

S.-L. Huang, MS, is Lecturer, School of Occupational Therapy, College of Medicine, National Taiwan University, and Senior Occupational Therapist, Division of Occupational Therapy, Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei, Taiwan.

C.-L. Hsieh, PhD, is Professor, School of Occupational Therapy, College of Medicine, National Taiwan University, and Senior Occupational Therapist, Division of Occupational Therapy, Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital.

R.-M. Wu, PhD, MD, is affiliated with the Department of Neurology, National Taiwan University Hospital, and the College of Medicine, National Taiwan University.

C.-H. Tai, MD, is affiliated with the Department of Neurology, National Taiwan University Hospital.

C.-H. Lin, MD, is affiliated with the Department of Neurology, National Taiwan University Hospital Yun-Lin Branch, Yun-Lin, Taiwan.

W.-S. Lu, MS, is Assistant Professor, School of Occupational Therapy, Chung Shan Medical University, and Senior Occupational Therapist, Occupational Therapy Room, Chung Shan Medical University Hospital, Taichung, Taiwan. Mailing address: No. 110, Sec. 1, Jianguo N. Rd, Taichung City 40201, Taiwan (Republic of China). Address all correspondence to Mr Lu at: adam222@ms39.hinet.net.

[Huang S-L, Hsieh C-L, Wu R-M, et al. Minimal detectable change of the Timed "Up & Go" Test and the Dynamic Gait Index in people with Parkinson disease. *Phys Ther*. 2011;91:114-121.]

© 2011 American Physical Therapy Association

Minimal Detectable Change of the Timed "Up & Go" Test and the Dynamic Gait Index in People With Parkinson Disease

Sheau-Ling Huang, Ching-Lin Hsieh, Ruey-Meei Wu, Chun-Hwei Tai, Chin-Hsien Lin, Wen-Shian Lu

Background. The minimal detectable change (MDC) is the smallest amount of difference in individual scores that represents true change (beyond random measurement error). The MDCs of the Timed "Up & Go" Test (TUG) and the Dynamic Gait Index (DGI) in people with Parkinson disease (PD) are largely unknown, limiting the interpretability of the change scores of both measures.

Objective. The purpose of this study was to estimate the MDCs of the TUG and the DGI in people with PD.

Design. This investigation was a prospective cohort study.

Methods. Seventy-two participants were recruited from special clinics for movement disorders at a university hospital. Their mean age was 67.5 years, and 61% were men. All participants completed the TUG and the DGI assessments twice, about 14 days apart. The MDC was calculated from the standard error of measurement. The percentage MDC (MDC%) was calculated as the MDC divided by the mean of all scores for the sample. Furthermore, the intraclass correlation coefficient was used to examine the reproducibility between testing sessions (test-retest reliability).

Results. The respective MDC and MDC% of the TUG were 3.5 seconds and 29.8, and those of the DGI were 2.9 points and 13.3. The test-retest reliability values for the TUG and the DGI were high; the intraclass correlation coefficients were .80 and .84, respectively.

Limitations. The study sample was a convenience sample, and the participants had mild to moderately severe PD.

Conclusions. The results showed that the TUG and the DGI have generally acceptable random measurement error and test-retest reliability. These findings should help clinicians and researchers determine whether a change in an individual patient with PD is a true change.



Post a Rapid Response to this article at:
ptjournal.apta.org

Gait and balance deficits are common in patients with Parkinson disease (PD).¹⁻³ Patients with PD are characterized clinically by movement-related symptoms, such as tremor, rigidity, slow movement (bradykinesia), and postural instability. Consequently, they experience difficulties in gait and gait-related activities (eg, turning and climbing stairs) that reduce their quality of life.³⁻⁵ To manage and monitor gait and balance deficits, clinicians need to routinely measure these characteristics of patients with PD. Furthermore, to interpret the results of the measurements, clinicians must determine whether change scores in gait and balance deficits represent true changes or are a result of measurement error.

Any measurement entails random measurement error. A measure without determined measurement error has limited score interpretability.^{6,7} The minimal detectable change (MDC) is the minimal amount of change between 2 points in time that indicates a true statistical change.^{6,8} The MDC ideally is different from the minimal important change (MID), which is the minimal “important or meaningful” change after an intervention from the viewpoint of a patient.⁹ The MDC is sometimes calculated to enhance interpretability. Difference scores that are smaller than the MDC can be attributed to random error at a certain confidence level (usually 95%).¹⁰ Clinicians and researchers can use the MDC as a threshold to determine whether a change score in a measure for an individual patient represents a true change or is within the bounds of random error.⁶ Thus, the MDC of a measure is crucial for the interpretation of data in both research and clinical settings.^{6,7}

The Timed “Up & Go” Test (TUG) measures functional mobility and gait speed,¹¹ and the Dynamic Gait Index (DGI), including 8 task-oriented items,

measures gait quality.¹² The TUG and the DGI provide a complementary and more comprehensive understanding of the characteristics of patients’ gait and balance control. Because of the relevance of the TUG and the DGI to the motor characteristics of patients with PD and their easy administration, both measures are being used increasingly to examine gait and balance deficits in patients with PD.¹³⁻¹⁶ The MDC of the TUG in patients with PD has been reported, but the results vary extensively (2–11 seconds).^{14,15} To our knowledge, however, the MDC of the DGI in patients with PD has not been examined. These shortcomings limit the interpretation of the change scores for both measures. Thus, the purpose of this study was to determine the MDCs of the TUG and the DGI in patients with PD.

Method

Participants

All of the participants were recruited from special clinics for movement disorders at the Department of Neurology, National Taiwan University Hospital, from February to October 2008. To minimize selection bias and the effect of cognitive impairment on the TUG and the DGI, the following criteria were used to determine whether patients could be included in this study: PD diagnosed by a movement disorder specialist, with a subsequent referral to occupational therapy; Hoehn-Yahr stages I to III¹⁷; a Mini-Mental State Examination score of greater than 20; and agreement to participate in this study and to sign consent forms as approved by the Medical Ethics Committee at the hospital. The exclusion criteria were: (1) a TUG score of more than 20 seconds in the first session because of safety concerns and (2) other diseases or injuries (eg, stroke, lower-extremity amputation) likely to affect balance function.

Eighty-five participants with PD were invited to participate in the

study. Seven participants were excluded because they had TUG scores of more than 20 seconds or had had a stroke and lost balance easily (n=2). Of the remaining 78 participants, 6 participants were lost to follow-up because of loss of contact or refusal to retest. The mean baseline scores of the TUG and the DGI for the remaining 72 participants and the 6 participants lost to follow-up were not statistically different ($P=.30$ for the TUG and $P=.24$ for the DGI).

The demographic characteristics and major comorbidity data (eg, hypertension, cardiovascular disease, diabetes mellitus, hyperlipidemia, hyperuricemia, cataracts) for the 72 participants were collected from medical records (Tab. 1). The participants’ mean age was 67.5 years, and 61% of the participants were men. The mean baseline scores of the TUG and the DGI were 11.8 seconds and 21.6 points, respectively.

Procedure

The participants were screened and invited to participate by the movement disorder specialists. When the participants agreed to join the research, they were scheduled for our assessments. All participants were assessed in person during the “on” status (about 1 hour after taking anti-PD medication, including levodopa [Sinemet* and Madopar[†]] or dopamine agonists [ergot and

* Merck & Co Inc, PO Box 4 WP39-206, West Point, PA 19486-0004.

[†] Roche Products Ltd, Hexagon Place, 6 Falcon Way, Welwyn Garden City, Hertfordshire, United Kingdom AL7 1TW.



Available With
This Article at
ptjournal.apta.org

• [Audio Abstracts Podcast](#)

This article was published ahead of print on October 14, 2010, at ptjournal.apta.org.

Table 1.
Demographic and Clinical Characteristics of the Participants^a

Characteristic	Participants Completing 2 Assessments (N=72)	Participants Lost to Follow-up (n=6)
Age (y), \bar{X} (SD)	67.5 (11.6)	64.4 (5.6)
Sex, men/women (n)	44/28	4/2
Hoehn-Yahr stage (no. of participants)		
I	17	3
II	33	2
III	22	1
Parkinson disease evolution time (range)	2 mo–15 y	6 mo–10 y
TUG baseline score (s), \bar{X} (SD)	11.8 (2.9)	9.9 (1.7)
DGI baseline score (points), \bar{X} (SD)	21.6 (2.6)	23.3 (1.2)
Days between 2 assessments, \bar{X} (SD)	13.5 (2.5)	
Comorbidity (no. of participants)		
Hypertension	23	0
Cardiovascular disease	8	1
Diabetes mellitus	8	0
Hyperlipidemia	6	0
Cataract	5	0
Benign prostatic hyperplasia	4	0
Cancer	2	0
Glaucoma	2	0
Osteoporosis	2	0
Depression	2	0
Hyperuricemia	1	0

^a TUG=Timed “Up & Go” Test, DGI=Dynamic Gait Index.

nonergot agonists]) by a single occupational therapist at the same place in 2 sessions about 2 weeks apart. The rater was familiar with the method and the sequence for evaluating TUG and DGI scores in this study. The participants used their regular walking aids (eg, canes, walkers) during the assessments in both test sessions. Changes in anti-PD medication were not allowed.

In the first session, half of the participants were administered the DGI before the TUG; the order was reversed for the other half to control for possible bias of the testing sequence. For safety, the participants were well instructed and allowed 1 practice trial before the formal TUG.

All of the participants performed the formal TUG once. Before the second session, the same rater confirmed that each participant had experienced no significant change (eg, medication, injury, disease progression) within the preceding 2 weeks.

Measure

The TUG is a mobility test that is used to measure the basic mobility skills of people who are elderly or have neurological conditions.^{11,14} It includes a sit-to-stand component as well as walking 3 m, turning, and returning to the chair. People perform these tasks using regular footwear and customary walking aids. The measured outcome is the time

in seconds to complete the entire sequence.

The DGI is a performance-based mobility test that is used to examine an individual’s ability to modify gait in response to task demands.¹² It consists of 8 common gait tasks: walking on a level surface, changing gait speed, walking with vertical and horizontal head turns, pivot turning, stepping over an obstacle, and ascending and descending stairs. These 8 items are scored on a 4-point ordinal scale, in which 0 represents severe impairment and 3 represents normal movement. The range is 0 to 24 points, with high scores indicating the ability to move normally.

Data Analysis

Data were analyzed with the SPSS 15.0 for Windows statistical program.[‡] To investigate the MDC, we calculated the intraclass correlation coefficient (ICC) first. The ICC (2,1), applied in this study, is commonly used to examine the extent of reproducibility between repeated measurements¹⁸ and to calculate the standard error of measurement (SEM), which, in turn, is used to calculate the MDC.^{6,10} The ICC (2,1) was computed with a random-effects 2-way analysis of variance for 2 test sessions, as follows:

$$ICC(2,1) = (BMS - EMS) / [BMS + EMS + 2(JMS - EMS) / n]$$

In this formula, BMS is the between-participants mean square (variability between participants), EMS is the error mean square (residual mean square), JMS is the observations mean square (variability between test sessions), and n is the number of participants. An ICC of greater than .80 indicates high reliability.¹⁹

[‡] SPSS Inc, 233 S Wacker Dr, Chicago, IL 60606.

The MDC (at a confidence level of 95%) was calculated from the SEM with the following formulas^{6,8}:

$$SEM = SD_{\text{baseline}} \times \sqrt{(1 - r_{\text{test-retest}})}$$

and

$$MDC = z \text{ score}_{\text{level of confidence}} \times \sqrt{2} \times SEM$$

In these formulas, SD_{baseline} is the standard deviation of the baseline, $r_{\text{test-retest}}$ is the coefficient of the test-retest reliability estimated from ICC (2,1) in this study, $z \text{ score}_{\text{level of confidence}}$ is the confidence interval (CI) from a normal distribution, and multiplying by the square root of 2 accounts for the extra uncertainty that arises when scores from measurements at 2 time points are used.

In addition, the MDC can be expressed as a percentage (MDC%), which is independent of the units of measurement. The MDC% can be used to determine a relatively true change after a treatment or between repeated measurements over time.²⁰ The MDC% also represents the relative amount of random measurement error. The MDC% is the MDC divided by the mean of all scores for the sample. An MDC% of less than 30 is considered acceptable, and an MDC% of less than 10 is considered excellent.²¹

Moreover, the reproducibility between 2 repeated measurements can be visualized by use of Bland-Altman plots with 95% limits of agreement.²² In the plots, the differences (d) between each pair of measurements are plotted against the mean of each pair of measurements. If the differences follow a standard normal distribution, then 95% of the differences will lie between $\bar{d} \pm 1.96SD_{\text{difference}}$ (ie, limits of agreement), where \bar{d} is

Table 2.

Test-Retest Reliability Values for the Timed “Up & Go” Test (TUG) and the Dynamic Gait Index (DGI) (N=72)^a

Test	First Session	Second Session	Difference	ICC (95% CI)	MDC (MDC%)
TUG	11.8 (2.9)	11.8 (3.4)	0.0 (2.0)	.80 (.70–.87)	3.5 (29.8)
DGI	21.6 (2.6)	21.5 (2.7)	–0.1 (1.5)	.84 (.76–.90)	2.9 (13.3)

^a Values are reported as means (standard deviations) unless otherwise indicated. ICC=intraclass correlation coefficient, CI=confidence interval, MDC= $z \text{ score}_{\text{level of confidence}} \times \sqrt{2} \times \text{standard error of measurement}$, MDC%=(MDC/mean of measurements) \times 100%.

the mean difference between the 2 test sessions and $SD_{\text{difference}}$ is the standard deviation of the differences.

These plots also can be used to illustrate heteroscedasticity, which represents a tendency: the differences between repeated measurements generally increase as the mean values of the measurements increase.²³ The possibility of heteroscedasticity can be examined on the basis of the association (ie, Pearson r) between the mean and the absolute difference of each pair of measurements. If r is greater than .3, then the data are heteroscedastic.²⁴

Systematic bias can be calculated from the 95% CI of \bar{d} on the basis of the standard error (SE) of \bar{d} with the following formulas:²³

$$95\% \text{ CI of } \bar{d} = \bar{d} \pm 1.96 \times SE$$

and

$$SE = SD_{\text{difference}}/\sqrt{n}$$

If 0 is included within the 95% CI, then it can be inferred that there is no significant systematic bias between measurements.

Results

The test-retest reliability values for the TUG and the DGI are shown in Table 2. The ICCs for the TUG and the DGI were .80 and .84, respectively, indicating that both measures had high test-retest reliability. The mean differences for the TUG and

the DGI were 0.0 and –0.1, respectively. The MDC of the TUG was 3.5 seconds, and the MDC% was 29.8, representing acceptable measurement error. The MDC of the DGI was 2.9 points, and the MDC% was 13.3, representing limited measurement error. However, 2 participants had TUG measurements of more than 20 seconds in the retest assessment (ie, 2 outliers). When we excluded data from both of those participants, the results were very similar to those reported above. The ICCs for the TUG and the DGI were .78 and .83, respectively. The MDC for the TUG was 3.4 seconds, and the MDC% was 29.6. The MDC of the DGI was 2.9 points, and the MDC% was 13.2.

The differences in scores are plotted against the mean scores of the 2 measurements for both the TUG and the DGI in Figure 1.²² The limits of agreement ranged from 3.9 to –3.9 seconds for the TUG and from 2.9 to –3.0 points for the DGI. The 95% CI of the mean difference for the TUG ranged from –0.4 to 0.5, and that for the DGI ranged from –0.4 to 0.3. Zero was included in the 95% CIs of the mean differences for both the TUG and the DGI, indicating that there was no significant systematic bias between the successive measurements.

In addition, the Pearson r values for the association between the mean and the absolute difference for the TUG and the DGI were .54 and –.41,

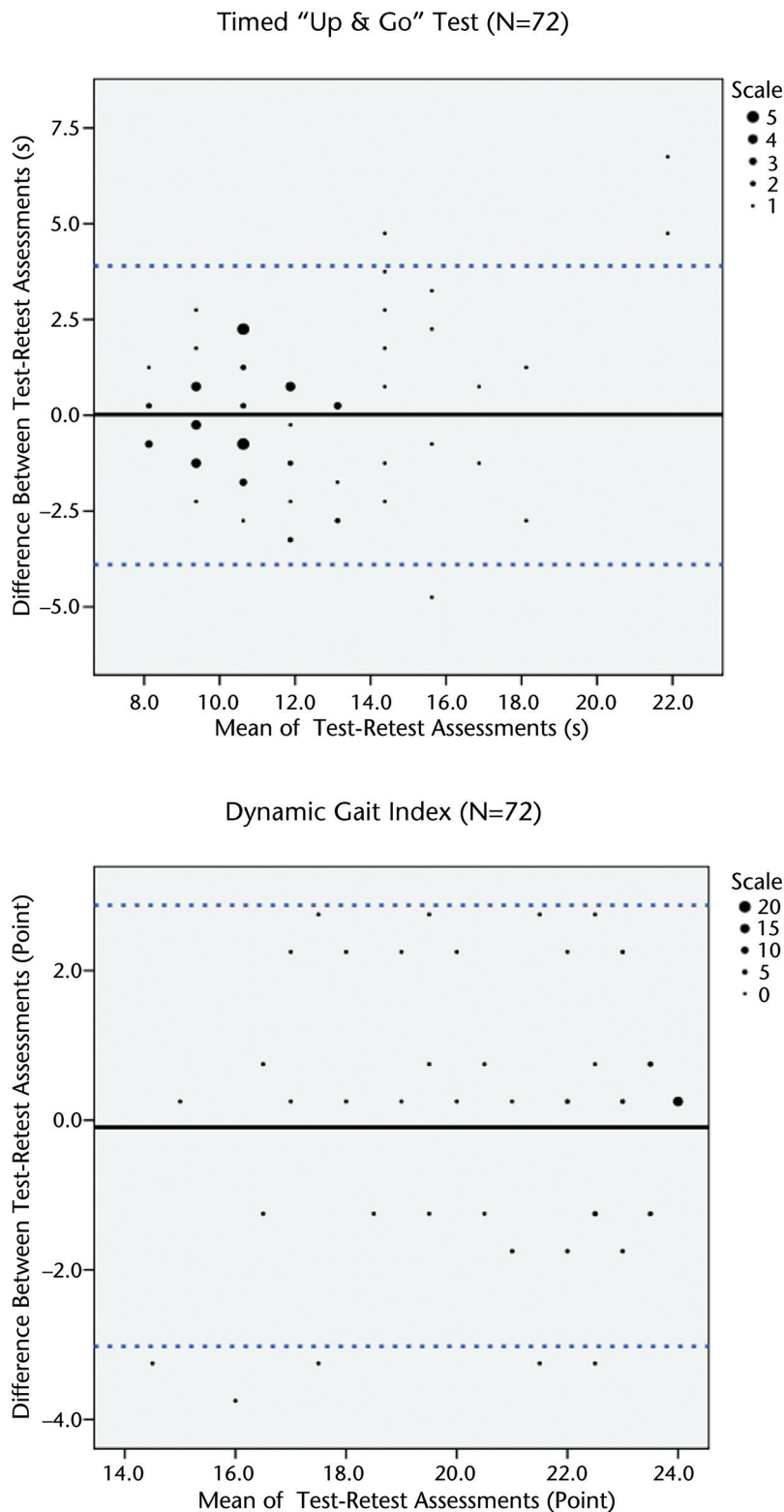


Figure 1. Bland-Altman method for plotting the difference scores against the mean scores of each pairing of the Timed "Up & Go" Test (N=72) and the Dynamic Gait Index (N=72). The solid line represents the mean of the differences. The 2 dashed lines define the limits of agreement (mean of the difference $\pm 1.96 \times SD$).

respectively. When we excluded data from the 2 participants whose TUG measurements in the retest assessment were more than 20 seconds (one of our exclusion criteria was a TUG score of more than 20 seconds in the first session, for safety), the Pearson r values for the TUG and the DGI were .34 and $-.43$, respectively.

The negative association ($-.41$) found for the DGI indicated that a higher score was correlated with less of a difference, a conclusion that appeared unreasonable. Thus, we inspected the raw scores of the DGI and found that 31 participants had maximum scores (24 points) in either or both of the 2 successive measurements. To remove the ceiling effect, we excluded data from these 31 participants and recalculated the reliability values for the DGI. We found that the negative association was no longer obvious (Pearson $r = -.28$) (Fig. 2) and that the ICC was .78, the mean difference was -0.2 , the MDC was 3.3 points, and the MDC% was 16.6.

Discussion

To examine measurement error, we used the MDC and the MDC%. The MDC represents the measurement error as an absolute value, whereas the MDC% is independent of the units of measurement and can be used to compare the amount of random error between measurements. In addition, the MDC can be viewed as the threshold of statistically significant change for an individual patient in a clinical setting.⁷ That is, if the magnitude of a change between successive measurements for an individual patient is more than the MDC of the measure, it can be concluded that the patient has made significant progress in the specific characteristic assessed by the measure.

The results of recent studies investigating the MDC of the TUG for pa-

tients with PD revealed wide variations. One study of 26 participants with PD, tested over a period of 7 days, revealed an MDC of 2 seconds.¹⁴ Another study of 37 community-dwelling adults with PD, tested over a period of 7 days, revealed an MDC of 11 seconds.¹⁵ These variations may have resulted from the small to moderate sample sizes in these studies. Our finding of an MDC of 3.5 seconds—a value lying between the values reported in the 2 earlier studies^{14,15}—seems more reliable because of the larger size of our sample^{20,25} and the even distribution among Hoehn-Yahr stages I to III in our sample. In addition, no MDC% values from the earlier studies were available for comparison. However, further studies with larger sample sizes or modified inclusion and exclusion criteria (eg, including patients with a first TUG score of more than 20 seconds) may be needed to validate our results.

The MDC% of the TUG was slightly less than 30, representing acceptable random measurement error. In the present study, however, to prevent fatigue, we measured the participants' performances on the TUG only once per session. More trials per session would increase the stability of the measurements and reduce the MDC and the MDC%.²⁵ Thus, the MDC of 3.5 seconds can be viewed as a high standard of random error for the TUG.

To our knowledge, this is the first study to report the MDC and the MDC% of the DGI. The DGI was developed to examine an older individual's ability to modify gait in response to task demands.^{12,26} Recently, the DGI was used to predict the probability of falls in patients with PD.^{13,27} However, MDC and MDC% values were not provided in previous studies. The results of the present study can be used as a reference for the measurement error

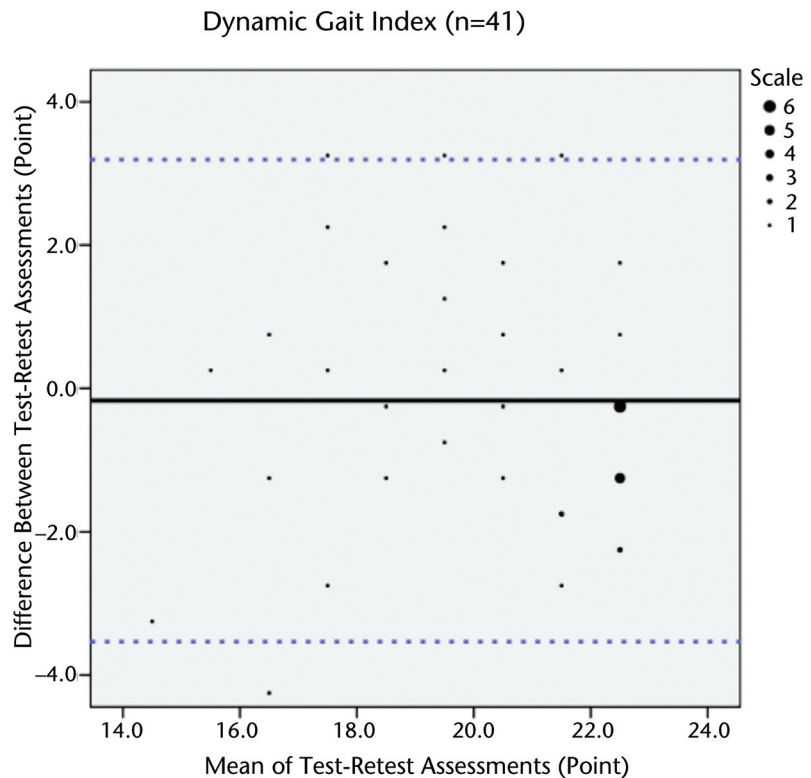


Figure 2.

Bland-Altman method for plotting the difference scores against the mean scores of each pairing of the Dynamic Gait Index (n=41). The solid line represents the mean of the differences. The 2 dashed lines define the limits of agreement (mean of the difference $\pm 1.96 \times$ SD).

of the DGI to help clinicians and researchers determine the true change between successive assessments for patients with PD.

We found that the Pearson r of the TUG was more than .3, implying the existence of heteroscedasticity. Because the difference and the mean of each pair of repeated measurements increased simultaneously, a fixed value for the MDC was not appropriate for all patients with varied walking performance. In such a situation, the MDC% is more appropriate than the MDC for interpreting a true change.²⁰ That is, the amount of random measurement error depends on the initial walking performance of the patient. According to Flansbjer et al,²⁰ a change exceeding the MDC% (ie, 29.8 for the TUG) of the initial test score for an individual patient

could be considered a true change. For example, the score of a patient with an initial TUG score of 11.8 seconds needs to improve by more than 3.5 seconds (11.8×0.298) to indicate a true change. These results should help clinicians interpret changes between test sessions for an individual patient.

The existence of heteroscedasticity also was originally found for the DGI (Pearson $r = -.41$); this result may have been caused by the notable ceiling effect for 31 participants (43% of the initial 72 participants). The Hoehn-Yahr stages for 26 of these 31 participants fell in stages I and II. To remove the possible influence of the ceiling effect, we recalculated the reliability values for the DGI by excluding data from participants with maximum scores in either of

the 2 sessions. We found that the heteroscedasticity was no longer obvious (Pearson $r = -.28$). In addition, the reliability values for the remaining 41 participants (ICC=.78, MDC=3.3 points, and MDC%=16.6) were slightly lower than those for the initial 72 participants (ICC=.84, MDC=2.9 points, and MDC%=13.3). These observations indicated that the notable ceiling effect for the DGI slightly increased the reliability values for the DGI. Thus, the reliability values for the 41 participants were more conservative, but also more appropriate, than those for the 72 participants. On the other hand, the use of the DGI in patients with PD but with better mobility is limited because of the ceiling effect.

For group comparisons (ie, research purposes), an individual-level MDC (MDC_{individual}) can be modified to a group-level MDC (MDC_{group}) with the formula $MDC_{group} = MDC_{individual} / \sqrt{n}$, where n is the size of the sample.²⁸ For example, if the MDC_{individual} of the DGI is 2.9 points, then the MDC_{group} of the DGI will be 0.5 point (for a sample size of 30); such an MDC_{group} is so trivial that it may be neglected. In research contexts, however, the MDC_{group} is negligible, given a substantial sample size, so it is seldom a concern.

Using the MDC_{individual} as a threshold to determine whether a change is true, researchers can report the proportion of the sample for which a change exceeds the MDC_{individual} of an outcome measure when investigating the effects of an intervention.⁶ Research reports regarding a significant change in a sample group after an intervention have revealed little information about the utility of the intervention to clinicians. The significant improvement of a group does not mean that all people in the group achieve real progress, and most people in the group may even fail to

improve. Thus, reporting the proportion of a study sample achieving a true change may help clinicians transfer research outcomes to clinical contexts, an advantage that may improve the utility of the results of a study.

The ICC represents the degree of reproducibility between 2 successive assessments. We found that the ICC of the DGI was .84, indicating high reliability.¹⁹ The ICC of the TUG was .80 (somewhat lower than the ICCs of .85 to .88 reported in previous studies^{14,15}), also representing high reliability.¹⁹ In addition, the mean difference in the test-retest assessments deviated insignificantly from 0, indicating that there was no systematic bias between the successive measurements of the TUG and the DGI. These results support the reproducibility of the TUG and the DGI between successive sessions of assessments for monitoring changes in patients' gait and balance control.

The Bland-Altman plot (Fig. 1) for the TUG scores showed that 2 participants took more than 20 seconds to perform this measure in the retest assessment. In addition, differences between test and retest measurements for both of these participants appeared to be "outliers." Thus, we deleted data from both participants and recalculated the reliability values for the TUG. We found slight variations for the ICC (which changed from .80 to .78), the MDC (which changed from 3.5 seconds to 3.4 seconds), and the MDC% (which changed from 29.8 to 29.6). The Pearson r (which changed from .54 to .34) was greater than .3, indicating that the heteroscedasticity still existed. Although these 2 participants took longer than 20 seconds to perform this measure in the retest assessment, their performance did not substantially influence the results.

To determine whether a change is true, in addition to the measurement error, one must consider fluctuations in the conditions of patients with PD resulting from the concentrations of anti-PD medications. In other words, to compare the performance of a patient with PD between 2 successive sessions, one must accurately arrange a consistent interval between medication ingestion and measure implementation.

In the present study, we estimated the MDCs of the TUG and the DGI, representing the extent of random error and a threshold of statistical significance. In clinical contexts, however, the MID,⁹ representing the degree of change that is meaningful to patients and relevant to clinicians, is equally critical for decision making in treatment planning. To enhance the applicability and interpretability of the TUG and the DGI, future investigations to estimate the MIDs of these 2 measures are warranted.

Our sample was a convenience sample, and our participants had mild to moderately severe PD (ie, Hoehn-Yahr stages I-III). In addition, we excluded participants whose TUG scores exceeded 20 seconds in the first session because of safety concerns. These characteristics of the sample may reduce the generalizability of our findings. Moreover, the Hoehn-Yahr scale provides only a general rating of the severity of PD. Using the Unified Parkinson Disease Rating Scale to rate the severity of disability would be far more informative. Future research with more patients and a more even distribution of disability, from mild to severe (eg, including patients with TUG scores of >20 seconds), may be needed to validate our findings.

Conclusion

The results of our research showed that the DGI and the TUG have generally acceptable random measure-

ment error and reliability in patients with PD. These results should help clinicians and researchers interpret changes in gait and balance deficits in patients with PD over time precisely and confidently.

Ms Huang, Dr Hsieh, and Mr Lu provided concept/idea/research design and writing. Ms Huang, Dr Wu, Dr Lin, and Mr Lu provided data collection. Ms Huang and Mr Lu provided data analysis and project management. Dr Hsieh provided fund procurement. Dr Wu and Dr Tai provided participants. Dr Wu provided facilities/equipment and institutional liaisons. Dr Hsieh and Dr Wu provided consultation (including review of manuscript before submission). The authors are grateful to all participants for their participation.

This study was approved by the Medical Ethics Committee of National Taiwan University Hospital.

This work was supported by the National Science Council (NSC96-2628-B-002-034-MY3) in Taiwan.

This article was submitted April 15, 2009, and was accepted August 13, 2010.

DOI: 10.2522/ptj.20090126

References

- Horak FB, Nutt JG, Nashner LM. Postural inflexibility in parkinsonian subjects. *J Neurol Sci.* 1992;111:46-58.
- Ashburn A, Stack E, Pickering RM, Ward CD. A community-dwelling sample of people with Parkinson's disease: characteristics of fallers and non-fallers. *Age Ageing.* 2001;30:47-52.
- Bloem BR, van Vugt JP, Beckley DJ. Postural instability and falls in Parkinson's disease. *Adv Neurol.* 2001;87:209-223.
- Smithson F, Morris ME, Iansek R. Performance on clinical tests of balance in Parkinson's disease. *Phys Ther.* 1998;78:577-592.
- Hagiwara N, Hashimoto T, Ikeda S. Static balance impairment and its change after pallidotomy in Parkinson's disease. *Mov Disord.* 2004;19:437-445.
- Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006;86:735-743.
- Jette AM, Tao W, Norweg A, Haley S. Interpreting rehabilitation outcome measurements. *J Rehabil Med.* 2007;39:585-590.
- Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. *Phys Ther.* 1998;78:1186-1196.
- Schünemann HJ, Guyatt GH. Commentary: goodbye M(C)ID! Hello MID, where do you come from? *Health Serv Res.* 2005;40:593-597.
- Schreuders TA, Roebroek ME, Goumans J, et al. Measurement error in grip and pinch force measurements in patients with hand injuries. *Phys Ther.* 2003;83:806-815.
- Podsiadlo D, Richardson S. The Timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc.* 1991;39:142-148.
- Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. *J Am Geriatr Soc.* 1986;34:119-126.
- Dibble LE, Lange M. Predicting falls in individuals with Parkinson disease: a reconsideration of clinical balance measures. *J Neurol Phys Ther.* 2006;30:60-67.
- Lim LI, van Wegen EE, de Goede CJ, et al. Measuring gait and gait-related activities in Parkinson's patients own home environment: a reliability, responsiveness and feasibility study. *Parkinsonism Relat Disord.* 2005;11:19-24.
- 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;88:733-746.
- Campbell CM, Rowse JL, Ciol MA, Shumway-Cook A. The effect of cognitive demand on Timed Up and Go performance in older adults with and without Parkinson disease. *Neurol Rep.* 2003;27:2-7.
- Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. *Neurology.* 1967;17:427-442.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420-428.
- Prince B, Makrides L, Richman J. Research methodology and applied statistics, part 2: the literature search. *Physiother Can.* 1980;32:201-206.
- Flansbjerg UB, Holmback AM, Downham D, et al. Reliability of gait performance tests in men and women with hemiparesis after stroke. *J Rehabil Med.* 2005;37:75-82.
- Smidt N, van der Windt DA, Assendelft WJ, et al. Interobserver reproducibility of the assessment of severity of complaints, grip strength, and pressure pain threshold in patients with lateral epicondylitis. *Arch Phys Med Rehabil.* 2002;83:1145-1150.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-310.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135-160.
- Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26:217-238.
- Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med.* 2000;30:1-15.
- Shumway-Cook A, Baldwin M, Polissar NL, Gruber W. Predicting the probability for falls in community-dwelling older adults. *Phys Ther.* 1997;77:812-819.
- Cakit BD, Saracoglu M, Genc H, et al. The effects of incremental speed-dependent treadmill training on postural instability and fear of falling in Parkinson's disease. *Clin Rehabil.* 2007;21:698-705.
- de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures: theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care.* 2001;17:479-487.