs, whereas *divergent validity* examines the extent to which a measure does not relate to other tests with different constructs. The extent to which a measure can classify individuals into distinct groups can also be used as evidence of validity; this is known as *discriminant validity.*[13]

Reliability data for some commonly used clinical measures have previously been reported for people with PD. For example, test–retest reliability of the TUG has been examined over short intervals[14–16] and was reported as moderate to excellent, depending on whether the individual was tested during the ''on'' phase or during the ''off'' phase. Similarly, test–retest reliability of the Berg Balance Scale ($n = 37$; subjects in early, middle, and later stages of PD)[17] and the Physical Performance Test ($n = 14$; subjects in middle-stage PD)[18] over a period of 7 days has been reported to be good. Intra-observer reliability of some self-report functional measures commonly used to evaluate individuals with PD, such as the Schwab & England Activities of Daily Living Scale (S&E) and the Northwestern University Disability Scale (NUDS), has also been reported;[19,20] however, despite the frequent use of these measures in clinical trials as determinants of treatment effectiveness, evaluation of test–retest reliability is generally lacking. The psychometric properties of other common clinical and psychosocial function measures have not been widely documented for the PD population. For example, although the psychometric properties of the Activities-specific Balance Confidence (ABC) Scale among healthy older adults[21–23] and individuals with other chronic conditions[24,25] have been reported, the validity of the scale for those with PD has not been reported. The incidence of falls in people with PD is as high as 68%,[26] and an association has been demonstrated between functional mobility and balance confidence;[27] furthermore, 42% of PD fallers were reported to have a fear of future falls.[28] Thus, determining the psychometric properties of the ABC in the PD population is an important undertaking.

The high variability of performance both within and between persons with PD makes assessing physical and psychosocial function a challenge. Because random fluctuations attributable to factors such as fatigue and emotional distress, as well as drug-induced performance fluctuations in the later stages of the disease, are common in people with PD, reliable measures are especially needed if physical and psychosocial function are to be assessed adequately in both clinical and research settings. To our knowledge, no other studies have published SEM and MDC values for the S&E, PD Questionnaire—Short Form (PDQ-8), or Stanford Self-Efficacy for Managing Chronic Disease 6-Item Scale (SSE). In the present study, we examined the internal consistency, test–retest reliability, and absolute reliability of six commonly used physical and psychosocial outcome measures in individuals with early and middle-stage PD. In addition, we evaluated the convergent and discriminant validity of the ABC by examining (1) the degree to which the scale correlated with TUG scores and with walking items included in the PDQ-8 and NUDS; and (2) the extent to which the ABC discriminated among Hoehn & Yahr Classification of Disability[29] stages.

METHODS

The study was approved by the Biomedical Research Ethics Board of the University of Saskatchewan, and all participants provided signed informed consent prior to participation. All testing was completed in the physical therapy research laboratory at the University of Saskatchewan.

Participants

Participants were recruited through advertisements placed in local newspapers; flyers distributed to local seniors' centres, physicians' offices, and hospitals; and announcements on local television news programmes and at Saskatoon PD support group meetings. Participants were included in the study if they had a clinical diagnosis of PD; were in stages 1, 2, or 3 of the Hoehn & Yahr Classification of Disability[29]; and were between 40 and 80 years of age. Exclusion criteria included a history of significant cardio-pulmonary insufficiency/illness or any other condition that severely limited physical activity and a history of neurological disorder other than PD. Because participants would be completing several self-report physical activity and psychosocial measures, we also did not include people with a diagnosis of dementia or significant cognitive impairments, or those with a diagnosis of depression or other psychiatric disorder. The Mini–Mental State Exam (MMSE)[30] and the Beck Depression Inventory—II (BDI-II)[31] were used to screen participants for these impairments. Individuals with MMSE scores below 20 (reflecting moderate to severe cognitive impairment) were excluded, as were individuals with BDI scores above 20 (reflecting moderate to severe depression).

Measures

The measures examined in this study included one observed performance measure, the TUG, and five self-report measures: the NUDS; the S&E; the ABC Scale; the PDQ-8; and the SSE. Two of these are PD-specific measures, while the remainder are non-condition-specific measures. All of the self-report measures are relatively easy to administer, and each takes about 5 to 10 minutes to complete.

The Timed Up-and-Go Test (TUG)

The TUG, a commonly used measure of mobility, evaluates the time required to rise from a chair, walk 3 m at a comfortable pace, turn, return to the chair, and sit down.[32] Participants completed two TUG trials. Using the method originally described by Podsiadlo and Richardson, the first trial was considered the practice trial, and only the second trial was recorded.[32] Test–retest reliability of the TUG in people with PD has been examined using short intervals, that is, five trials with a 2-minute rest between trials ($r = 0.80$–0.98 at end of dose; $r = 0.73$–0.99 at peak dose);[14] within the same day (intra-class correlation coefficient (ICC[2,1]) = 0.94);[15] and over a 1-week period (ICC[3,2] = 0.85).[17]

The Northwestern University Disability Scale (NUDS)

The NUDS is a five-item self-report measure of walking, dressing, eating, hygiene, and speech performance[33] that is frequently used in PD clinical trials. Four subscales are scored on an ordinal scale of 0 (unable, no function) to 10 (normal); eating (including chewing/swallowing ability, diet, etc.) and feeding (including cutting food, use of cup, etc.) are each scored using a scale from 0 (complete assistance) to 5 (normal), and the two scores are combined into one sub-scale score out of 10. Sub-scale scores are added to compute a total NUDS score. Validity of the NUDS has been reported as moderate to good.[34] Interrater reliability of the NUDS total score for resident neurologists inexperienced in the use of the scale has been reported to be $k = 0.50$, with interrater reliability of individual items ranging from 0.22 (speech) to 0.65 (dressing).[19] Others have reported interrater reliability, as measured by k_w, at 0.77.[35] There are no reports of the test–retest reliability or internal consistency of the NUDS.

The Schwab & England Activities of Daily Living Scale (S&E)

The S&E is a single-item self-rated global measure of overall level of functional independence.[36] The individual is asked to rate his or her function using an 11-point

scale (10% increments), from 100% (completely independent; able to do all chores without slowness, difficulty, or impairment; essentially normal; unaware of any difficulty) to 0% (vegetative functions such as swallowing, bladder and bowels are not functioning; bedridden). The S&E, a component of the Unified Parkinson's Rating Scale (UPDRS), has become a standard PD assessment tool and has been used in hundreds of studies.[37] The construct validity of the S&E has been reported to be adequate,[38] and interrater reliability among patient, caregiver, and physician has been found to be moderate (ICC = 0.60).[20]

The Activities-specific Balance Confidence (ABC) Scale

The ABC Scale is a 16-item self-report questionnaire that assesses balance confidence in performing various tasks.[21] The scale is based on Bandura's concept of self-efficacy, defined as a belief in oneself and in one's perceived ability to perform a specific task.[39] Respondents estimate on a scale of 0% (no confidence) to 100% (complete confidence) how confident they are that they can perform various activities, such as picking a slipper up off the floor or walking on a slippery surface, without losing their balance or becoming unsteady. The individual item scores are summed and divided by 16 to yield an overall mean balance confidence score.[21] The psychometric properties of the ABC have been evaluated in frail elders[21–23] and in populations with specific medical diagnoses, including people post-stroke[24] and post-amputation.[25] Within the elderly population, internal consistency has been reported to be high ($\alpha = 0.96$);[21] 2-week test–retest reliability has been reported to be 0.92,[21,22] and discriminant[22,23] validity have been reported to be strong. Steffen and Seney reported the 1-week test–retest ICC of the ABC as 0.94, for a group of 36 individuals with PD in Hoehn & Yahr stages 1 to 4.[17]

The Parkinson's Disease Questionnaire—Short Form (PDQ-8)

The PDQ-8, a PD-specific eight-item health-status questionnaire, was developed from the PDQ-39[40] to decrease the burden for respondents.[41] The eight items ask whether the individual has experienced the following in the past month because of having PD: embarrassment; difficulty with physical complaints (painful muscle cramps and pains); difficulty getting around in public places, dressing, and communicating; and problems with depression, close relationships, and concentration. Each item is scored on an ordinal scale from 0 (never or not at all) to 4 (always or cannot do at all). Scores are summed, and the total score can range from 0 (normal) to 32 (worst disability).

The PDQ-8 items have been validated,[42] and psychometric properties of the nested version[43] and non-nested versions[44,45] of the PDQ-8 have been reported (in the nested version, the PDQ-8 was not administered as a separate instrument; rather, the PDQ-39 was administered, after which the eight items of the PDQ-8 were analyzed). Cronbach's alpha scores for items in the nested and non-nested versions ranged from 0.75[43] to 0.83[41] and from 0.81[42] to 0.84[45] respectively. The 2-week test–retest reliability of nested PDQ-8 is good (ICC = 0.80);[43] proxy reliability between patient and caregiver has also been assessed (ICC = 0.84).[45]

The Stanford Self-Efficacy for Managing Chronic Disease 6-Item Scale (SSE)

The SSE asks the respondent to rate his or her confidence on an ordinal scale from 0 (not at all confident) to 10 (totally confident) in managing different aspects of the disease.[46] The SSE includes four items related to confidence in keeping fatigue, physical discomfort or pain, emotional distress, and other symptoms or health problems from interfering with activities the person wants to do; one item relates to confidence in engaging in tasks and activities to manage the health condition; and one item relates to confidence in engaging in tasks, other than taking medications, to decrease the effects of the disease on the person's daily life. Scores are summed and divided by 6 for an overall self-efficacy score. Internal consistency of the scale has been tested in 605 study participants with a variety of chronic diseases ($\alpha = 0.91$).[47]

Procedures

Participants completed a self-report demographic and health questionnaire prior to initial testing. Information collected included age, education, medications, activity level in and around the home and community, a brief fall history, and use of health care services.

Participants then completed the following tests, in the following standardized order, on two occasions (designated as Time 1 and Time 2): TUG, NUDS, S&E, ABC, PDQ-8, and SSE. Participants were given brief rest periods as needed throughout the testing period and between the two TUG trials (practice trial and recorded trial). Participants were evaluated twice, by the same trained evaluator and at approximately the same time of the day, to control for time effects. The mean number of days between Time 1 and Time 2 was 12.9 ± 5.1 days.

Data Analysis

The Statistical Package for the Social Sciences, version 13.0 (SPSS Inc., Chicago, IL) was used for data analysis, and an alpha level of $p < 0.05$ was considered significant. Descriptive statistics were calculated. Cronbach's alpha was used to examine the internal consistency (i.e., homogeneity) of the multi-item measures (NUDS, ABC, PDQ-8, and SSE). Test–retest reliability coefficients for all six measures were determined using intra-class corre-

lation Model 3 (ICC[3,1]).[11] An ICC is the preferred estimate of reliability because it reflects both correlation and agreement;[11] with Model 3, each participant is assessed by the same raters, and the tested raters are considered the only raters of interest.[48] Test–retest reliability is considered poor if coefficients are ≤ 0.50, moderate if coefficients are between 0.50 and 0.75, and good if values are > 0.75.[11] While coefficients ≥ 0.70 are considered satisfactory for group-level comparisons, for individual comparisons and clinical decision making, reliability coefficients ≥ 0.90 may be preferred to ensure valid interpretation of findings.[11] In other words, clinicians who wish to evaluate an individual patient before and after treatment to determine whether a change has occurred (i.e., whether the treatment was effective) should consider using only outcome measures that have reliability coefficients exceeding 0.90. However, using measures with reliability coefficients ≥ 0.70 would suffice if clinicians want to evaluate whether treatment was effective, or whether observed change is true change as opposed to chance fluctuations, in a group of patients.

Absolute reliability, the measure of how an individual score varies on repeated measurement, was determined for all outcome measures using the standard error of measurement (SEM).[12] Minimal detectable change (MDC), the minimal amount of change that is not due to variation in measurement,[12] at a 95% confidence interval (MDC_{95}) was calculated for all outcome measures using the SEM, by means of the following equation:

$$MDC_{95} = SEM \times \sqrt{2} \times 1.96^{12}$$

Convergent validity, the extent to which measures relate to other tests of the same constructs,[11] was examined by determining the correlations between the ABC and the TUG and between the ABC and item 1 of the PQD-8 (*difficulty getting around in public places*) and the NUDS walking sub-scale scores, using Pearson product–moment and Spearman rank-correlation coefficients. Discriminant validity of the ABC was assessed by examining how scores at Time 1 discriminated the participants among Hoehn & Yahr's stages of disease progression, using an ANOVA and Scheffé comparison, the most rigorous post hoc comparison test.[11]

RESULTS

The 6 female and 18 male community-dwelling, ambulatory adults with PD who participated in the study ranged in age from 50 to 80 years (mean = 64.6 ± 8.0), and a majority (70.8%) had some post-secondary education. Thirteen participants (54.2%) were in Hoehn & Yahr stage 1, 6 (25.0%) in stage 2, and 5 (20.8%) in stage 3 of disease progression. Time since diagnosis ranged from less than 1 year to 20 years (mean = 4.5 ± 4.3). Mean MMSE score was 27.4 ± 2.5. More than three-quarters of the participants (78.3%) were taking a PD medication, and 44.4% were taking two or more PD medications. See Table 1 for demographic and health data of the participants.

Descriptive statistics for all measures at Time 1 and Time 2 and mean differences at Time 1 and 2 are shown in Table 2. Cronbach's alpha coefficients for the NUDS, ABC, PDQ-8, and SSE were 0.47, 0.92, 0.72, and 0.91

Table 1 Participant Demographic and Health Information ($n = 24$)

	n *(%)*	*Mean (SD) Range*
Age (years)		64.9 (8.0) 40–80
Sex		
Female	6 (25)	
Male	18 (75)	
Highest level of education completed		
Some elementary	2 (8.3)	
Some secondary	3 (12.5)	
Completed secondary	2 (8.3)	
Some college/university	4 (16.7)	
Completed college/university	10 (41.7)	
Graduate degree	3 (12.5)	
Living arrangement		
Lives with others	22 (91.7)	
Lives alone	2 (8.3)	
Employment status		
Retired	18 (75)	
Working part time	1 (4.2)	
Working full time	5 (20.8)	
Years since diagnosis		4.5 (4.3) 0–20
H & Y		
Stage 1	13 (54.2)	
Stage 2	6 (25.0)	
Stage 3	5 (20.8)	
MMSE score		27.4 (2.5) 26–30**
BDI score		9.0 (5.6) 0–19
PD medications*		
L-dopa	5 (20.8)	
Sinemet	3 (12.5)	
Amantadine	12 (50)	
Mirapex	7 (29.2)	
Requip	3 (12.5)	
Number of falls in past 2 weeks		0.25 (1.0) 0– 5
Number of near-falls in past 2 weeks		0.54 (2.0) 0–10

* Medication percentages do not total 100% because 44.4% of subjects taking two or more PD medications.

** Scores reflect age- and education-adjusted MMSE scores.

H & Y = Hoehn & Yahr stages of disease progression; MMSE = Mini–Mental State Exam scores; BDI = Beck Depression Index; PD = Parkinson disease

Table 2 Descriptive Data for the Outcome Measures at Time 1 and Time 2 ($n = 24$)

Measure	*Time 1* *Mean (SD)* *Median (Range)*	*Time 2* *Mean (SD)* *Median (Range)*	*Score Differences* *Time 1–Time 2* *Mean (SD)*	*Possible Range of Scores for Measure*
TUG	10.6s (3.7) 9.4s (6.5–20.3)	10.3s (2.5) 10.2s (6.8–17.9)	0.33 (2.48)	
NUDS	46.9 (2.4) 47 (41–50)	46.3 (3.4) 47.5 (36–50)	0.63 (2.75)	0–50
S&E	90.4 (7.9) 90 (70–100)	89.2 (8.4) 90 (70–100)	1.25 (6.30)	0–100
ABC	91.0 (9.0) 91 (66.9–100)	90.3 (8.4) 90 (71.3–100)	0.70 (5.67)	0–100
PDQ-8	25.9 (4.2) 26 (16–32)	24.9 (5.0) 25.5 (14–32)	1.04 (2.77)	0–32
SSE	8.2 (1.5) 8.5 (4.5–10.0)	8.2 (1.6) 8.4 (4.3–10.0)	0.12 (1.14)	0–10

TUG = Timed up-and-go test; NUDS = Northwestern University Disability Scale; S&E = Schwab & England Activities of Daily Living Scale; ABC = Activities-specific Balance Confidence Scale; PDQ-8 = Parkinson's Disease Questionnaire—Short Form; SSE = Stanford Self-Efficacy for Managing Chronic Disease 6-Item Scale

Table 3 Test–Retest and Absolute Reliability of Outcome Measures

Measure	*ICC*	*95% CI*	*SEM**	*SEM 95% CI***	*MDC_{95}****
TUG (s)	0.69	0.41, 0.85	1.75	3.43	4.85
NUDS	0.56	0.21, 0.78	1.94	3.80	5.38
S&E	0.70	0.43, 0.86	4.45	8.72	12.33
ABC	0.79	0.57, 0.90	4.01	7.86	11.12
PDQ-8	0.82	0.63, 0.91	1.96	3.84	5.43
SSE	0.72	0.44, 0.87	0.81	1.59	2.25

* SEM = SD of the difference scores divided by $\sqrt{2}$
** SEM 95% CI was calculated as follows: SEM $\times$ 1.96
*** $MDC_{95} = SEM \times \sqrt{2} \times 1.96$
TUG = Timed up-and-go test; NUDS = Northwestern University Disability Scale; S&E = Schwab & England Activities of Daily Living Scale; ABC = Activities-specific Balance Confidence Scale; PDQ-8 = Parkinson's Disease Questionnaire—Short Form; SSE = Stanford Self-Efficacy for Managing Chronic Disease 6-Item Scale; ICC[3,1] = intra-class correlation coefficient; SEM = standard error of measurement

respectively. For both the ABC and the SSE, the alpha would not have changed substantially with the deletion of any items; for the NUDS, Cronbach's alpha would increase to 0.57 if the speech item were deleted, and Cronbach's alpha for the PDQ-8 would increase to 0.78 if the item related to muscle cramps and pain were deleted.

Test–retest reliability (ICC[3,1]) and absolute reliability (SEM) findings for all outcome measures are summarized in Table 3. The inter-item ICC for the NUDS walking sub-scale was 0.55 (95% CI: 0.19–0.77), and ICCs for the other sub-scales ranged from −0.03 (*hygiene,* 95% CI: −0.42–0.37) to 0.79 (*dressing,* 95% CI: 0.19–0.77). The inter-item ICC for item 1 of the PDQ-8 was 0.70 (95% CI: 0.42–0.86), and ICCs for the other PDQ-8 items ranged from 0.51 (*muscle cramps and pain,* 95% CI: 0.15–0.76) to 0.83 (*unable to communicate,* 95% CI: 0.65–0.92). For the SSE, inter-item test–retest reliability ranged from 0.55 (*do things other than take medications,* 95% CI: 0.19–0.76) to 0.70 (*keep fatigue from interfering,* 95% CI: 0.42–0.62).

Descriptive statistics for individual ABC items and ICCs for the ABC are reported in Table 4. Only three items had mean scores below 90% confidence: *stand on a chair and reach for something*; *step onto or off an escalator while holding onto parcels such that you cannot hold onto the railing*; and *walk outside on icy sidewalks.* Inter-item test–retest reliability for ABC items ranged from 0.25 (95% CI: −0.16–0.59) for item 14 (*step onto or off an escalator while you are holding onto a railing*) to 0.82 (95% CI: 0.62–0.92) for item 15 (*step onto or off an escalator while you are holding onto parcels*).

ABC scores were significantly correlated with TUG scores ($r = -0.44$, $p = 0.03$), the walking sub-scale of the NUDS ($r_s = 0.48$, $p = 0.02$), and item 1 (mobility) of the PDQ-8 ($r_s = 0.51$, $p = 0.01$). Overall mean (SD), median, and range of ABC scores for each stage of disease progression are shown in Table 5. For those in

Table 4 Descriptive Statistics and Test–Retest Reliability of Activities-specific Balance Confidence (ABC) Scale Items

ABC Scale Item	*"How confident are you that you will not lose your balance or become unsteady when you …"*	*Mean (SD)*	*Median (Range)*	*ICC**	*95% CI*
1	… walk around the house?	97.9 (5.1)	100 (80–100)	0.44	0.05–0.71
2	… walk up or down stairs?	90.4 (12.3)	100 (60–100)	0.68	0.39–0.85
3	… bend over and pick up a slipper from the front of a closet floor?	90.8 (17.7)	100 (30–100)	0.65	0.34–0.83
4	… reach for a small can off a shelf at eye level?	98.3 (3.8)	100 (90–100)	0.31	−0.10–0.63
5	… stand on your tiptoes and reach for something above your head?	91.7 (10.1)	95 (70–100)	0.75	0.49–0.88
6	… stand on a chair and reach for something?	83.9 (19.0)	90 (30–100)	0.61	0.28–0.82
7	… sweep the floor?	97.4 (6.2)	100 (80–100)	0.53	0.17–0.77
8	… walk outside the house to a car parked in the driveway?	97.1 (5.5)	100 (80–100)	0.50	0.13–0.75
9	… get into or out of a car?	92.9 (9.1)	100 (70–100)	0.46	0.08–0.73
10	… walk across a parking lot to the mall?	96.7 (7.0)	100 (80–100)	0.57	0.23–0.79
11	… walk up or down a ramp?	95.0 (8.9)	100 (70–100)	0.67	0.37–0.84
12	… walk in a crowded mall where people rapidly walk past you?	91.3 (11.5)	100 (70–100)	0.56	0.21–0.78
13	… are bumped into by people as you walk through the mall?	90.4 (11.6)	95 (70–100)	0.60	0.27–0.81
14	… step onto or off an escalator while you are holding onto a railing?	90.8 (19.5)	100 (10–100)	0.25	−0.16–0.59
15	… step onto or off an escalator while holding onto parcels such that you cannot hold onto the railing?	80.0 (22.2)	90 (20–100)	0.82	0.62–0.92
16	… walk outside on icy sidewalks?	82.5 (18.7)	90 (40–100)	0.53	0.16–0.76

* ICC[3,1]

Table 5 Activities-specific Balance Confidence (ABC) Scale Scores by Stage of Disease

Hoehn & Yahr Stage	*ABC Mean (SD)*	*ABC Median*	*ABC Range*
Stage 1 ($n = 13$)	94.9 (5.8)	97.5	81.9–100
Stage 2 ($n = 6$)	90.9 (7.7)	90.6	79.4–100
Stage 3 ($n = 5$)	81.0 (10.6)	78.8	66.9–95.6

stage 1, mean ABC scores for individual items ranged from 86.00 ± 18.05 (item 4) to 100 ± 0.0 (Item 10). Scores for those in stage 2 ranged from 75.00 ± 17.61 (item 15) to 98.33 ± 4.08 (item 4). For individuals in stage 3, mean ABC scores for individual items ranged from 60.00 ± 21.60 (items 6 and 15) to 96.00 ± 5.50 (item 1). The ANOVA result was significant ($F = 6.40$, $p = 0.007$); the mean differences in ABC scores between disease stages and corresponding p-values were as follows: stage 1 and stage 2 = 3.97 (95% CI: −5.63–13.56, $p = 0.56$); stage 1 and stage 3 = 13.90 (95% CI: 3.67–24.14, $p = 0.007$); stage 2 and stage 3 = 9.94 (95% CI: −1.84–21.71, $p = 0.109$).

DISCUSSION

Estimates of consistency are generally considered good, and a scale is considered adequate, when Cronbach's alpha is > 0.70.[49] Internal consistency of three of the four multi-item self-report measures studied (ABC, PDQ-8, and SSE) was good to excellent; the ABC Scale and the SSE demonstrated the greatest homogeneity, which suggests that each participant's responses were consistent across the items for these two measures. The Cronbach's alpha would have changed minimally with the deletion of any item for the ABC and SSE, which again indicates strong item consistency. Our internal-consistency findings for the ABC and SSE are similar to those reported in other diagnostic populations.[21,47] Cronbach's alpha for the PDQ-8 in our study is lower than those previously reported for both nested[43] and non-nested[42,45] versions of this measure; administering the PDQ-8 nested within the original PDQ-39 may influence participants' responses and, in turn, alter the psychometric properties of the nested version relative to the independently administered version of the measure.[43] With respect to range, the item-total correlations of the non-nested PDQ-8 items in our study were similar to those reported in previous studies with larger sample sizes;[42] in our study, however, the item-total correlations for several items were higher—for example, item 1 (mobility) and item 5 (communication).

Although the NUDS is often used in PD clinical trials, reports on its psychometric properties are very limited. Validation of this scale has been attempted by comparing the NUDS, originally developed as a measure of disability, with impairment measures.[34] Typically, scores for each of the sub-scales are summed to provide one overall NUDS score, although the scores do not represent continuous data. In our study, the NUDS had the lowest Cronbach's alpha value of all outcome measures evaluated for internal consistency, a level well below that considered adequate for a measure. This is not

surprising, as the NUDS consists of five items, each measuring a different aspect of PD-related function—some at the impairment level and others at the activity level. In addition, one of the sub-scales consists of two separate functions (eating and feeding), the scores of which are added together. The speech item had a very low item-total correlation mean and thus did not correlate with the overall scale. Deletion of the speech item would substantially increase the alpha value of the measure, although not to the level considered adequate for a scale.

We chose a 2-week time frame to examine test–retest reliability because we believed this would be long enough to limit memory or learning effects but short enough to avoid actual changes in ability and function related to the disease, and thus in participants' responses. In previous studies,[14–16] test–retest reliability of the TUG was examined within shorter time frames than that used in this study. Although clinicians may re-evaluate TUG within a treatment session or before and after a treatment session, clinicians also re-evaluate over longer time frames in order to assess retention of changes in performance or treatment effectiveness. As noted above, we had participants complete one practice trial and recorded their time on a second trial, the methodology originally described by Podsiadlo and Richardson.[32] Our ICC of 0.69 for the TUG is considered moderate; however, we found that taking the average of the two trials would increase the test–retest ICC to 0.76 (95% CI: 0.51–0.89). Similarly, Steffen and Seney found reproducibility over a 1-week period of two averaged trials of the TUG (after one practice trial) to be good (ICC = 0.85).[17] These findings suggest that for people with PD, at least two TUG trials should be averaged to enhance reliability.

We found no reports in the literature of test–retest reliability for the NUDS, S&E, or SSE. In our study, the test–retest reliability ICC of the NUDS was low to moderate, while the values obtained for the S&E, the non-nested PDQ-8, and the SSE were moderate to good. The ICC for the PDQ-8 in our study is comparable to the 2-week test–retest reliability previously reported[43] for the non-nested version of the PDQ-8.

The test–retest reliability of the ABC in our study was good, but lower than has been reported in other studies with other populations—older adults (2-week interval, $n = 21$, $r = 0.91$);[21] people at least 1 year post-stroke (4-week interval, $n = 24$, ICC = 0.85);[24] and people post-amputation (4-week interval, $n = 50$, ICC = 0.91).[25] In addition, inter-item test–retest ICCs for individual ABC items were much lower than values previously reported for a stroke population.[24] Given the nature of PD, it would not be unusual to see more variability in scores for individuals with PD than in groups of individuals with more stable conditions.

Steffen and Seney reported 1-week test–retest ICC of the ABC as 0.94;[17] however, their participants were not comparable to those in our study: average disease duration was more than three times as long (14 ± 6 years), and distribution of disability was wider (three participants were in Hoehn & Yahr stage 1, seven in stage 2, nine in stage 3, and eight in stage 4).[17] During the administration of the ABC, we found that several participants needed to be redirected to the main question posed by the ABC; that is, they had to be reminded to distinguish between their *level of balance confidence* in performing each task and their *usual level of participation* in each activity. This observation suggests that consistent interpretation of this central concept is necessary to ensure reliable responses over time, which may require that the questionnaire be completed during an in-person interview.

Although the test–retest ICC values we obtained for the averaged TUG, ABC, S&E, PDQ-8, and SSE are considered moderate to good from a statistical perspective and acceptable for group-level comparisons, they do not meet the 0.90 level considered acceptable for individual comparisons.[11] An ICC is influenced by error as well as by sources of variation attributable to respondents, raters, and trials. We designed our study to limit variation due to trials (e.g., same time of day, same order of tests), and the rater was the same for both trials. It is unlikely that overall ability and function would change significantly in 2 weeks; however, individuals' perceptions or self-assessments of ability may have changed as a result of sensitization to abilities produced by the first day of testing. The very narrow range of measurement scores among participants in our sample more than likely influenced the results. Demonstrating reliability requires a certain amount of variability among participants' scores;[11] in other words, homogeneity of scores affects reliability.[11]

For every score derived from a measure, there is a range of scores within which the actual score will lie. Thus, in addition to measuring the reliability of a clinical measure, assessing the consistency or stability of repeated responses (*response stability*) over time is important.[11] The standard error of measurement (SEM)—that is, absolute reliability—is commonly used to express response stability and is calculated from test–retest reliability findings.[12] The SEM, which is influenced by error variation only, is expressed in the same units as the original measurement.[12] Using the SEM, we calculated the MDC_{95} for all outcome measures. Clinicians can use the SEM to determine the range of scores that can be expected on retesting and the MDC statistic for a particular measure to interpret whether a change in score can be considered true change resulting from an intervention or whether it is attributable to measurement error.[11,12] Because the test–retest ICC values obtained in our study for the averaged TUG, as well as for the ABC, S&E, PDQ-8, and SSE, are considered acceptable only for group-level comparisons, clinicians should exercise caution when using our MDCs to determine change in a particular individual.

Minimal detectable change values for the TUG and ABC have been reported previously, but no previous research has reported MDC values for the NUDS, S&E, PDQ-8, or SSE. Steffen and Seney reported a MDC_{95} of 13 points for the ABC and 11 seconds for the TUG.[17] The TUG value is much greater than our finding. Both our study and the Steffen and Seney study[17] included participants in several Hoehn & Yahr stages; determining the MDC statistic for outcome measures for each of the stages may provide more clinically meaningful information.

Like Hatch et al.,[27] we found a negative correlation between ABC scores and recorded TUG time, which suggests that balance confidence and functional mobility are inversely related; in our study, however, the relationship ($r = -0.44$, $p = 0.03$) was not as strong as previously reported.[27] We also found that the ABC was correlated with items within other measures that examine walking ability (the walking sub-scale of the NUDS and item 1 of the PDQ-8), providing evidence that balance confidence and walking ability are related. Similar findings have been reported in studies of older adults: older adults with a fear of falling restricted their activity and reported declines in mobility,[50,51] and reduced physical function has been found in older people with low fall-related self-efficacy.[52] Individuals with PD who have decreased balance confidence may similarly restrict their activity levels and experience declines in mobility.

ABC scores differed significantly between stage 1 and stage 3 of disease progression, which indicates that the ABC was able to discriminate between these two stages but not between consecutive stages. The clinical hallmark of stage 3 is the onset of postural instability, as measured by the pull test (item 30 of the UPDRS).[53] Our findings suggest that the ABC can detect large differences in disease progression but does not detect more subtle differences. The very small sample sizes for stage 2 and stage 3, and the lack of variability in scores both overall and within each disease stage, may have influenced our results and limited our ability to evaluate the discriminative ability of the ABC: participants in stage 2 scored themselves below 80% confidence for only one ABC item, and those in stage 3 scored themselves below 80% confidence for only four items.

The measures we chose to examine represent evaluation across a variety of constructs, from quality of life (PDQ-8) and self-efficacy (ABC, SSE) to activity (NUDS, S&E) and impairments (ABC, TUG), all of which are important when evaluating outcomes of wellness interventions for persons in the earlier stages of PD. For example, self-efficacy, an individual's perceived capability within a specific domain of activities,[39] is gaining in importance as a concept to be considered and evaluated when examining the effectiveness of wellness interventions for chronic conditions, especially if interventions include education or home-based programmes.[54] The concept of self-efficacy may provide useful insight into the psychological or cognitive–motivational aspects of a person's functional and participation abilities. Perceived ability is more predictive of behaviour than is actual ability, and low self-efficacy can result in avoidance of tasks.[39]

The distribution of participants among stages 1–3 of disease progression, with the majority in stages 1 and 2 (and $>50\%$ in stage 1), is consistent with the level of disability that would be expected in community-dwelling, ambulatory persons with PD. We examined people in the early and middle stages of PD because we anticipated that they would have relatively stable symptoms and would be on relatively stable medication protocols. We were also interested in using these outcome measures in a study examining the effectiveness of community-based group wellness programmes geared toward those in the early and middle stages of PD. Although we expected stability of symptoms, we did not expect such a limited range of measurement scores, as indicated by group mean scores in each of the disease stages. For example, the mean score for each ABC item for our sample ranged from 80% to 98%. These mean scores are much higher, and the range is much more limited, than those previously reported;[55] a previous study reported mean scores for individual ABC items ranging from 43% to 81% in a sample of 58 individuals with PD (mean age $= 66.2 \pm 9$ years; years since diagnosis $= 6.5 \pm 4.9$; disease severity not reported).[55]

For all self-report measures except the PDQ-8, the large majority of participants scored themselves toward the higher end of the range, regardless of disease stage, balance problems (as indicated by reported number of falls and near falls in the past 2 weeks), and walking speed (as indicated by the TUG). All measures except the TUG were self-report. Self-report measures are very important to consider, but they may not reflect actual performance, which can be influenced by a range of factors. It is possible that the participants did not want to indicate their true level of function to the evaluator or that they did not understand the intent of the questions being asked or of the concepts being investigated (e.g., balance confidence; see discussion of ABC Scale above). Leritz et al. found that although there were no differences between patient and caregiver ratings on the motor and activities of daily living portions of the UPDRS, patients significantly under-reported levels of impaired function relative to their caregivers on the Activities of Daily Living Scale and the S&E, regardless of cognitive status (as measured by the MMSE) and regardless of side of PD disease focus.[56] Although the individuals in Leritz et al.'s study were at a more advanced stage of disease progression (Hoehn & Yahr stage 3 and above; mean years with PD $= 14.13 \pm 7.84$ for those with left PD and 13.00 ± 3.88 for those with right PD)[56] than those in our study, it is possible that individuals in

our study may also have overestimated their performance on some measures.

LIMITATIONS

This study had several limitations, some of which are described above. The sample ($n = 24$) was relatively small, and the majority of these participants rated themselves toward the high end on all but one of the self-report scales. These ceiling effects limited the range of PD-related impairments, activity limitations, and participation restrictions observed in the study, and the lack of variability among participants' scores may have influenced the results. More than half the sample was in stage 1 of the Hoehn & Yahr Classification of Disease Progression, which further limited the range of PD-related problems observed and reduced the generalizability of our results to persons who are farther along in the disease process. In addition, five of the six measures used in the study were self-report tools, all of which use ordinal rating scales. The use of ordinal measures also likely contributed to ceiling effects. If we had included more performance measures with continuous scales, such as the TUG, there might have been more variation in scores among participants.

CONCLUSION

The high variability of performance both within and among persons with PD makes assessing physical and psychosocial function over time a challenge. Our results suggest that the ABC, S&E, PDQ-8, and SSE measures have moderate to excellent internal consistency and can provide reliable test–retest values in community-dwelling populations with PD. Results also suggest that the ABC is a valid measure for use with PD populations in the earlier stages of the disease. We do not recommend that the NUDS be used for either individual or group-level comparisons because of its poor homogeneity and low to moderate test–retest reliability.

The SEMs and MDCs for the measures examined in this study may assist therapists in determining whether change in a group of individuals with PD, such as those participating in community wellness programmes, is due to testing error or is the result of any interventions. Additional studies with larger sample sizes that allow for the calculation of values for each stage of PD disability are needed.

KEY MESSAGES

What Is Already Known on This Topic

Intra-observer reliability of some self-report functional measures commonly used to evaluate individuals with PD has been reported, but test–retest reliability evaluation of these measures is generally lacking. The psychometric properties of common clinical and psychosocial function measures (such as the ABC Scale) that may be useful for documenting change in people with PD have not been examined in this patient population.

What This Study Adds

The ABC, S&E, PDQ-8, and SSE are acceptable tests and measures to use for group-level comparisons in people with PD. The TUG is also acceptable for group-level comparisons over time; however, when the TUG is used to examine change in people with PD, at least two trials should be averaged to decrease measurement error. The lack of homogeneity and reliability of the NUDS makes it unacceptable for use for both individual- and group-level comparisons.

REFERENCES

1. Lang AE, Lozano AM. Parkinson's disease—first of two parts. N Engl J Med. 1998;339:1044–53.
2. Obeso JA, Grandas F, Vaamonde J, Luquin MR, Artieda J, Lera G, et al. Motor complications associated with chronic levodopa therapy in Parkinson's disease. Neurology. 1989;39:11–9.
3. Sage JI, Mark MH. Basic mechanisms of motor fluctuations. Neurology. 1994;44:S10–4.
4. Marsden CD, Parkes JD. Success and problems of long-term levodopa therapy in Parkinson's disease. Lancet. 1977;1:345–9. doi:10.1016/S0140-6736(77)91146-1
5. Nutt JG, Carter JH, Lea ES, Woodward WR. Motor fluctuations during continuous levodopa infusions in patients with Parkinson's disease. Mov Disord. 1997;12:285–92. doi:10.1002/mds.870120304
6. Shulman LM, Taback RL, Rabinstein AA, Weiner WJ. Non-recognition of depression and other non-motor symptoms in Parkinson's disease. Parkinsonism Relat Disord. 2002;8:193–7. doi:10.1016/S1353-8020(01)00015-3
7. Witjas T, Kaphan E, Azulay JP, Blin O, Ceccaldi M, Pouget J, et al. Nonmotor fluctuations in Parkinson's disease. Neurology. 2002;59:408–13.
8. Ewing Garber C, Friedman JH. Effects of fatigue on physical activity and function in patients with Parkinson's disease. Neurology. 2003;60:1119–24.
9. Sullivan KL, Ward CL, Hauser RA, Zesiewicz TA. Prevalence and treatment of non-motor symptoms in Parkinson's disease [abstract]. Parkinsonism Relat Disord. 2007;13:545. doi:10.1016/j.parkreldis.2006.10.008
10. Friedman JH, Fernandez HH. The nonmotor problems of Parkinson's disease. Neurologist 2000;6:8–27.
11. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2000.
12. Stratford PW. Getting more from the literature: estimating the standard error of measurement from reliability studies. Physiother Can. 2004;56:27–30. doi:10.2310/6640.2004.15377
13. Streiner D, Norman G. Health measurement scales: a practical guide to their development and use. 3rd ed. Oxford: Oxford University Press; 2003.
14. Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the timed ''up & go'' test in people with Parkinson disease. Phys Ther. 2001;81:810–8.
15. Brusse KJ, Zimdars S, Zalewski KR, Steffen TM. Testing functional performance in people with Parkinson disease. Phys Ther. 2005;85:134–41.

16. Morris ME, Matyas TA, Iansek R, Summers JJ. Temporal stability of gait in Parkinson's disease. Phys Ther. 1996;76:763–89.
17. Steffen T, Seney M. Test–retest reliability and minimal detectable change on balance and ambulation tests, the 36-Item Short-Form Health Survey, and the Unified Parkinson Disease Rating Scale in people with parkinsonism. Phys Ther. 2008;88:733–46. doi:10.2522/ptj.20070214
18. Paschal K, Oswald A, Siegmund R, Siegmund S, Threlkeld AJ. Test–retest reliability of the physical performance test for persons with Parkinson disease. J Geriatr Phys Ther. 2006;29:82–6.
19. Geminiani G, Cesana BM, Tamma F, Contri P, Pacchetti C, Carella F, et al. Interobserver reliability between neurologists in training of Parkinson's disease rating scales: a multicenter study. Mov Disord. 1991;6:330–5. doi:10.1002/mds.870060411
20. McRae C, Diem G, Vo A, O'Brien C, Seeberger L. Reliability of measurements of patient health status: a comparison of physician, patient and caregiver ratings. Parkinsonism Relat Disord. 2002;8:187–92. doi:10.1016/S1353-8020(01)00009-8
21. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) scale. J Gerontol A–Biol. 1995;50A:M28–34.
22. Myers AM, Fletcher PC, Myers AH, Sherk W. Discriminative and evaluative properties of the Activities-specific Balance Confidence (ABC) scale. J Gerontol A–Biol. 1998;53A:M287–94.
23. Myers AM, Powell LE, Maki BE, Holliday PJ, Brawley LR, Sherk W. Psychological indicators of balance confidence: relationship to actual and perceived abilities. J Gerontol A–Biol. 1996;51A:M37–43.
24. Botner EM, Miller WC, Eng JJ. Measurement properties of the Activities-specific Balance Confidence scale among individuals with stroke. Disabil Rehabil. 2005;27:156–63. doi:10.1080/09638280400008982
25. Miller WC, Deathe AB, Speechley M. Psychometric properties of the Activities-specific Balance Confidence scale among individuals with a lower-limb amputation. Arch Phys Med Rehabil. 2003;84:656–61. doi:10.1016/S0003-9993(02)04807-4
26. Wood BH, Bilclough JA, Bowron A, Walker RW. Incidence and prediction of falls in Parkinson's disease: a prospective multidisciplinary study. J Neurol Neurosur PS. 2002;72:721–5. doi:10.1136/jnnp.72.6.721
27. Hatch J, Gill-Body KM, Portney LG. Determinants of balance confidence in community-dwelling elderly people. Phys Ther. 2003;83:1072–9.
28. Bloem BR, Grimbergen YAM, Cramer M, Willemsen M, Zwinderman AH. Prospective assessment of falls in Parkinson's disease. J Neurol. 2001;248:950–8. doi:10.1007/s004150170047
29. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology. 1967;17:427–42.
30. Folstein MF, Folstein SE, McHugh PR. Mini–Mental State: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98. doi:10.1016/0022-3956(75)90026-6
31. Beck AT, Steer RA, Brown GK. Manual for Beck Depression Inventory II (BDI-II). San Antonio, TX: Psychology Corp.; 1996.
32. Podsiadlo D, Richardson S. The timed ''up & go'': a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39:42–8.
33. Canter CJ, De La Torre R, Mier M. A method for evaluating disability in patients with Parkinson's disease. J Nerv Mental Dis. 1961;133:143–7.
34. Henderson L, Kennard C, Crawford TJ, Day S, Everitt BS, Goodrich S, et al. Scales for rating motor impairment in Parkinson's disease: studies of reliability and convergent validity. J Neurol Neurosur PS. 1991;54:18–24. doi:10.1136/jnnp.54.1.18
35. Ginanneschi A, Degl'Innocenti F, Maurello MT, Magnolfi S, Marini P, Amaducci L. Evaluation of Parkinson's disease: a new approach to disability. Neuroepidemiology. 1991;10:282–7. doi:10.1159/000110285
36. Schwab R, England A. Projection technique for evaluating surgery in Parkinson's disease. In: Gillingham J, Donaldson I, editors. Third symposium on Parkinson's disease. Edinburgh: Livingstone; 1969.
37. Ramaker C, Marinus J, Stiggelbout AM, van Hilten BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord. 2002;17:867–76. doi:10.1002/mds.10248
38. Forjaz MJ, Martinez-Martin P. Metric attributes of the Unified Parkinson's Disease Rating Scale 3.0 Battery: part II, construct and content validity. Mov Disord. 2006;21:1892–8. doi:10.1002/mds.21071
39. Bandura A. Self-efficacy mechanism in human agency. Am Psychol. 1982;37:122–47. doi:10.1037/0003-066X.37.2.122
40. Peto V, Jenkinson C, Fitzpatrick R, Greenhall R. The development and validation of a short measure of functioning and well-being in patients with Parkinson's disease. Qual Life Res. 1995;4:241–8.
41. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The PDQ-8: development and validation of a short-form Parkinson's disease questionnaire. Psychol Health. 1997;12:805–14. doi:10.1080/08870449708406741
42. Tan LCS, Lau PN, Aua WL, Luo N. Validation of PDQ-8 as an independent instrument in English and Chinese. J Neurol Sci. 2007;255:77–80. doi:10.1016/j.jns.2007.01.072
43. Tan LCS, Luo N, Nazri M, Li SC, Thumboo J. Validity and reliability of the PDQ-39 and the PDQ-8 in English-speaking Parkinson's disease patients in Singapore. Parkinsonism Relat Disord. 2004;10:493–9. doi:10.1016/j.parkreldis.2004.05.007
44. Jenkinson C, Fitzpatrick R. Cross-cultural evaluation of the short form 8-item Parkinson's Disease Questionnaire (PDQ-8): results from America, Canada, Japan, Italy and Spain. Parkinsonism Relat Disord. 2007;13:22–8. doi:10.1016/j.parkreldis.2006.06.006
45. Martinez-Martin P, Benito-Leon J, Alonso F, Catalan MJ, Pondal M, Zamarbide I. Health-related quality of life evaluation by proxy in Parkinson's disease: approach using the PDQ-8 and EuroQol 5D. Mov Disord. 2004;19:312–8. doi:10.1002/mds.10656
46. Ware JE Jr, Nelson EC, Sherbourne CD, Stewart AL. Preliminary tests of a 6-item general health survey: a patient application. In: Stewart AL, Ware JE Jr, editors. Measuring functioning and well-being: the medical outcomes study approach. Durham, NC: Duke University Press; 1992. p. 291–303.
47. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program for patients with chronic disease. Effect Clin Pract. 2001;4:252–62.
48. Shrout PE, Fleiss JL. Intraclass correlation: uses in assessing rater reliability. Psychol Bull. 1979;86:420–8. doi:10.1037/0033-2909.86.2.420
49. Nunnally JC Jr. Psychometric theory. 2nd ed. New York: McGraw-Hill; 1978.
50. Vellas B, Cayla F, Bocquet H, de Pemille F, Albarede JC. Prospective study of restriction of activities in old people after falls. Age Ageing. 1987;16:189–93.
51. Vellas BJ, Wayne SJ, Romero LJ, Baumgartner RN, Garry PJ. Fear of falling and restriction of mobility in elderly fallers. Age Ageing. 1997;26:189–93. doi:10.1093/ageing/26.3.189
52. Tinetti ME, Mendes de Leon GF, Doucette JT, Baker DI. Fear of falling and fall-related efficacy in relationship to functioning among community-living elders. J Gerontol. 1994;49:M140–7.
53. Fahn S, Elton R. Unified Parkinson's disease rating scale. In: Fahn S, Marsden D, Calne D, editors. Recent developments in Parkinson diseases. London: Macmillan; 1987. p. 153–63.
54. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. J Am Med Assoc. 2002;288:2469–75. doi:10.1001/jama.288.19.2469
55. Adkin AL, Frank JS, Jog MS. Fear of falling and postural control in Parkinson's disease. Mov Disord. 2003;18:496–502. doi:10.1002/mds.10396
56. Leritz E, Loftis C, Crucian G, Friedman W, Bowers D. Self-awareness of deficits in Parkinson disease. Clin Neuropsychol. 2004;18:352–61. doi:10.1080/1385404049052412