Published in final edited form as:

Muscle Nerve. 2018 January ; 57(1): 136–139. doi:10.1002/mus.25700.

RELIABILITY OF THE TRIPLE TIMED-UP-AND-GO (3TUG) TEST

Donald B. Sanders, MD¹, Jeffrey T. Guptill, MD¹, Kathy L. Aleš, MD², Lisa D. Hobson-Webb, MD¹, David P. Jacobus, MD², Riaz Mahmood, MS, BSN³, Janice M. Massey, MD¹, Melissa M. Pittman, BSN, RN¹, Kristi Prather, MPH⁴, Shruti M. Raja, MD¹, Eric Yow, MS⁴, and Vern C. Juel, MD¹

¹Neuromuscular Section, Department of Neurology, Duke University Medical Center, Durham, NC, USA

²Jacobus Pharmaceutical Company, Inc. Princeton, NJ, USA

³Campbell University School of Osteopathic Medicine, Lillington, NC, USA

⁴Duke Clinical Research Institute, Durham, NC, USA

Abstract

Introduction—We report the reliability of a new measure, the Triple Timed-Up-and-Go (3TUG) test, for assessing clinical function in patients with Lambert-Eaton myasthenia (LEM).

Methods—Intra-rater reproducibility and inter-rater agreement of the 3TUG were assessed in 25 control subjects, 24 patients with non-LEM neuromuscular disease and 12 LEM patients. The coverage probability (CP) method was the primary measure of reproducibility and agreement. The *a priori* acceptable range was <20% difference in 3TUG times and a CP 0.90 confirmed agreement.

Results—CP values > 0.90 for intra-rater and inter-rater tests confirmed acceptable reproducibility and agreement for all groups.

Discussion—The 3TUG is a quick, non-invasive and reproducible measure that is easy to perform, measures clinically important weakness in LEM patients and requires little training. Further evaluation in a larger number of LEM patients is in progress to validate the 3TUG as a clinical measure in LEM.

Keywords

Timed up-and-go test; Lambert-Eaton myasthenia; outcome measures; reliability; 3,4-diaminopyridine; coverage probability

Introduction

No single outcome measure has been used or observed to quantify the functional disability in Lambert-Eaton myasthenia (LEM). In this paper, we report the intra-rater reliability

(reproducibility) and inter-rater agreement of a new measure, the Triple Timed-Up-and-Go (3TUG) test, for assessing clinical function in patients with LEM.

Timed-Up-and-Go (TUG) tests measure the time it takes a subject to arise from a chair, walk a short distance and return to the chair; these tests have been used mainly to assess mobility in Parkinsonism and the elderly.¹⁻³ Their test-retest reproducibility is high (intra-class correlation of 0.985–0.988, $p < 0.001$), and it has been suggested that a change of even 3.5⁴ to 4⁵ seconds, or 29.8%,⁴ may represent a clinically significant change.

The 3TUG test was developed to provide a simple, non-invasive measure of disease severity in patients with LEM, who typically have weakness in thigh and hip girdle muscles that causes difficulty walking, climbing stairs and arising from a chair. To assess the potential effect on the TUG of neuromuscular fatigue or facilitation, which are characteristic of LEM, the 3TUG test consists of 3 laps, performed as follows: The subject is seated in a standard 18" high straight-backed armchair. The floor 3 meters from the front legs of the chair is marked with a line of colored tape and the center of the line is marked with an "X." Subjects are instructed to get up from the chair, walk at their normal pace to the line, step on the X, turn around, walk back to the chair, turn around and sit down. This is repeated 3 times without rest. Each lap ends when the subject's back contacts the chair back and the patient is instructed either to begin the next lap or that the test is complete. The 3TUG time is the average of the 3 lap times.

The 3TUG test was successfully used as an entry criterion and as the primary outcome measure in a recently-completed prospective clinical trial of 3,4-diaminopyridine free base (DAP) in LEM, the DAPPER study.^{6,7} To further evaluate the 3TUG test for clinical use, we undertook to assess its reproducibility and agreement when measured by different examiners and when repeated by the same examiner.

Materials and Methods

Study subjects consisted of 25 controls, adults with no history of neuromuscular disease or other difficulty walking, and 24 patients with a non-LEM neuromuscular disease. To minimize the risk of falls during 3TUG testing, patients were not asked to participate in the study if they had fallen within the preceding three months, or if they or their doctor felt there was any physical risk to performing the 3TUG test. The enrolled subjects first performed 3 laps without timing to minimize the effect of learning, followed by a timed trial (Test 1), a 5 minute rest period and a second timed trial (Test 2). All testing was performed with the same assistive device, if any, that the subject normally used at that time of day. Both 3TUG tests were recorded with a video camera placed to capture the full length of the walking course.

Laps were timed "live" independently by two observers; one observer read the instructions for the first test and the other observer did so for the second test. Laps were also timed from the video by an independent observer who was blinded to the live readings and the order in which the 3TUG tests had been performed.

Data were also obtained from the live and blind recorded readings of 3TUG tests performed by 12 LEM patients during the baseline observation phase of the DAPPER study, during

which they were taking a constant dose of DAP.⁶ 3TUG times obtained on two consecutive days at the same time of day and interval after DAP were compared.

This study was approved by the Duke University Institutional Review Board for Clinical Investigations, and all subjects provided informed consent for their participation.

Data Analysis

Intra-rater (test-retest) reproducibility was assessed in all 3 cohorts by comparing each subject's 3TUG time for Test 1 with their time for Test 2 as measured by each observer.

Inter-observer agreement was assessed in all 3 cohorts by comparing the times recorded by each observer with the time(s) of the other observer(s), e.g., observer A vs observer B; A vs C; and B vs C.

The coverage probability (CP) method was used as the primary assessment of reproducibility and agreement.⁸ CP is the probability that the difference between two paired observations is within a pre-established acceptable range. The CP is calculated as the number of observed differences within the acceptable range divided by the total number of comparisons; a CP of 1.0 indicates 100% agreement between the observations. Confidence intervals for the CP were calculated using a binary generalized linear mixed model to account for the correlated outcomes.

Based on reported characteristics of the TUG test in non-LEM patients,^{1–5} the primary endpoint in the DAPPER trial was pre-defined as a deterioration of 30% in the 3TUG time upon withdrawal of study drug.⁷ For the current study, it was determined *a priori* based on the experience of the study neuromuscular physicians that the acceptable range would be a 20% difference in 3TUG times and that agreement would be demonstrated by a CP 0.90, i.e., 90% of all differences are within the acceptable range.⁴

The CP method was selected as the primary evaluation because of its simplicity and broad applicability to different patient populations and clinical settings. 95% binomial confidence intervals for a CP=0.90 were calculated at various sample sizes to estimate precision and to guide the enrollment targets for the control subjects and neuromuscular disease patients. Bland-Altman plots of test-retest results were produced to provide visual validation of the agreement data (Figure 1).⁹ In these plots, differences on the y-axis are plotted against averages on the x-axis to confirm the CP agreement assessments.

Results

Intra-rater (Test-Retest) Reproducibility (Table 1)

Control subjects (Table S1)—The mean percent difference between Test 1 and Test 2 among the 3 observers was 1.54 and none of the pairs exceeded a 20% difference, giving a CP of 1.0 and demonstrating agreement. A Bland-Altman plot (Figure 1A) confirms the CP agreement assessment.

Neuromuscular disease patients (Table S2)—Of the 24 subjects with a non-LEM neuromuscular disease, 16 had myasthenia gravis, 5 had Charcot-Marie-Tooth Disease, 1 had CIDP and 2 had myotonic dystrophy type 1. The mean percent difference between the 2 tests among the 3 observers for the 72 pairs was 1.90 and none of the differences exceeded 20%, again giving a CP of 1.0, and demonstrating agreement. A Bland-Altman plot (Figure 1B) confirms the CP agreement assessment.

DAPPER LEM patients (Table S3)—Among the 12 LEM patients, the mean 3TUG time on Day 0 was 9.37 sec and on Day 1 was 8.96 sec. The difference exceeded 10% in 5 of 24 pairs and exceeded 20% in 2 pairs, resulting in a CP of 0.92, which is above the pre-established threshold of 0.90 for acceptable agreement. A Bland-Altman plot (Figure 1C) confirms the CP agreement assessment.

Inter-rater Agreement (Table 2)

The average difference in 3TUG times measured by different observers was very small in all 3 groups. The percent difference did not exceed 20% (or even 10%) for any of the pairs, resulting in a CP of 1.0 in all groups.

Discussion

The 3TUG was developed to be a measure of dysfunction that could be used to identify LEM patients who benefit from DAP. In this study we have demonstrated that the 3TUG has excellent agreement among independent observers; that within-subject reproducibility is excellent in control subjects and in a selected group of patients with non-LEM neuromuscular disease; and that agreement and reproducibility are acceptable in a small group of patients with LEM. The results also indicate that the 3TUG can be reliably scored from video recordings, which makes it suitable for clinical trials with a central reader.

The use of 3 repetitions of the TUG was based on concern that neuromuscular fatigue or facilitation could affect LEM patients to different degrees and thus potentially adversely affect the ability of the test to determine the effect of treatment among patients with different levels of disease severity.

The 3TUG is a quick, non-invasive, reliable and reproducible measure that is easy to perform, measures clinically important weakness in LEM patients and requires little training. Based on these characteristics, the 3TUG was selected as the primary outcome measure for the DAPPER trial of DAP in LEM and to predict responsiveness of patients in that trial.^{6,7} The expected variability in 3TUG time in patients with LEM on stable doses of medication is <20%. Further evaluation in a larger number of LEM patients with different levels of severity is in progress to confirm this, to determine the value of 3 repetitions, and to validate the 3TUG as a clinical measure in LEM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research reported in this publication was supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR001117.

Dr. Guptill is supported by the National Institute of Neurological Disorders And Stroke of the National Institutes of Health under Award Number K23NS085049.

The DAPPER trial was sponsored by Jacobus Pharmaceutical Co., Inc (JPC), which provided 3,4-diaminopyridine free base for the LEM patients in this study.

The authors confirm that they have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Dr. Sanders is a paid consultant to JPC. Drs. Aleš and Jacobus are employees of JPC. Dr. Juel was site investigator for the DAPPER trial, which was sponsored by JPC.

Abbreviations

CP	coverage probability
DAPPER	prospective trial of 3,4-diaminopyridine free base in Lambert-Eaton myasthenia
LEM	Lambert-Eaton myasthenia
Min	minimum
Max	maximum
NMD	neuromuscular disease
Obs	observer
Blind	blinded observer

References

1. Bischoff HA, Stahelin HB, Monsch AU, Iversend MD, Weyh A, von Dechend M, Akos R, Conzelmann M, Dick W, Theiler R. Identifying a cut-off point for normal mobility: a comparison of the timed "up and go" test in community-dwelling and institutionalised elderly women. *Age Aging*. 2003; 32:315–320.
2. 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. [PubMed: 11175678]
3. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: six-minute walk test, Berg balance scale, timed up & go test and gait speeds. *Phys Ther*. 2002; 82:128–137. [PubMed: 11856064]
4. Huang SL, Hsieh C-L, Wu R-M, Tai C-H, Lin C-H, Lu W-S. 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. [PubMed: 20947672]
5. Ries JD, Echternach JL, Nof L, Blodgett MG. Test-retest reliability and minimal detectable change scores for the timed "up & go" test, the six minute walk test, and gait speed in people with Alzheimer disease. *Phys Ther*. 2009; 89:569–579. [PubMed: 19389792]
6. Sanders DB, Jacobus LR, Aleš KL, Jacobus DP. DAPPER Study Team. Predicting responsiveness to study drug before randomization in the DAPPER trial of 3,4-diaminopyridine in Lambert-Eaton Myasthenic Syndrome (Abstract). *Neurology*. 2015; 84:P7.066.

7. Sanders DB, Juel VC, Harati Y, Smith AG, Peltier AC, Marburger T, Lou JS, Pascuzzi R, Richman D, Xie T, Jacobus LR, Jacobus DP. DAPPER Study Team. Results from the DAPPER study: inpatient double-blind, placebo-controlled withdrawal study of 3,4-diaminopyridine in patients with Lambert-Eaton myasthenic syndrome (Abstract). *Muscle Nerve*. 2015; 52:S13. [PubMed: 26384854]
8. Barnhart HX, Yow E, Crowley AL, Daubert MA, Rabineau D, Bigelow R, Pencina M, Douglas PS. Choice of agreement indices for assessing and improving measurement reproducibility in a core laboratory setting. *Stat Methods Med Res*. 2014
9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; i:307–310.

Figure 1. Test-Retest Comparisons

Bland-Altman plots of data from Test 1 and Test 2. The top solid line in each plot shows the bias.

A. Controls. Data from 75 paired 3TUG times measured by 2 live and 1 blinded observer on 25 control subjects. All test-retest differences are less than 1.25 seconds and 71/75 are within $\pm 1.96\text{SD}$ of the mean. The measurement for Test 1 was on average 0.13 sec greater than that for Test 2.

B. Neuromuscular disease. Data from 72 paired 3TUG times for Test 1 and Test 2 measured by 2 live and 1 blinded observer on 24 patients with non-LEM neuromuscular disease. All test-retest differences are less than 2 seconds and 67/72 are within $\pm 1.96SD$ of the mean. The measurement for Test 1 was on average 0.22 sec greater than that for Test 2.

C. LEM. Data from 24 paired 3TUG times for Day 0 and Day 1 from a live and a blinded observer on 12 LEM patients. All test-retest differences are less than 2 seconds and 22/24 are within $\pm 1.96SD$ of the mean. The measurement for Test 1 was on average 0.42 sec greater than that for Test 2.

Table 1

Intra-rater reproducibility results (test-retest)

	Controls	NMD	LEM DAPPER
No. of Subjects	25	24	12
No. of Pairs	75	72	24
Age, years			
Males, N (%)	17 (68)	15 (63)	4 (33)
Mean (SD)	47 (15)	61 (24)	56 (15)
Min, Max	24, 72	19, 86	35, 72
Females, N (%)	8 (32)	9 (38)	8 (67)
Mean (SD)	41 (15)	54 (19)	57 (20)
Min, Max	23, 70	25, 77	23, 83
Test 1			
Average, sec	8.15	10.25	9.37
Min, Max, sec	5.71, 10.23	8.29, 15.36	6.37, 15.27
Test 2			
Average, sec	8.01	10.03	8.96
Min, Max, sec	5.95, 10.59	7.77, 13.70	6.03, 13.73
Difference (Test 1–Test 2)			
Average (bias), sec	0.13	0.22	0.42
Min, Max, sec	−0.66, 1.22	−1.33, 1.95	−0.68, 1.90
Mean % Difference	1.54	1.90	4.07
CP (95% CI)			
20% difference	1.0 (NE)	1.0 (NE)	0.92 (0.59, 0.99)
10% difference	0.96 (0.77, 0.99)	0.83 (0.67, 0.93)	0.79 (0.51, 0.93)

Abbreviations: CP, coverage probability; DAPPER, prospective trial of 3,4-diaminopyridine free base in Lambert-Eaton myasthenia; F, female; LEM, Lambert-Eaton myasthenia; M, male; Min, minimum; Max, maximum; NE, not estimable; NMD, neuromuscular disease; SD, standard deviation.

Table 2

Inter-rater agreement results

Item	Controls		NMD		LEM DAPPER	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
No. of Subjects	25	25	24	24	12	12
No. of Pairs	75	75	72	72	12	12
Difference in Pairs						
Average (bias), sec	-0.01	-0.02	-0.04	0.00	-0.01	-0.02
Min, sec	-0.28	-0.46	-0.48	-0.49	-0.18	-0.09
Max, sec	0.26	0.51	0.26	0.33	0.07	0.07
Mean % Difference	-0.14	-0.37	-0.38	-0.07	0.23	-0.22
CP (95% CI)						
20% difference	1.0 (NE)	1.0 (NE)	1.0 (NE)	1.0 (NE)	1.0 (NE)	1.0 (NE)
10% difference	1.0 (NE)	1.0 (NE)	1.0 (NE)	1.0 (NE)	1.0 (NE)	1.0 (NE)

Abbreviations: CP, coverage probability; DAPPER, prospective trial of 3,4-diaminopyridine free base in Lambert-Eaton myasthenia; LEM, Lambert-Eaton myasthenia; Min, minimum; Max, maximum; NE, not estimable; NMD, neuromuscular disease.