Phase-Level Analytical Validation of a Smartphone-Based Timed Up and Go (TUG) Assessment in Multiple Sclerosis

Abreu, M[1]; Clancy, C[1]; Lutz, J[1,2,3]

[1] Beats Medical, Dublin, Ireland
[2] Boston University, Chobanian & Avedisian School of Medicine, Boston, USA
[3] Luja Consulting LLC, Pittsfield, USA

Methodological Issue Being Addressed

Smartphone-based mobility assessments offer a scalable approach for decentralized clinical trials and real-world monitoring. However, analytical validation of derived mobility endpoints remains challenging.

For the Timed Up and Go (TUG), total duration is the primary outcome: accurate estimation from smartphone sensors is an important benchmark for broader use in clinical studies while also enabling evaluation of more granular movement features.

Establishing agreement between smartphone-derived movement states and observable functional mobility behaviors requires structured reference methods capable of linking sensor outputs to real-world movement events.

This work presents a framework for phase-level analytical validation of a smartphone-based TUG, anchored in accurate estimation of total test duration, and extended to phase-level characterization using synchronized video annotation.

Introduction

The TUG is a widely used assessment of functional mobility across neurological populations, including MS (1). Instrumented and smartphone-based implementations can extend the assessment beyond total duration to features such as turning and gait characteristics, while also enabling remote and bring-your-own-device (BYOD) assessment. However, validating these derived phase-level measures against observable functional events remains challenging. We evaluated a structured video-based approach to validate total TUG duration and phase-level smartphone-derived metrics in individuals with multiple sclerosis (MS).

Methods

Participants

9 individuals with MS
Recruited through the Portuguese Multiple Sclerosis Society (SPEM).

44±10 years
Mean age across the cohort.

67% female

Participants represented a heterogeneous mobility cohort with some participants using walking aids.

Assessment

Smartphone-based TUG assessments were completed by patients in a clinic setting using a standard smartphone (Android).

Video Reference

Synchronized video recordings were independently annotated by 3 raters using a literature-informed annotation framework to provide a reference for smartphone-derived measures (2,3). Key TUG phases included sit-to-stand, walking, turning. Step counts were annotated as well. Inter-rater reliability was assessed using the intraclass correlation coefficient (ICC).

Analysis

Agreement between smartphone-derived and video-derived measurements was evaluated using mean absolute error (MAE), Spearman correlation (ρ), and Bland-Altman agreement analyses. Analyzable sample size varied by endpoint because of incomplete video capture (n=8–9).



Good Morning
Day 1
Select to start
Mobility
Dexterity
Speech
Hi there.
Let's complete your scheduled Nine-Hole Peg Test. Do you want to start now?
Next
Start
Your impact in 7 days (0%)
Home

TUG Phases

Results

ICC: High inter-rater reliability was observed for total TUG duration (**ICC=0.99**), walking duration (**ICC=0.95**), and sit-to-stand duration (**ICC=0.80**), supporting the annotation framework as a reference method.

Total TUG duration: close agreement with video measurements (MAE=0.41 seconds; bias=-0.40 seconds; Spearman ρ=1.00).

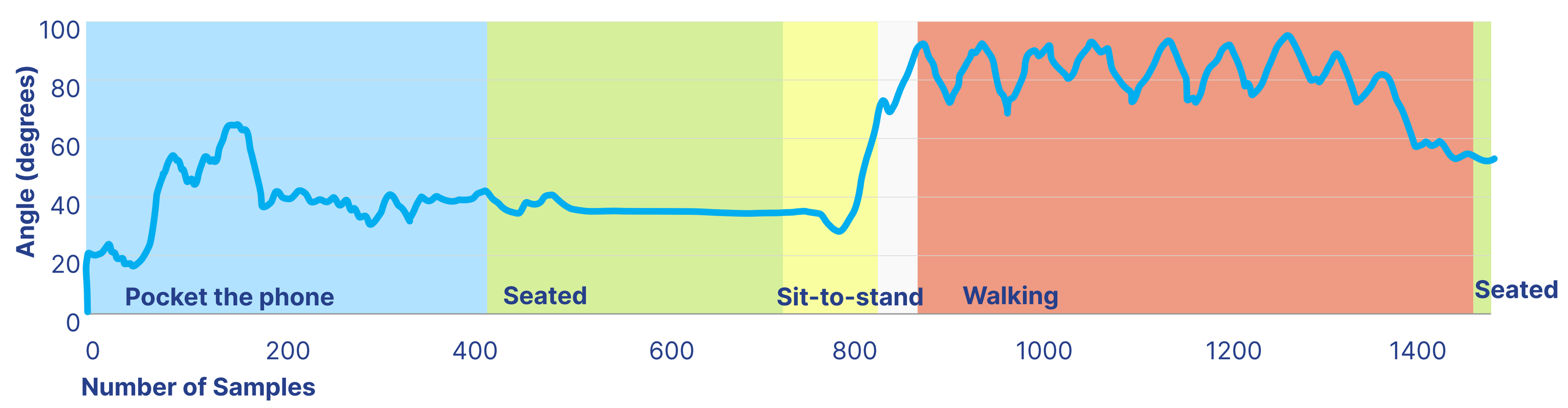
Walking duration: strong agreement (bias=-0.16 seconds; Spearman ρ=1.00), with all observations within Bland-Altman limits of agreement.

Sit-to-stand duration: MAE of 0.36 seconds and minimal bias (0.05 seconds), although with greater variability due to the short duration of this phase.

Step count: correlated strongly with video reference (Spearman ρ=0.89, p=0.001).

Turn detection: 16 of 18 turns identified, complete detection in 8 of 9 patient assessments.

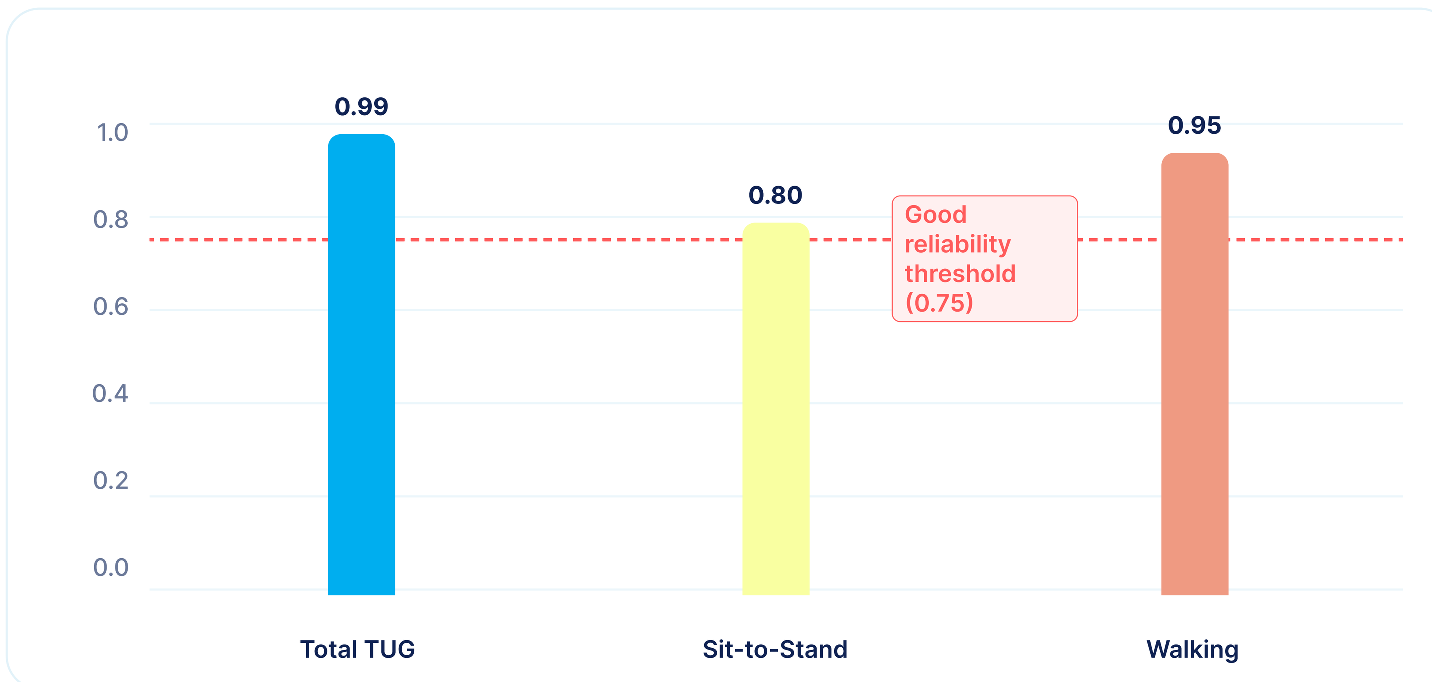
Example Smartphone-Derived TUG Phase Classification



Angle (degrees)
100
80
60
40
20
0
Pocket the phone
Seated
Sit-to-stand
Walking
Seated
0
200
400
600
800
1000
1200
1400
Number of Samples

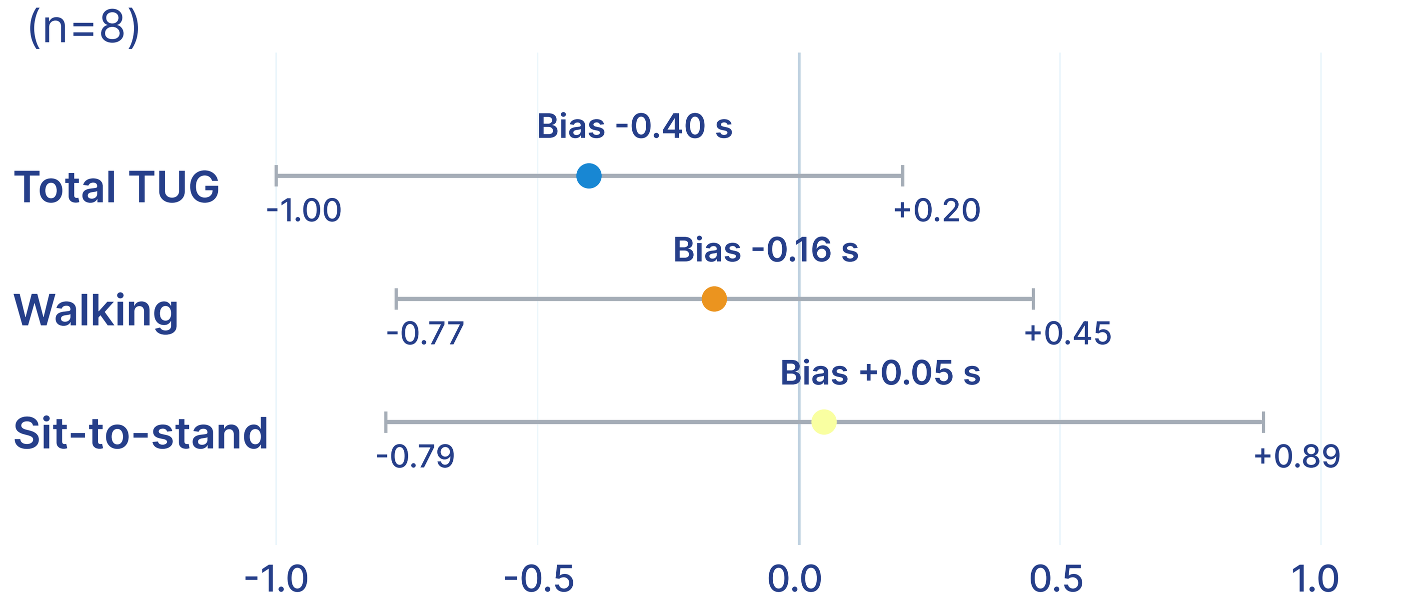
Inter-Rater Reliability (ICC)

Intraclass Correlation Coefficients across TUG sub-phases



1.0
0.8
0.6
0.4
0.2
0.0
0.99
0.80
0.95
Good reliability threshold (0.75)
Total TUG
Sit-to-Stand
Walking

Bland-Altman Agreement Across TUG Measures (n=8)



Bias -0.40 s
Total TUG
-1.00
+0.20
Bias -0.16 s
Walking
-0.77
+0.45
Bias +0.05 s
Sit-to-stand
-0.79
+0.89
-1.0
-0.5
0.0
0.5
1.0

Dot = mean bias | Line endpoints = 95% limits of agreement

Smartphone – video difference (seconds)

Conclusions

• Structured video annotation provided a reproducible reference for phase-level validation of smartphone-derived TUG outcomes.
• Supports robust estimation of the primary endpoint (TUG duration) and evaluation of additional mobility features, including turning and transitional movements.
• Tested across heterogeneous presentations, including patients using walking aids.
• Supports continued development of smartphone-based mobility assessments for decentralized and hybrid CNS clinical trials

Limitations

• Small, single-site sample (N=9) assessed in a clinic-based, researcher-observed setting; performance under unsupervised home-use conditions was not evaluated.
• Sensor agreement was assessed on a single Android device; cross-platform performance and reproducibility across neurological disorders remain to be established.

References:

1 Sebastião E, Sandroff BM, Learmonth YC, Motl RW. Validity of the Timed Up and Go Test as a Measure of Functional Mobility in Persons With Multiple Sclerosis. Arch Phys Med Rehabil. 2016;97(7):1072–1077. doi:10.1016/j.apmr.2015.12.031.

2 Ortega-Bastidas P, Gómez B, Aqueveque P, Luarte-Martínez S, Cano-de-la-Cuerda R. Instrumented Timed Up and Go Test (iTUG)—More Than Assessing Time to Predict Falls: A Systematic Review. Sensors. 2023;23(7):3426. doi:10.3390/s23073426

3 Johnston W, Judice PB, Molina García P, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. Br J Sports Med. 2021;55(14):780–793. doi:10.1136/bjsports-2020-103147.

Acknowledgements/Disclosures: We thank the Portuguese Multiple Sclerosis Society (SPEM) and its clinical team for supporting participant recruitment and study conduct. This work was funded by Beats Medical. One or more authors report potential conflicts which are described in the program.

SPEM SOCIEDADE PORTUGUESA DE ESCLEROSE MÚLTIPLA

BEATS MEDICAL