Research Report

Reliability of Measurements Obtained With the Timed "Up & Go" Test in People With Parkinson Disease

Background and Purpose. The Timed "Up & Go" Test (TUG) is used to measure the ability of patients to perform sequential locomotor tasks that incorporate walking and turning. This study investigated the retest reliability, interrater reliability, and sensitivity of TUG scores in detecting changes in mobility in subjects with idiopathic Parkinson disease (PD). **Subjects.** The performance of 12 people with PD was compared with that of 12 age-matched comparison subjects without PD. **Methods.** The subjects with PD completed 5 trials of the TUG after withdrawal of levodopa for 12 hours ("off" phase of the medication cycle) as well as an additional 5 trials 1 hour after levodopa was administered ("on" phase of the medication cycle). They were scored on the Modified Webster Scale at both sessions. The comparison subjects also performed 5 TUG trials. All trials were videotaped and timed by 2 experienced raters. The videotape was later rated by 3 experienced clinicians and 3 inexperienced clinicians. **Results.** For the subjects with PD, within-session performance was highly consistent, with correlations (r) ranging from .80 to .98 for the "off" phase and from .73 to .99 for the "on" phase. The performance of the comparison subjects across the 5 trials was also highly consistent (r=.90–.97). Comparisons showed differences between trials 1 and 2 on the TUG for both groups. Removal of data for trial 1 (the practice trial) further enhanced retest reliability. There was close agreement in TUG scores among raters despite different levels of experience (intraclass correlation coefficient [3,1]=.87–.99). Mean TUG scores were different between the "on" and "off" phases of the levodopa cycle and between subjects with PD and comparison subjects during the "on" phase. **Conclusion and Discussion.** Retest reliability and interrater reliability of the TUG measurements were high, and the measurements reflected changes in performance according to levodopa use. The TUG can also be used to detect differences in performance between people with PD and elderly people without PD. [Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the Timed "Up & Go" Test in people with Parkinson disease. *Phys Ther.* 2001;81:810–818.]

Key Words: *Gait, Levodopa, Locomotion, Mobility, Parkinson disease, Physical therapy, Turning.*

Susan Morris

Meg E Morris

Robert Iansek

Physical therapists play a major role in assessing the ability of people with Parkinson disease (PD) to perform complex motor tasks that are routinely performed in everyday life.[1–4] They also teach people strategies for moving more easily.[5–7] Physical therapists need reliable and valid measurements that can reflect mobility during goal-directed locomotor tasks in people with PD because these tasks are often difficult for people with PD.[8,9]

The "Get-up and Go" Test[10] was designed to measure mobility in elderly people and has been advocated as a useful tool for quantifying locomotor performance in people with PD.[11] This test requires people to stand up from a chair, walk 3 m, turn around, walk back to the chair, and sit down again. The subject is videotaped, and then mobility is rated on a 5-point scale, ranging from "normal" to "severely abnormal." Mathias et al[10] found only moderately good correlations (r=.21–.75) between "Get-up and Go" Test scores and laboratory measurements of gait and balance in elderly people (mean age=73.8 years, range=52–94 years), although there were differences in ratings between medical practitioners and physical therapists. One of the difficulties with this test is that guidelines for rating the severity of abnormality are not provided by the test developers, other than to say that "severely abnormal means that the subject appears at risk of falling during the test performance."[10(p387)] The lack of guidelines may account for why there is wide interrater variability in scores and difficulty grading the performance for intermediate zones of the scale.

To increase the reliability of the measurements while ensuring that the test continued to be quick and easy to administer, Podsiadlo and Richardson[12] modified the "Get-up and Go" Test to incorporate a timed component. When validating the Timed "Up & Go" Test (TUG), a selection of elderly patients (aged 60–90 years) with stroke, PD, arthritis, cerebellar disorders, and

S Morris, MAppSc, BAppSc(PT), Grad Dip (PT), is a private practitioner in Melbourne, Victoria, Australia.

ME Morris, PhD, MAppSc, BAppSc(PT), Grad Dip (Gerontology), is Professor and Head, School of Physiotherapy, La Trobe University, Victoria 3086, Australia (m.morris@latrobe.edu.au). Address all correspondence to Dr Morris.

R Iansek, PhD, FRACP, is Neurologist and Director, Movement Disorders Program, Kingston Centre, Cheltenham, Victoria 3192, Australia.

Susan Morris and Meg Morris provided concept/research design, writing, and data analysis. Susan Morris provided data collection and project management, and Meg Morris provided fund procurement. Robert Iansek provided subjects and administrative support.

The authors thank the Geriatric Research Unit at Kingston Centre for its support, the participants of this study for their time and commitment given, Ms Winnie Tang, PT, for her valued work as research assistant, the physical therapists at Kingston Centre for their support and comments on the manuscript, and Dr Thomas Matyas of La Trobe University.

This experiment was conducted in partial fulfillment of the requirements for Susan Morris' postgraduate diploma at the School of Physiotherapy, La Trobe University.

This study was approved by the ethics committees at La Trobe University and Kingston Centre.

This research was undertaken with the assistance of a grant from the Victorian Branch Research Committee, a professional body within the Australian Physiotherapy Association.

This article was submitted November 10, 1999, and was accepted August 30, 2000.

general deconditioning followed the same procedure. The task, however, was timed using a stopwatch. Excellent agreement in timed scores was found between raters (intraclass correlation coefficient [ICC]=.99) as well as for the same rater during consecutive clinic visits (ICC=.99). Moreover, the TUG times correlated moderately well with gait speed ($r=-.55$), scores on the Berg Balance Scale ($r=-.72$), and the Barthel Activities of Daily Living Index ($r=-.51$). Although Podsiadlo and Richardson included data obtained for 10 patients with PD in their analysis, the PD data were not presented separately, and the retest reliability and interrater reliability of the TUG scores for patients with idiopathic PD were not documented.

Although Martinez-Martin et al[13] showed that TUG scores in a sample of 50 subjects with PD ranged from 5 to 31, with a mean of 14.6 (SD 6.9), they did not evaluate the reliability and validity of the measurements for this patient population. Berg et al[14] found good correlations between the TUG and the Berg Balance Scale ($r=-.76$) and Tinetti Balance Scale ($r=.74$), yet they confined their analyses to older people without PD. Other researchers[14–20] have used the TUG to measure mobility in elderly people (aged 60–99 years). Although observations suggest that the TUG may be a useful test that is quick and easy to administer, a systematic and controlled trial examining the validity of inferences from the measurements for people with PD has not yet been conducted.

The purposes of our investigation were: (1) to quantify the reliability of measurements obtained with the TUG and (2) to examine whether the measure could be used to detect differences in motor performance across the PD medication cycle. We measured retest reliability of measurements obtained with the TUG over 5 trials in subjects with PD 12 hours after withdrawal of levodopa and then 1 hour after the first dose was given. Reliability over 5 trials also was examined for age- and sex-matched comparison subjects. Interrater reliability of TUG scores was examined for experienced and inexperienced raters. In addition, to assess the discriminative ability of the measure, we examined whether the TUG scores could be used to differentiate between the performance of subjects with PD 12 hours after withdrawal of medication ("off" phase of the medication cycle) and their performance 1 hour after the next dose of levodopa was administered ("on" phase of the medication cycle). The ability to use the measure to detect differences in performance between subjects with PD and comparison subjects also was examined.

Method

Subjects

Twenty-four subjects over the age of 50 years were recruited for the study. The subjects had a mean age of 65.5 years (SD=10.5, range=50–81). Twelve subjects had idiopathic PD and were recruited from the Movement Disorders Clinic at Kingston Centre, Cheltenham, Victoria, Australia. Subjects were hospital inpatients undergoing detailed assessment by the treating neurologist. We excluded subjects younger than 50 years of age because we believe the literature shows that early-onset PD tends to be more aggressive than late-onset PD, with consequent increases in movement variability.[21,22] In addition, 12 comparison subjects were selected from the volunteer services division at Kingston Centre. They were matched for height, sex, and age because these factors are known to affect gait speed.[23]

To be included in the study, subjects had to be able to walk at least 10 m unassisted and without orthoses. We chose this distance because the TUG involves a walk of 6 m. Subjects also had to be willing and able to give informed consent according to the Declaration of Helsinki (1964). Subjects were excluded if they had visual impairment or musculoskeletal, neurological, or cardiovascular disorders that affected locomotion. Subjects with diabetes also were excluded because there was a fasting period from 8:00 PM the day before testing was completed until approximately 10:00 AM the next day. Fasting was necessary because food digestion can have a variable effect on the absorption of levodopa. Subjects who scored less than 29 on the Short Test of Mental Status[24] also were excluded because this score predicts dementia in 95% of cases.[24] In order to keep within-group variability to a minimum, subjects with severe dystonia or dyskinesia (≥7 on the dyskinesia subsection) as judged by use of the United Parkinson's Disease Rating Scale (UPDRS)[25] also were excluded from the study.[26] The characteristics of the subjects with PD and the comparison subjects are summarized in Tables 1 and 2.

The subjects with PD (5 men, 7 women) ranged in age from 50 to 81 years and in height from 146 to 185 cm ($\overline{X}$=165.3, SD=10.2). The comparison subjects were age matched (±2 years) and height matched (±10 cm). In the sample, the mean age was 68.8 years (SD=10.4) for the subjects with PD and 69.9 years (SD=11.1) for the comparison subjects. The mean duration of PD was 9.6 years (SD=6.4, range=1–26). The height of the comparison subjects ranged from 152 to 178 cm ($\overline{X}$=165.7, SD=6.85). The mean Modified Webster Scale[27] score at the end of the dose was 14.2 (SD=4.4, range=9–23). The mean Modified Webster Scale scores at peak dose was 10.2 (SD=4.9, range=3–17).

Table 1.
Characteristics of Subjects With Parkinson Disease (PD)

Subject No.	Age (y)	Sex	Height (m)	Modified Webster Scale[27] Score "Off" Phase	Modified Webster Scale[27] Score "On" Phase	Duration of PD (y)	Medication	Dosage (mg/d)
1	59	F	1.65	18	15	26	Sinemet liquid[a]	10 mL/h
2	76	F	1.46	17	15	14	Eldepryl[b]	10
							Madopar[c]	600/150
							Madopar HBS[c]	300/75
3	71	F	1.55	23	16	6	Permax[d]	0.5
							Madopar	600/150
							Eldepryl	10
4	50	F	1.64	11	7	11	Madopar	1,200/300
							Eldepryl	5
5	51	F	1.64	9	3	12	Permax	1
							Madopar	100/25
							Sinemet CR[a]	500/125
6	71	M	1.85	14	14	12	Sinemet CR	1,300/325
							Sinemet	150/37.5
							Antadine[e]	200
7	69	M	1.67	15	8	5	Madopar	600/150
							Permax	0.1
8	76	M	1.73	15	7	10	Madopar HBS	400/100
							Madopar	450/112.5
							Permax	0.75
9	81	M	1.76	19	17	1	Madopar	300/75
10	81	F	1.60	10	8	6	Sinemet	500/125
							Sinemet CR	200/50
11	71	M	1.70	10	7	15	Sinemet	1,250/125
							Permax	3
							Artane[f]	2
12	70	F	1.59	9	5	7	Sinemet	1,500/150

[a] Merck Sharp and Dome, 54-68 Ferndell St, South Granville, New South Wales 2142, Australia.
[b] Reckitt and Colman, 44 Wharf Rd, West Ryde, New South Wales 2114, Australia.
[c] Roche, 4-10 Inman Rd, Dee Why, New South Wales 2099, Australia.
[d] Eli Lilly, 112 Wharf Rd, West Ryde, New South Wales 2114, Australia.
[e] Boots, 21 Loyalty Rd, North Rock, New South Wales 2151, Australia.
[f] Lederle, 5 Gibbon Rd, Baulkham Hills, New South Wales 2153, Australia.

Table 2.
Characteristics of Comparison Subjects

Subject No.	Age (y)	Sex	Height (m)
1	61	F	1.66
2	74	F	1.60
3	71	F	1.58
4	50	F	1.66
5	50	F	1.60
6	71	M	1.78
7	70	M	1.70
8	74	M	1.66
9	83	M	1.72
10	83	F	1.65
11	79	M	1.63
12	73	F	1.52

The Modified Webster Scale[27] is a test based on the original Webster Scale developed for PD. In addition to the 10 items contained in the original scale (bradykinesia, rigidity, posture, gait, arm swing while walking, tremor, facies, seborrhea, speech, self-care), 2 items have been added. These items are balance and rising from a chair.[27] All items are rated on a 4-point scale (0–3), giving a possible score of 36. Increasingly higher scores indicate increasing degree of disability. Measurements obtained with the scale have been demonstrated to have good reliability and criterion-related validity by several studies.[5,26]

Apparatus

Administration of the TUG requires subjects to stand up from a chair, walk 3 m, turn around, walk back to the chair, and sit down.[12] The starting position was standardized so that the subjects commenced the test with their feet flat on the floor and their arms resting on the armrests. Two physical therapists with more than 7 years of experience in treating patients with neurological problems timed the performance for each trial using stopwatches. Raters were not advised prior to testing

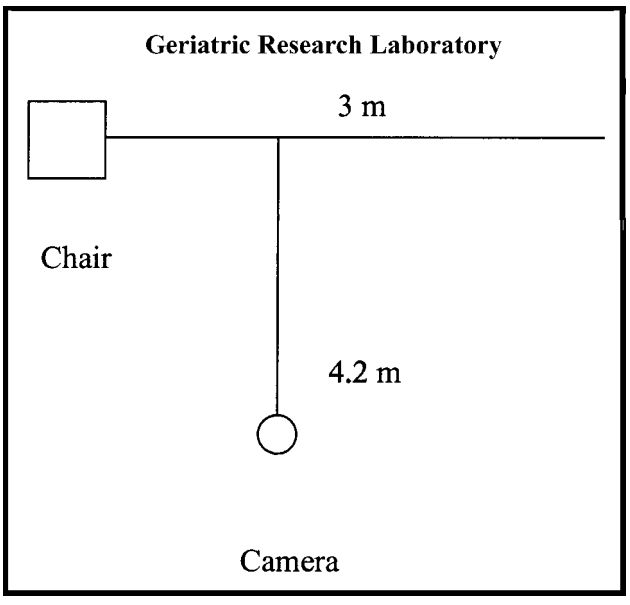

Figure.
Schematic representation of the videotaping setup.

whether the subjects had PD or not. Timing commenced on the word "start" (spoken by SM) and ceased once the subject had stopped moving, after sitting down again in the armchair. This procedure differed from that of Podsiadlo and Richardson,[12] who commenced timing on the word "go." We believe that the use of the word "go" could act as an external auditory cue in people with PD,[1] which could introduce error that would affect reliability.

Handheld stopwatches (Micronta Sports Timer*) were used by the 2 raters. The error that has been attributed to timing using a stopwatch for measurements of gait speed is 1.1%.[28] A firm chair with arms (seat height=50 cm, seat depth=47 cm, seat width=50 cm, armrest height=65 cm) was placed at the start of a 3-m walkway indicated by a length of electrician tape attached to a linoleum floor. Videotaping equipment was arranged on the right-hand side, 4.2 m from the marked walkway (Figure). A wide-angled lens (60° arc) was used so that the test could be videotaped without moving the videocamera. Subjects walked in their regular footwear. Shoes were not removed for testing because subjects in this age group infrequently walk without footwear.

Procedure

Testing took place in the morning in the Gait Laboratory at Kingston Centre, where all subjects were inpatients. Subjects with PD were timed on 5 trials of the TUG at least 12 hours after their last evening dose of PD medication, prior to receiving their first morning dose. Therefore, testing commenced at 6:00 AM if the last medication time was at 6:00 PM or at 8:00 AM if the last dose was at 8:00 PM. Subjects with PD were then timed on 5 trials of the TUG 1 hour after their first morning dose of levodopa was administered. The comparison subjects performed 5 trials of the TUG at a suitable time of convenience. Based on our experience, we believe that 5 trials provided a representative sample of performance while avoiding fatigue. The trials were separated by a 2-minute rest interval.

For the TUG, each subject was instructed to "stand up and walk at a comfortable speed (preferred speed) to the end of the tape, turn around, walk back to the chair, and sit down." Subjects were permitted one practice trial to familiarize them with the procedure before testing commenced. All trials were videotaped.

To investigate interrater reliability of the TUG scores, we examined the agreement between scores for 3 experienced raters and 3 inexperienced raters. The experienced raters were senior physical therapy clinicians (minimum experience=10 years), and the inexperienced raters were newly registered nurses. Each rater viewed the sequence of performances for the 12 subjects with PD and the 12 comparison subjects. The order of the videotape sequence was randomized in order to reduce the likelihood that raters would anticipate a particular score for a given individual. Raters viewed the videotape independently at least 1 week after testing. The raters were instructed to time each trial using a stopwatch. They were instructed to commence timing with the first forward movement of the head and trunk (due to low audibility of a small number of the videotapes) and to cease timing when the subject was stationary in the chair after completing the test. Despite this departure from the standard procedure of commencing timing on the word "start," there remained excellent agreement among all raters. Raters were permitted to replay trials if they were unable to record a time on the first viewing. None of the raters reported replaying the videotapes for any of the trials. Each videotape took approximately 1 hour to view.

Strategy for Statistical Analysis

For the subjects with PD, correlations between the mean TUG scores for trials 1 and 2, trials 2 and 3, trials 3 and 4, and trials 4 and 5 for the "off" and "on" phases of the levodopa cycle were examined using the Pearson product moment correlation coefficient (r). Prior literature on gait in people with PD has reported Pearson product moment correlation coefficients to summarize retest reliability[10,12]; thus, we used this statistic to compare our results with previous findings. The same procedure was used to examine correlations from one trial to the next for the comparison subjects. The consistency of the data across the 5 trials of the TUG for the subjects with PD

* Tandy Electronics, 91 Kurrajong Ave, Mt Druitt, New South Wales 2770, Australia.

Table 3.
Means and Standard Deviations for Trials 1 to 5 on the Timed "Up & Go" Test (in Seconds) for Subjects With Parkinson Disease (PD) and Comparison Subjects

	Subjects With PD (n=12)								
	"Off" Phase[a]			"On" Phase[b]			Comparison Subjects (n=12)		
Trial	$\bar{X}$	SD	Range	$\bar{X}$	SD	Range	$\bar{X}$	SD	Range
1	21.01	12.65	10.31–44.62	13.78	2.99	10.28–18.37	10.17	1.75	7.53–13.94
2	18.12	8.37	10.28–37.00	13.46	2.89	9.84–18.31	9.75	1.59	7.29–11.88
3	16.50	5.45	10.19–25.39	14.64	4.35	10.00–22.56	9.70	1.63	7.28–12.66
4	16.92	8.03	9.97–38.85	13.06	4.54	9.41–25.25	9.52	1.47	7.03–11.84
5	15.15	5.48	9.65–29.03	13.83	4.41	9.54–25.54	9.52	1.38	7.28–12.03

[a] "Off" phase=tested 12 hours after withdrawal of PD medication.
[b] "On" phase=1 hour after first morning dose of PD medication.

was then examined using a one-way analysis of variance (ANOVA) and a series of planned comparisons using the *t* statistic. The Bonferroni correction was used to reduce the chance of a type I error, which could occur when multiple *t* tests are performed.[29] The same procedure was used for the comparison subjects to provide reference data against which to interpret the PD results.

Intraclass correlation coefficients (3,1)[30] were used to examine the relationship between the scores recorded by the experienced and inexperienced raters. Intraclass correlation coefficients were calculated using the following formula:

$$ICC\,(3,1)=\frac{BMS-EMS}{BMS+(k-1)\,EMS}$$

where *BMS* refers to the between-subjects mean square, *EMS* is the residual mean, and *k* is the number of raters.

Analyses of variance were used to assess differences between experienced and inexperienced raters for mean TUG scores in the "off" and "on" phases for PD subjects as well as for the comparison subjects. Each of these analyses was confined to trials 2, 3, and 4 because there were differences in ratings between trials 1 and 2 as well as between trials 4 and 5. We viewed trial 1 as a practice trial, and we thought that the final trial was excessively fast due to practice effects.

In order to investigate aspects of sensitivity of the TUG in discriminating changes in performance, *t* tests were used to examine the differences between the means for trials 2, 3, and 4 in the "off" and "on" phases. In a similar way, the means for trials 2, 3, and 4 at peak dose for the subjects with PD were compared with the means of trials 2, 3, and 4 for the comparison subjects.

Finally, the mean "off" phase TUG scores were correlated with the Modified Webster Scale scores obtained in the "off" phase, and the mean "on" phase TUG scores were correlated with the Modified Webster Scale scores obtained in the "on" phase. We used SPSS for Windows 6.0 (1992) statistical software[†] for data analyses.

Table 4.
Pearson Product Moment Correlations Between Trials for Subjects With Parkinson Disease (PD) ("Off" and "On" Phases of the Medication Cycle) and Comparison Subjects[a]

	Subjects With PD (n=12)		Comparison Subjects (n=12)
	End of Dose	Peak Dose	
Trial 1–trial 2	.96	.96	.90
Trial 2–trial 3	.90	.73	.96
Trial 3–trial 4	.80	.82	.96
Trial 4–trial 5	.98	.99	.97

[a] All correlations are statistically significant at $P<0.05$, $df=11$.

Results

Change Across Trials and Retest Reliability

For the subjects with PD in the "off" phase of the medication cycle, there was no difference in performance across the middle 3 trials (trials 2–4) of the TUG ($F=0.6$; $df=2,22$; $P=.555$, power=0.139) (Tab. 3). The mean TUG value for trials 2 through 4 was 17.18 (SD=7.3, range=9.97–28.88). In contrast, the mean value for trial 1 was relatively high (approximately 21), and the mean value for trial 5 was relatively low (approximately 15). A repeated-measures ANOVA for "off" phase performance in subjects with PD showed a change in TUG times across the 5 trials ($F=3.03$; $df=4,44$; $P=.027$, partial $\lambda^2=0.216$). The comparisons we planned showed differences in means between trials 1 and 2 ($t[11]=1.9$, $P=.0415$, $\alpha=.025$) and trials 4 and 5 ($t[11]=2.13$, $P=.0285$, $\alpha=.025$). There were no differences between trials 2 and 3 and trials 3 and 4. Removal of the first and last trials enhanced the reliability of the

[†] SPSS Inc, 444 N Michigan Ave, Chicago, IL 60640.

Table 5.
Intraclass Correlation Coefficients for Experienced and Inexperienced Raters for the Timed "Up & Go" Test During the "Off" and "On" Phases of the Medication Cycle in Subjects With Parkinson Disease

Phase of Medication Cycle	Experienced Raters (n=3)	Inexperienced Raters (n=3)
"Off"	.99	.87
"On"	.99	.99

measurements because trial 1 was abnormally slow and trial 5 was faster than typical, possibly due to the effects of practice.

For subjects with PD tested in the "on" phase, there was a greater degree of consistency in the results compared with the "off" phase. The mean TUG time for the PD group in the "on" phase was 13.72 seconds (SD=3.9, range=9.41–25.54). The ANOVA showed no change in timed scores from trial 1 to trial 5 (F=0.67; *df*=4,44; *P*=.613, power=0.202) or from trial 2 to trial 4 (F=1.10; *df*=2,22; *P*=.350, power=0.218). Moreover, our planned comparisons showed no differences in performance on the TUG between trials 1 and 2 (t[11]=1.37, *P*=.98) or between trials 4 and 5 (t[11]=1.34, *P*=.208, α=.025).

Although there was a change in TUG times across the 5 trials in the comparison subjects (F=3.86; *df*=4,44; *P*=.009, partial λ^2=0.26), this was due to differences between trials 1 and 2 (t[11]=1.94, *P*=.039) (Tab. 3). There was no change in performance across trials 2 through 5 (F=1.69; *df*=3,33; *P*=.188). Thus, removal of the first trial enhanced the reliability of the measurements.

Table 4 summarizes correlations between trials for the subjects with PD in the "off" and "on" phases, showing strong, positive linear relationships between consecutive TUG trials. Pearson correlation coefficients (r) ranged from .73 to .99. Similarly, for the comparison subjects, there were strong, positive linear relationships between consecutive TUG trials (r=.90–.97).

Interrater Reliability

Intraclass correlation coefficients were calculated to investigate the agreement between experienced and inexperienced raters in both phases of the levodopa cycle. As shown in Table 5, all of the ICCs were ≥.87, indicating a high degree of agreement for this test across different conditions.

Using the mean TUG scores for subjects with PD at the end of dose for trials 2 through 4, the ICCs showed no difference among the 3 experienced raters (ICC [3,1]=.999). Similarly, there was a very high degree of agreement among the 3 experienced raters for TUG scores for trials 2 through 4 at peak dose in the levodopa cycle (ICC [3,1]=.998). Consistent with these findings, the inexperienced raters showed close agreement for the mean scores of trials 2 through 4 in the "off" phase (ICC [3,1]=.87) and the "on" phase (ICC [3,1]=.999). Overall, all ICCs were .99, except for an ICC of .87 that was calculated for the inexperienced raters timing the TUG during the "off" phase of the medication cycle.

Table 6.
Grand Means and Standard Deviations of All Trials of the Timed "Up & Go" Test (in Seconds) Recorded by Experienced and Inexperienced Raters During the "Off" and "On" Phases of the Medication Cycle for the Subjects With Parkinson Disease

	"Off" Phase		"On" Phase	
Raters[a]	$\bar{X}$	SD	$\bar{X}$	SD
1E	16.71	1.94	12.81	1.14
2E	16.45	1.90	12.60	1.08
3E	16.36	2.50	12.67	1.14
4I	16.52	1.92	12.68	1.07
5I	16.40	1.90	12.68	1.14
6I	16.76	2.15	13.10	1.21

[a] E=experienced rater, I=inexperienced rater.

To further investigate agreement between experienced and inexperienced raters, the mean scores of the TUG recorded by the experienced raters was correlated with the mean scores recorded by the inexperienced raters in the "off" and "on" phases of the medication cycle. There was a strong, positive linear relationship between the experienced and inexperienced raters for both of these conditions (ICC [3,1]=.998 in both cases). The grand means and standard deviations for all trials of the TUG also showed a high degree of agreement between raters for both "off" and "on" phases (Tab. 6). There were moderately strong, positive linear correlations between the Modified Webster Scale scores and the TUG scores in the "off" phase (r=.62) and in the "on" phase (r=.52).

Sensitivity of the TUG for Detecting Change

Planned comparisons using independent-samples t tests were used to investigate whether the mean TUG scores for trials 2 through 4 in the "off" and "on" phases for the subjects with PD were different. These comparisons showed a difference across the stages of the medication cycle (t[11]=2.4, *P*=.035). The correlation between "off" and "on" phase scores showed a moderately strong, positive linear relationship (r=.74, n=12, *P*=.003). Fifty-five percent of the variation in the "on" phase scores could be explained by variation in the "off" phase scores.

The TUG scores also showed differences between the means of trials 2 through 4 in the subjects with PD at the end of dose ($\bar{X}$=17.18) and the means of trials 2 through 4 in the comparison subjects ($\bar{X}$=9.66)

(t[12.13]=3.78, P=.003). In addition, the TUG was able to discriminate between the means of trials 2 though 4 in the subjects with PD in the "on" phase and the means of trials 2 through 4 in the comparison subjects (t[14.92]=3.79, P=.002).

Discussion

This is the first study to investigate the retest reliability, interrater reliability, and sensitivity of the measurements obtained with the TUG in people with idiopathic PD. Performance times were very stable within each session, provided that subjects were allowed one "practice" trial prior to measurement. Performance was more variable after levodopa had been withdrawn for 12 hours and was greater in the subjects with PD than in the comparison subjects.

The finding that TUG times were most consistent at peak dose of the levodopa cycle supports our previous findings on the repeatability of the temporal and spatial measures of gait in subjects with PD.[26] Using footswitch stride analysis to measure a series of 10-m samples of walking, we showed that people with mild to moderately severe PD had highly consistent short-stepped, slow gait patterns when they were tested 1 to 3 hours after their morning dose of levodopa. However, when they were tested 30 minutes prior to the next dose, we found considerable variability and poor correlations between repeat measures of performance. Thus, 2 patterns of mobility are evident in people with PD. One pattern is consistent and repeatable and occurs at peak dose in the medication cycle, and the other pattern is variable and inconsistent and corresponds to the end-of-dose phase of the medication cycle.[1,26]

When performing neurological assessments for patients with PD, we believe it is important for the physical therapist to sample performance in both the "off" and "on" phases, so that the full spectrum of mobility disorders can be documented. Physical therapy interventions may be more effective if they are tailored to changing mobility status.[1] For example, when a person is in the "off" phase, he or she can be instructed to use one set of movement strategies to enhance motor performance. We believe this instruction could include the use of external cues, mental rehearsal, visualization, and deliberately focusing attention on the key components of the movement to be performed. When a person is in the "on" phase, he or she may require fewer or different strategies. According to Morris,[1] some patients do not require any physical therapy in that locus of the medication cycle. Optimal physical therapy ensures that neurological assessments and treatments are closely linked, so that changes in rehabilitation and medication can be closely charted using clinical measures that have demonstrated reliability and validity.[3] The TUG appears ideally suited to this task because it has strong interrater and retest reliability and is able to clearly differentiate different phases of the levodopa cycle.

To enhance the retest reliability of measurements in this study, 5 TUG trials were performed and the means of trials 2 through 4 were used for analysis. In both the subjects with PD and the comparison subjects, the times on the first trial were generally increased in relation to the mean. The high retest reliability of TUG times we found is comparable to the results of other studies. For example, Podsiadlo and Richardson[12] reported retest reliability to be very high (ICC=.99) although they did not document when testing occurred in relation to the medication cycle. Similarly, Thompson and Medley[18] found retest reliability (r) for the TUG in elderly people without health problems to range between .81 and .99.

The interrater reliability of the measurements was excellent for both experienced and inexperienced raters, with ICCs (3,1) ranging from .87 to .99. Again, this finding is in agreement with the findings of Podsiadlo and Richardson,[12] who noted that reliability (ICC) in a mixed group of raters was .99 for the TUG.

We demonstrated that the TUG scores could be used to clearly differentiate between the performance of subjects with PD and that of comparison subjects, even when the subjects with PD were in the "on" phase of the medication cycle. When in the "on" phase, subjects with PD were still considerably slower than their age-matched counterparts. The TUG scores could also be used to detect differences between the scores obtained from the subjects when they were medicated and the scores obtained when they were not medicated. In the nonmedicated state, performance was slower. In the only other study that investigated mobility across the medication cycle,[31] a variation of the "Get-up and Go" Test did not reflect the treatment effects of levodopa. However, the test protocol used by Bowes et al[31] relied on manual calculation of the step length, mean double support time, and speed as recorded by a gait assessment trolley, which was a small measurement cart attached to the patient by fine leads. The encumbrance produced by the leads might have increased measurement error. These gait variables could be used to discriminate between the medicated and nonmedicated states, but their reliability was poor (r=.24–.53). Furthermore, Bowes and colleagues recorded at 2, 4, and 6 hours after levodopa medication was given, which falls outside the "on" period used in our investigation (60–90 minutes) and, therefore, makes direct comparison of results difficult.

There are a number of limitations of this study that need to be acknowledged. There was a small sample, due to limited access to patients, which limits the generalizability of findings. We, therefore, consider this a pilot study. Subjects also were recruited from a sample of conve-

nience from Kingston Centre in Australia and may not be representative of all people with PD or elderly people without health problems in general. Future investigations would benefit from random sampling wherever possible. The subjects with PD in this study were classified as having mild to moderate PD according to the Modified Webster Scale. The sample did not include people who were more severely affected by PD, people in the earliest stages of the disease, or people with dystonia or dyskinesia. The presence of these movement disorders would be expected to increase variability of performance, thus reducing the reliability of repeat measurements. Despite these limitations, we believe that our study provides evidence establishing the validity of TUG scores in people with idiopathic PD, warranting replication by clinicians in other countries using larger samples of patients.

Conclusion

Physical therapists are key members of the health team working with people who have PD. We believe that physical therapists will have an increasing need for clinical measures that have reliability and validity and that are quick and easy to administer in both institutional and community settings. We argue that the TUG is useful for the measurement of mobility in people with mild to moderate PD because it fulfills these requirements. The TUG may be particularly well suited for the quantification of disorders resulting in poor sequencing of well-learned motor skills, which is a problem in people with PD.

References

1 Morris ME. Movement disorders in people with Parkinson disease: a model for physical therapy. *Phys Ther.* 2000;80:578–597.

2 Morris ME, Churchyard A, Iansek R. How to conduct a dose response trial of Parkinson's disease medication. *Australian Journal of Physiotherapy.* 1998;44:131–133.

3 Schenkman M, Cutson T, Kuchibhatla M, et al. Reliability of impairment and physical performance measures for persons with Parkinson's disease. *Phys Ther.* 1997;77:19–27.

4 Smithson F, Morris ME, Iansek R. Performance on clinical tests of balance in Parkinson's disease. *Phys Ther.* 1998;78:577–592.

5 Behrman AL, Teitelbaum P, Cauraugh JH. Verbal instructional sets to normalise temporal and spatial gait variables in Parkinson's disease. *J Neurol Neurosurg Psychiatry.* 1998;65:580–582.

6 Banks M. Physiotherapy benefits patients with Parkinson's disease. *Clin Rehabil.* 1989;3:11–16.

7 Burleigh-Jacobs A, Horak FB, Nutt JG, Obeso JA. Step initiation in Parkinson's disease: influence of levodopa and external sensory triggers. *Mov Disord.* 1997;2:206–215.

8 Benecke R, Rothwell JC, Dick JPR, et al. Performance of simultaneous movements in patients with Parkinson's disease. *Brain.* 1986;109: 739–757.

9 Benecke R, Rothwell JC, Dick JPR, et al. Disturbance of sequential movements in patients with Parkinson's disease. *Brain.* 1987;110: 361–379.

10 Mathias S, Nayak U, Isaacs B. Balance in elderly patients: the "Get-up and Go" Test. *Arch Phys Med Rehabil.* 1986;67:387–389.

11 Morris ME, Iansek R. *Parkinson's Disease: A Team Approach.* Melbourne, Victoria, Australia: Southern Health Care Network; 1997.

12 Podsiadlo D, Richardson S. The timed "Up and Go" test. *Arch Phys Med Rehabil.* 1989;67:387–389.

13 Martinez-Martin P, Urra D, del Ser Quijano T, et al. A new clinical tool for gait evaluation in Parkinson's disease. *Clin Neuropharmacol.* 1997;20:183–194.

14 Berg KO, Maki BE, Williams KI, et al. Clinical and laboratory measures of postural balance in an elderly population. *Arch Phys Med Rehabil.* 1992;73:1073–1080.

15 Salgado R, Lord S, Packer J, Ehrlich F. Factors associated with falling in elderly hospital patients. *Gerontology.* 1994;40:325–331.

16 Fleming K, Evans J, Weber D, Chutka D. Practical functional assessment of elderly persons: a primary-care approach. *Mayo Clin Proc.* 1995;70:890–910.

17 McMurdo M, Johnstone R. A randomized controlled trial of a home exercise programme for elderly people with poor mobility. *Age Ageing.* 1995;24:425–428.

18 Thompson M, Medley A. Performance of community dwelling elderly on the Timed Up and Go test. *Physical and Occupational Therapy in Geriatrics.* 1995;13:17–29.

19 MacRae P, Asplund L, Achnelle J, et al. A walking program for nursing home residents: effects on walk endurance, physical activity, mobility, and quality of life. *J Am Geriatr Soc.* 1996;44:175–180.

20 Skelton D, McLaughlin A. training functional ability in old age. *Physiotherapy.* 1996;82:159–167.

21 Quinn N, Critchley P, Marsden CD. Young-onset Parkinson's disease. *Mov Disord.* 1987;2:73–91.

22 Kostic V, Przedborski S, Flaster E, Sternic N. Early development of levodopa-induced dyskinesias and response fluctuations in young-onset Parkinson's disease. *Neurology.* 1991;41:202–205.

23 DuChatinier K, Molen N, Rozendal R. Step length, step frequency and temporal factors of the stride in normal humans. *K Ned Akad van Weten.* 1970;73:214–227.

24 Kokmen E, Naessens J, Offord K. A Short Test of Mental Status: description and preliminary results. *Mayo Clin Proc.* 1987;62:281–288.

25 Fahn S, Marsden CD, Calne D, Goldstein M. *Recent Developments in Parkinson's Disease.* Florham Park, NJ: MacMillan Healthcare Information; 1987.

26 Morris ME, Matyas TA, Iansek R, Summers JJ. Temporal stability of gait in Parkinson's disease. *Phys Ther.* 1996;76:763–789.

27 Kempster P, Frankel J, Bovingdon M, et al. Levodopa peripheral pharmacokinetics and duration of motor response in Parkinson's disease. *J Neurol Neurosurg Psychiatry.* 1989;52:718–723.

28 Stevens J. *Effects of Instructions and Surfaces on the Gait Speed of Stroke Patients* [master's preliminary]. Melbourne, Victoria, Australia: La Trobe University; 1993.

29 Portney LG, Watkins MP. *Foundations of Clinical Research Applications to Practice.* East Norwalk, Conn: Appleton & Lange; 1993.

30 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.

31 Bowes S, Clark P, Charlett A, et al. Objective outcome criteria in trials of anti-Parkinson therapy in the elderly: sensitivity, specificity, and reliability of measures of brady- and hypokinesia. *Br J Clin Pharmacol.* 1991;31:295–304.