

Reliability of functional outcome measures in adults with neurofibromatosis I

SAGE Open Medicine

Volume 6: 1–7

© The Author(s) 2018

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/2050312118786860

journals.sagepub.com/home/smo

Rebecca L Mullin^{1,2} , John F Golding³, Rebecca Smith^{1,2}, Victoria Williams¹, Mary Thomas¹ and Rosalie E Ferner¹

Abstract

Objectives: To determine intra-rater and inter-rater reliability of functional outcome measures in adults with neurofibromatosis I and to ascertain how closely objective and subjective measures align.

Methods: A total of 49 ambulant adults with neurofibromatosis I aged 16 years and over were included in this observational study: median age 31 years (range: 16–66 years), 29 females, 20 males. Participants were video-recorded or photographed performing four functional outcome measures. Four raters from the neurofibromatosis centre multi-disciplinary team independently scored the measures to determine inter-rater reliability. One rater scored the measures a second time on a separate occasion to determine intra-rater reliability. The measures evaluated were the functional reach, timed up and go, 10 m walk and a modified nine-hole peg tests. Participants also completed a disease-specific quality-of-life questionnaire.

Results: Inter-rater reliability and intra-rater reliability scores (intra-class coefficient) were similar for each outcome measure. Excellent rater agreement (intra-class coefficient, $r \geq 0.9$) was found for the functional reach, timed up and go and the 10 m walk tests. Rater agreement was good for the modified nine-hole peg test: intra-class coefficient $r = 0.75$ for intra-rater reliability and 0.76 for inter-rater reliability. The timed up and go and the 10 m walk tests correlated highly with perceived mobility challenges in the quality-of-life questionnaire.

Conclusion: The functional reach, timed up and go and 10 m walk tests are potentially useful outcome measures for monitoring neurofibromatosis I treatment and will be assessed in multi-centre and longitudinal studies.

Keywords

Neurofibromatosis I, outcome measure, rater reliability, standard error measurement, minimal detectable change

Date received: 7 December 2017; accepted: 5 June 2018

Introduction

Neurofibromatosis 1 (NF1) is a common, inherited disease associated with benign and malignant peripheral nerve sheath tumours.¹ The complications of NF1 are variable, unpredictable and widespread, and cognitive impairment is common.¹ Challenges with functional tasks such as walking, balance or using the hands are common in NF1.² They may arise from tumours causing pressure on peripheral nerves or the spinal cord, central nervous system tumours or skeletal abnormalities. The relationship between objective performance and subjective experience is also important. Functional challenges may be amenable to medical, surgical or physical interventions and there is a need for robust functional outcome measures in this patient group to assess treatment efficacy. To the best of our knowledge, there has been no systematic evaluation of functional, motor outcome measures in adults with this disease.

An essential requirement of robust outcome measurements is that they are reliable.³ Reliability is defined as ‘the degree to which measurement is free from measurement error’.⁴ It is important to evaluate properties such as

¹National Centre for Neurofibromatosis, Guy’s Hospital, Guy’s and St Thomas’ NHS Foundation Trust, London, UK

²Department of Physiotherapy, Guy’s and St Thomas’ NHS Foundation Trust, London, UK

³Department of Psychology, University of Westminster, London, UK

Corresponding author:

Rebecca L Mullin, National Centre for Neurofibromatosis, Guy’s Hospital, Guy’s and St Thomas’ NHS Foundation Trust, Great Maze Pond, London SE1 9RT, UK.

Email: Rebecca.mullin@gstt.nhs.uk



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

reliability within the target population, as variability within a disease strongly influences outcome measurement results.³

Inter-rater reliability requires the same group of subjects to be measured at the same time by different observers, and intra-rater reliability considers the same subjects and the same observer with measurements taken at different time points.⁵ Absolute reliability is expressed as the standard error of measurement (SEM) and this can be calculated from rater reliability. Minimal detectable change (MDC) describes the minimal amount of change in the instrument score to be sure that the score change is not attributable to measurement error. This may be calculated from SEM.⁶

The INF1-QOL questionnaire (impact of NF1 on quality-of-life questionnaire) is a validated, reliable disease-specific questionnaire.² Responders categorise problems as no issues, mild, moderate or severe, in the 14-item, self-report, quality-of-life questionnaire. It includes two functional domains: walking and using the hands.

Advances in molecular biology have facilitated the development of novel therapy to include drugs that have the potential to treat symptomatic neurofibromas which has accelerated the quest for functional outcome measures. The primary aim of this study was to evaluate inter- and intra-rater reliability of four commonly used gait, balance and hand function outcome measures in adults with NF1. From these data, we calculated the SEM and MDC. The secondary aim of this study was to correlate patients' perceived mobility and upper limb function as rated through the INF1-QOL questionnaire with their objective functional outcome measurement scores.

Methods

Guy's and St. Thomas' NHS Foundation Trust is a national centre for the diagnosis, management and support of 1150 people with NF1.

All adults (aged 16 years and over) with NF1 who attended their clinic appointments during the 4-month recruitment period (May–September 2015) were approached by letter inviting them to take part in this observational study. We aimed to recruit 50 participants as recommended for a reliability study.³ At the time of their appointment, the treating clinician (doctor or nurse) confirmed they met the inclusion/exclusion criteria and ascertained whether they wished to take part or not.

To be included in this study, the participants needed to meet the following requirements: have a clinical diagnosis of NF1, be aged 16 years or over, have sufficient cognition to provide informed consent, not have significant mobility or balance impairments that are unrelated to their NF1 and be able to walk more than 10m without physical assistance (may use walking aids).

Written consent was collected by the researcher and the participant was given a unique alphanumeric research identification code. Participants provided demographic information

and completed a NF1 quality-of-life patient-reported outcome measure (INF1-QOL). Each participant completed three repetitions of each of the chosen outcome measures while being video recorded or photographed by the researcher.

Ethical permission was granted by National Research Ethics Service- Hampstead, reference 15/LO/1084.

Outcome measurement selection

A review of the evidence base identified that a wide variety of motor performance outcome measures have undergone metric evaluation in comparable cohorts such as chronic pain, community dwelling older adults with multi-morbidity, spinal cord injury, stroke and multiple sclerosis (MS). The research team chose to evaluate motor performance outcome measures for walking, balance and use of the hands, based on the functional challenges identified by people who have NF1 in a pre-study focus group and functional challenges identified in the INF1-QOL questionnaire.² The four selected outcome measures were chosen based on a high rate of rater reliability in comparable conditions, and following communication with clinicians and researchers in this field, who specified that the outcome measurements needed to be quick and easy to perform and interpret in the outpatient clinic environment to ensure long-term uptake into practice.

The functional reach test assesses standing balance. In the functional reach test,⁷ the participant stands parallel to a wall with arms at 90° of shoulder flexion and reaches forward as far as they can without taking a step. A photograph was taken at the furthest point that the participant was able to reach and measurements are recorded to 1 mm.

The timed up and go test assesses functional mobility. In the timed up and go test,⁸ participants stand from a chair, walk 3 m, turn around and return to the chair. Measurements are recorded to milliseconds.

The 10 m walk test assesses functional mobility and gait speed. In the 10 m walk test,⁹ participants walk at their normal speed along a measured walkway. Measurements are recorded to milliseconds.

The modified nine-hole peg test assesses upper limb function through dexterity. In the modified nine-hole peg test, the participant takes pegs from a bowl and places them into the holes of a peg board. Measurements are recorded to milliseconds on a digital stopwatch.

Rating process

Video recordings and photographs of participants performing the outcome measurement tests were immediately transferred to a secure electronic location. Four raters watched and rated the videos and photographs separately to assess inter-rater reliability and they posted their scores for each test into a sealed box. One of the raters rated the photographs and video recordings a second time, to assess intra-rater reliability. The rater team comprises four experienced members

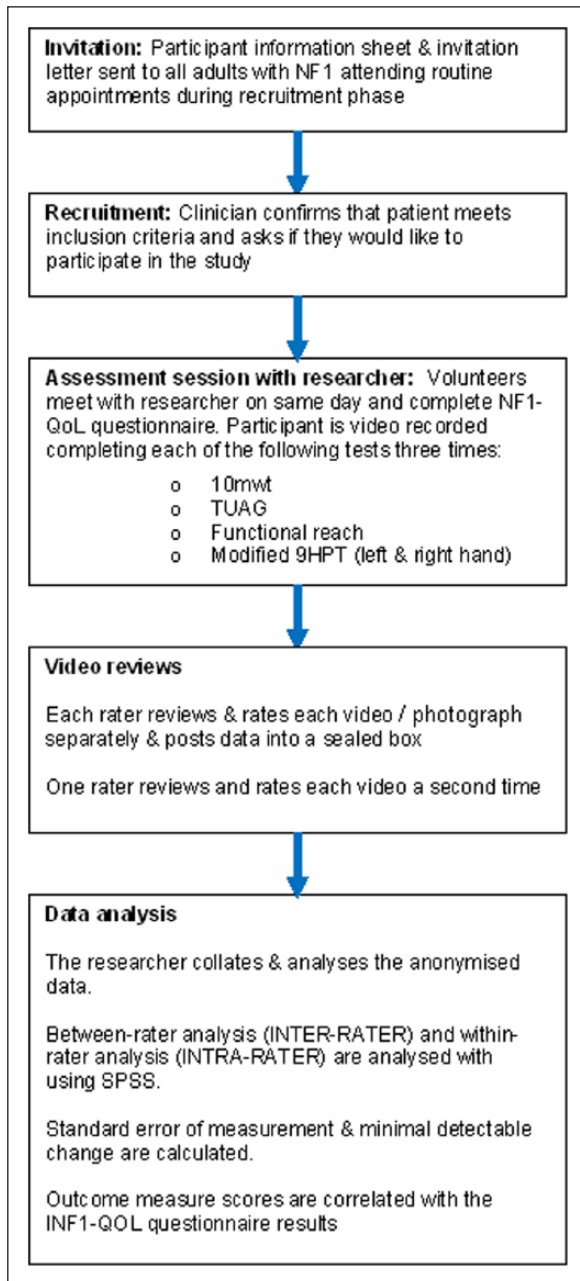


Figure 1. Study flow diagram.

of the NF1 multi-disciplinary team including doctors, a nurse and a physiotherapist. The researcher collated these data onto a spreadsheet. Data were transferred to SPSS for statistical analysis (Figure 1).

Bias

Several steps were taken to counter bias in this study. The researcher was not involved in the recruitment process or as a video rater to reduce the risk of researcher bias such as selection bias. The intra-rater reliability tester was instructed to watch the videos a second time, only after they had

watched all 49 sets of the videos through once to ameliorate recall bias. Outcome measurements were completed with the same researcher (R.M.) with standardised instructions to reduce the risk of performance bias. Videos were taken of the outcome measurement sessions and used for analysis to ensure all raters saw the same test after being provided with training on how to interpret findings to reduce the risk of detection bias.

Statistical analysis

Data from all measures were analysed using the IBM statistical package for social sciences (SPSS) version 23. A two-way mixed effects model was used to calculate intra-class coefficient (ICC 3,1) and evaluate relative intra-rater reliability of the 10 m walk test, timed up and go test, functional reach test and nine-hole peg test. A two-way random-effects model was used (ICC 2,1) to evaluate inter-rater reliability of the 10 m walk test, functional reach test, timed up and go and nine-hole peg test (see Table 2). The statistical analysis processes align with other studies investigating inter- and intra-rater reliability of the selected functional outcome measures.^{10–12}

The ICC is a number between 0 and 1: 1 represents perfect reliability with no measurement error and 0 represents no reliability. Values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability.¹³

The SEM, absolute inter- and intra-rater reliability, was calculated for each measure in adults with NF1 with the following equation

$$SEM = SD\sqrt{1-r}$$

MDC was calculated for each measure from the SEM using the following equation

$$MDC = SEM \times 1.96 \times \sqrt{2}$$

where MDC is the minimal detectable change, SEM is the standard error of measurement and r is the reliability (ICC).

Results

A total of 85 adults with NF1 were sent invitation letters for this study: 15 did not wish to take part, 14 did not meet the eligibility criteria and 7 had not received the participant information sheet before their appointment (more than 24 h). Thus, 49 ambulant adults with NF1 volunteered and participated in the study.

There were 29 females and 20 males in this study with a mean age of 31 years (range: 16–66 years). The range of scores for the INF1-QOL was 1–26, mean score was 9 (where 0 indicates no difficulty and 42 indicates severe difficulty in all 14 domains).

Table 1. Rater reliability: intra-class coefficient (ICC) scores with 95% confidence intervals (95% CI) for outcome measures.

Functional test	Intra-rater reliability		Inter-rater reliability	
	ICC	95% CI	ICC	95% CI
Functional reach	0.90	0.86, 0.94	0.90	0.87, 0.94
Timed up and go	0.97	0.96, 0.98	0.97	0.96, 0.98
10m walk	0.99	0.98, 0.99	0.95	0.93, 0.97
10m walk speed	0.99	0.98, 0.99	0.96	0.94, 0.96
Modified nine-hole peg test	0.75	0.67, 0.83	0.76	0.69, 0.84

Table 2. Mean scores for each outcome measure with standard deviation, range, SEM and MDC scores, calculated from inter-rater reliability.

Functional test	Mean score	1 Standard deviation	Range (min–max)	SEM	MDC
Functional reach (cm)	31.98	9.68	4.00–38.43	2.92 cm	8.08 cm
Timed up and go (s)	11.63	7.94	5.5–49.82	1.03 s	2.86 s
10m walk time (s)	6.73	3.80	3.76–26.23	0.87 s	2.40 s
10m walk speed (m/s)	1.70	0.45	0.39–2.62	0.09 m/s	0.26 m/s
Modified nine-hole peg test (s)	18.36	5.10	11.47–53.62	2.79 s	7.73 s

SEM: standard error of measurement; MDC: minimal detectable change.

Table 1 details the ICC for intra-rater and inter-rater reliability for each of the four outcome measurements with 95% confidence intervals (95% CI). Reliability (ICC) was excellent with low measurement error and tight 95% CIs for the functional reach, timed up and go and 10m walk test time and speed. The modified nine-hole peg test had lower ICCs and wider 95% CI than the other three measures. For each outcome measure, ICC and 95% CI values were comparable for inter- and intra-rater reliability.

Table 2 details the mean score and range for each of the above tests alongside clinically important MDC. There was a more or less continuous distribution for all measures. The wide range of times was not simply due to outliers but reflected the clinical heterogeneity of NF1.

Table 3 details the correlations between each functional outcome measure and the INF1-QOL questionnaire which measured patient reported, disease-specific quality of life. Pearson correlations were computed between each functional test and the total INF1-QOL questionnaire scores, subsections for question 7 for walking and question 8 for hand function. As can be seen, all functional tests correlated significantly with the INF1-QOL total score. For question 7 'walking', the best correlations were for the two measures of mobility. By contrast, the correlations with question 8 'hand function', were largely non-significant.

Discussion

In this study, we evaluated the inter- and intra-rater reliability of a set of functional outcome measures in 49 adults with NF1 who were representative of the typical ranges of disease severity seen in a published study of quality of life in adults with NF1.² The functional reach test, timed up and go test

and 10m walk tests demonstrated excellent reliability for both inter- and intra-rater reliability. Interestingly, the modified nine-hole peg test demonstrated lower inter- and intra-rater reliability than the other measures tested.

The reliability scores (ICC) for inter- and intra-rater reliability of the functional reach test were excellent. They align with high levels of inter- and intra-rater reliability for normal¹⁴ and frail elderly adults.¹⁵ The SEM calculated from these findings was similar to multiple different clinical populations, with published standard errors in measurement between 1.86 and 2.91 cm for spinal cord injuries,¹⁶ stroke¹¹ and peripheral vestibular disorders.¹⁷ The MDC for this measure was variable between different clinical populations: from 5 cm in spinal cord injury patients to 11.5 cm in Parkinson's.^{18,19} Relative to the mean functional reach score, an MDC of 8.08 cm in NF1 was deemed acceptable.

Inter- and intra-rater reliability scores for the timed up and go test were also excellent. They align with high levels of inter- and intra-rater reliability for healthy, normal older adults.²⁰ There are no published data on SEM or MDC for the timed up and go test so currently, it is not possible to compare our findings against other clinical groups. Both SEM and MDC scores were deemed acceptable in NF1.

Inter- and intra-rater reliability scores for the 10 m walk test were also excellent. They align with healthy adults, spinal cord injuries, stroke and traumatic brain injuries with ICC greater than 0.9.^{21–24} The SEM aligns with comparable clinical populations including spinal cord injuries,¹⁶ strokes¹¹ and geriatrics.²⁵ An MDC of 0.26 m/s is similar to MDC scores for similar clinical populations.

The reliability scores (ICC) for the modified nine-hole peg test were 0.75 and 0.76 for intra-rater and inter-rater reliability, respectively. This is lower than the classic nine-hole

Table 3. Correlation Pearson (r) between each functional test and subsections of the INFI-QOL questionnaire with significance level.

Functional test	Correlation with questionnaire total score, r	Correlation with question 7 walking, r	Correlation with question 8 hand function, r
Functional reach (cm)	-0.32*	-0.47**	-0.19 ns
Timed up and go (s)	0.43**	0.71**	0.25 ns
10m walk time (s)	0.42**	0.73**	0.18 ns
10m walk speed (m/s)	-0.51**	-0.71**	-0.23 ns
Modified nine-hole peg test (s)	0.36*	0.48**	0.32*

ns: not significant.

* $p < 0.05$; ** $p < 0.01$.

peg test²⁶ when used in healthy adults and in multiple sclerosis, where reliability scores were greater than 0.9.^{27,28} We cannot ignore that this may be because we used a modified version of the test, but based on rater feedback we suggest that this test may be difficult for people with NF1 because the test requires sustained concentration. Individuals with NF1 often have cognitive impairment including difficulty with concentration and planning meaning that the start and end points of the test were difficult to determine because some participants hesitated before continuing the task. There may also be a muscle force component.²⁹

High levels of rater reliability indicate consistency in interpretation, within rater and between raters. MDC scores calculated from data collected within this study provide the clinician with an important marker that may assist with their decision-making. The next stage of metric evaluation of these outcome measures, test-retest reliability, will reveal how stable the measures are over a time period where NF1 symptoms remain stable.

There was an overall correlation between the functional outcome measurement scores with the total score for the quality-of-life questionnaire (INFI-QOL), but greater correlations with sub-items. The highest level of correlation was between the mobility outcome measures (timed up and go and 10m walk test) and question 7 of the INFI-QOL measure which related to walking. The functional reach test is less closely aligned with any questions in the INFI-QOL questionnaire but still achieves statistical significance (-0.47). This may be because the functional reach specifically evaluates standing balance, a factor not specifically targeted within the INFI-QOL questionnaire. As balance and falls were raised as an important concern for people with NF1 in the pre-study focus group, this measure may still be of benefit and deserves further exploration as part of future trials. Interestingly, there was a small but significant correlation (-0.48) between the modified nine-hole peg test and question 8 of the INFI-QOL measure which relates to hand function. As cognitive processing contributes to the time taken to perform this test and rater reliability is not as good as the other measures, it does not appear to be as useful an outcome measure for assessing upper limb function in this patient group and measures such as grip dynamometry may be more appropriate.

Limitations of study

To our knowledge, this is the first study to evaluate reliability of functional outcome measures through use of videos. By video recording and photographing participants, we could be certain that all raters analysed the same test, from the same angle and it also ensured that the researcher could continue to stand close to the participant for balance and mobility testing and ensure safety as per routine clinical practice. We acknowledge that outcome measurement assessments are not normally conducted through video media but this testing regime was acceptable to participants and the research team who needed to fulfil research duties around their clinical commitments. We chose to evaluate reliability of outcome measures across a variety of professions (doctors, physiotherapists, nurses) to ensure that raters represented the multi-disciplinary team. We were limited by the time available to carry out the study but as outlined, this did not impact on the value of our data. Assessment of other upper limb outcome measures would have proved fruitful and this will be the focus of future work.

Originally we aimed to recruit 50 individuals for the study, but achieved 49 participants. Although this might appear to be a limitation, in practical terms it did not alter the reliability estimates. This is because the reliabilities that we observed were predominantly 0.9 or better apart from the modified nine-hole peg test which was 0.75–0.76. Indeed, Shoukri et al.³⁰ demonstrate that sample size theory demonstrates that our study is adequately powered for reliabilities in the range observed in our study.³⁰ Many reliability studies, using similar methods, have employed as few as 16 participants to evaluate reliability.¹²

Conclusion

There is a need for reliable functional outcome measures to monitor treatment and to evaluate novel therapy in NF1 adults. The functional reach, timed up and go and 10m walk tests had excellent inter- and intra-rater reliability and were quick and easy to perform in a clinic setting. Furthermore, these tests correlated highly with perceived functional challenges of mobility in the INFI-QOL questionnaire. The modified nine-hole peg test was slightly less reliable and

other upper limb measures such as dynamometry should be evaluated in this group of patients. Our future aims will be to evaluate these motor outcome measures in multi-centre and longitudinal studies. We will also use them as tools for assessing patient outcome of therapeutic interventions.

Acknowledgements

The authors acknowledge Alexandra Curtis and Rona Inniss for their valuable contributions in the write up phase of this study.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval

Ethical approval for this study was obtained from National Research Ethics Service- Hampstead reference 15/LO/1084.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Informed consent

Written informed consent was obtained from all participants in this study, including ages 16–18 years. In the United Kingdom, young people over the age of 16 years are deemed able to provide informed consent when ‘Gillick competent’. This includes all studies that do not involve clinical trials of an interventional product (CTIMPS)*. <http://www.hrdecisiontools.org.uk/consent/principles-children-EngWalesNI.html>

Trial registration

This study is registered at ClinicalTrials.gov identifier: NCT02479360.

ORCID iD

Rebecca L Mullin  <https://orcid.org/0000-0002-2571-104X>

References

1. Ferner RE and Gutmann DH. Neurofibromatosis type 1 (NF1): diagnosis and management. *Handb Clin Neurol* 2013; 115: 939–955.
2. Ferner RE, Thomas M, Mercer G, et al. Evaluation of quality of life in adults with neurofibromatosis 1 (NF1) using the Impact of NF1 on Quality Of Life (INF1-QOL) questionnaire. *Health Qual Life Outcomes* 2017; 15(1): 34.
3. De Vet HCW, Terwee CB, Mokkink LB, et al. *Measurement in medicine: a practical guide*. Cambridge: Cambridge University Press, 2011.
4. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; 63(7): 737–745.
5. Stokes EK. *Rehabilitation outcome measures*. Edinburgh: Churchill Livingstone, 2011, pp. 27–34.
6. Stratford PW, Binkley JM and Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther* 1996; 76(10): 1109–1123.
7. Duncan PW, Weiner DK, Chandler J, et al. Functional reach: a new clinical measure of balance. *J Gerontol* 1990; 45(6): M192–M197.
8. Podsiadlo D and Richardson S. The timed ‘Up & Go’: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991; 39(2): 142–148.
9. Macfarlane PA and Looney MA. Walkway length determination for steady state walking in young and older adults. *Res Q Exerc Sport* 2008; 79(2): 261–267.
10. Botolfson P, Helbostad JL, Moe-Nilssen R, et al. Reliability and concurrent validity of the Expanded Timed Up-and-Go test in older people with impaired mobility. *Physiother Res Int* 2008; 13(2): 94–106.
11. Flansbjerg UB, Holmbäck AM, Downham D, et al. Reliability of gait performance tests in men and women with hemiparesis after stroke. *J Rehabil Med* 2005; 37(2): 75–82.
12. Poncumhak P, Saengsuwan J, Kamruecha W, et al. Reliability and validity of three functional tests in ambulatory patients with spinal cord injury. *Spinal Cord* 2013; 51(3): 214–217.
13. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15(2): 155–163.
14. Bennie S, Bruner K, Dizon A, et al. Measurements of balance: comparison of the Timed ‘Up and Go’ test and Functional Reach test with the Berg Balance Scale. *J Phys Ther Sci* 2003; 15(2): 93–97.
15. Thomas JI and Lane JV. A pilot study to explore the predictive validity of 4 measures of falls risk in frail elderly patients. *Arch Phys Med Rehabil* 2005; 86(8): 1636–1640.
16. Burns AS, Delparte JJ, Patrick M, et al. The reproducibility and convergent validity of the walking index for spinal cord injury (WISCI) in chronic spinal cord injury. *Neurorehabil Neural Repair* 2011; 25(2): 149–157.
17. Mann GC, Whitney SL, Redfern MS, et al. Functional reach and single leg stance in patients with peripheral vestibular disorders. *J Vestib Res* 1996; 6(5): 343–353.
18. 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(1): 19–24.
19. Lynch SM, Leahy P and Barker SP. Reliability of measurements obtained with a modified functional reach test in subjects with spinal cord injury. *Phys Ther* 1998; 78(2): 128–133.
20. Hofheinz M and Schusterschitz C. Dual task interference in estimating the risk of falls and measuring change: a comparative, psychometric study of four measurements. *Clin Rehabil* 2010; 24(9): 831–842.
21. Bohannon RW. Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. *Age Ageing* 1997; 26(1): 15–19.
22. Bowden MG and Behrman AL. Step Activity Monitor: accuracy and test-retest reliability in persons with incomplete spinal cord injury. *J Rehabil Res Dev* 2007; 44(3): 355–362.
23. Wolf SL, Catlin PA, Gage K, et al. Establishing the reliability and validity of measurements of walking time using

- the Emory Functional Ambulation Profile. *Phys Ther* 1999; 79(12): 1122–1133.
24. Tyson S and Connell L. The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. *Clin Rehabil* 2009; 23(11): 1018–1033.
 25. Perera S, Mody SH, Woodman RC, et al. Meaningful change and responsiveness in common physical performance measures in older adults. *J Am Geriatr Soc* 2006; 54(5): 743–749.
 26. Mathiowetz V, Volland G, Kashman N, et al. Adult norms for the Box and Block Test of manual dexterity. *Am J Occup Ther* 1985; 39(6): 386–391.
 27. Oxford Grice K, Vogel KA, Le V, et al. Adult norms for a commercially available Nine Hole Peg Test for finger dexterity. *Am J Occup Ther* 2003; 57(5): 570–573.
 28. Erasmus LP, Sarno S, Albrecht H, et al. Measurement of ataxic symptoms with a graphic tablet: standard values in controls and validity in Multiple Sclerosis patients. *J Neurosci Methods* 2001; 108(1): 25–37.
 29. Stevenson DA, Allen S, Tidyman WE, et al. Peripheral muscle weakness in RASopathies. *Muscle Nerve* 2012; 46(3): 394–399.
 30. Shoukri MM, Asyali MH and Donner A. Sample size requirements for the design of reliability study: review and new results. *Stat Methods Med Res* 2004; 13: 251–271.