Psychometric properties of a core set of measures of balance for people with cerebellar ataxia secondary to multiple sclerosis

Abstract:

Objective: To examine the reliability, validity and interpretability of four clinical measures in assessing the severity of balance dysfunction among people with cerebellar ataxia (CA) secondary to multiple sclerosis (MS).

Design: Cross sectional observation study.

Setting: Data collected across four outpatient clinics in New Zealand and United States of America.

Participants: Sixty consecutive participants with CA secondary to MS.

Main outcome measures: Balance was assessed and video-recorded using the Berg Balance Scale (BBS), Timed Up and Go (TUG) test, the posture and gait sub-component of the International Co-operative Ataxia Rating Scale (PG-ICARS) and gait, stance and sit sub-components of the Scale for the Assessment and Rating of Ataxia (SARABal). The videos were later used to estimate reliability. The Barthel Index, Expanded Disability Status Scale (EDSS), ICARS and SARA were assessed and disease duration recorded.

Results: Reliability was good for all four measures (range between ICC 0.95 and 0.99). Internal consistency was moderate to good for all four measures (α range 0.72-0.94), moderate to good correlation between the measures of balance (ρS range 0.72-0.85) and poor to moderate correlation with disease severity (EDSS), functional independence (Barthel Index) and disease duration (ρS range -0.37 to 0.76). Minimal Detectable Change (MDC) was derived for BBS (3), PG-ICARS (2) and SARABal (2). Measures were able to discriminate between assistive walking device users and non-users.

Conclusions: All four measures showed good reliability and acceptable validity; however, owing to the item repetition in scoring of the PG-ICARS and moderate construct, criterion and convergent validity of the TUG, the BBS and SARABal are recommended for balance assessment in clinical practice for people with CA secondary to MS.

Key words: Reliability, Validity, Multiple sclerosis, Cerebellar ataxia

List of abbreviations:

BBS- The Berg Balance Scale

CA- cerebellar ataxia

EDSS- Expanded Disability Status Scale

ICARS- International Co-operative Ataxia Rating Scale

ICC- intra class correlation coefficient

MDC- Minimal Detectable Change

MS- Multiple sclerosis

PG-ICARS- The Posture and Gait sub-component of the International Co-operative Ataxia Rating Scale

SARA- Scale for the Assessment and Rating of Ataxia

SARABal- The gait, stance and sit sub-components of the Scale for the Assessment and Rating of Ataxia

TUG- The Timed Up and Go

α- Cronbach alpha

ρS- Spearman correlation coefficient

Damage, disease or dysfunction of the cerebellum or in the region of the cerebellar peduncles results in cerebellar ataxia (CA). Multiple sclerosis (MS) is one of the leading cause for CA and reports confirm 37% of people with MS have persistent cerebellar involvement.[1]. The prevalence of MS is high in New Zealand (NZ) (73.1/100000).[2] Problems with balance and gait are characteristic features of CA.[3] In the clinical setting, evaluating the influence of specific interventions on balance and function requires measurement tools that are both valid and reliable,

and have good clinical utility[4] in that, they are quick and easy to perform without the need for sophisticated equipment.[5]

Choosing an appropriate measure that captures severity of balance problems, monitors the progress of disease, and evaluates treatment effects on balance among people with CA is challenging as currently there is no recommended set of assessment tools. Cerebellar-specific measures are available but their clinical utility is limited as they are time-consuming.[6,7] There also appears to be a lack of awareness of these cerebellar-specific measures among clinicians, especially physiotherapists.[8] Studies in the past have recommended measures of balance in MS,[9-11] however they did not focus on CA.

The posture and gait sub-component of the International Co-operative Ataxia Rating Scale (PG-ICARS) is reported an appropriate measure in terms of best psychometric property estimates.[12] The gait, stance and sit sub-components of the Scale for the Assessment and Rating of Ataxia (SARABal), the Berg Balance Scale (BBS) and the Timed up and go (TUG) test are reported as suitable choice of clinicians for the assessment of balance in CA.[13] These four measures are quick and easy to perform, do not require sophisticated equipment, available free of cost and training of the assessor is not required. However formal validation is required prior to recommending them as core set of clinical measures. The purpose of this study is to determine the inter-rater, intra-rater reliability, internal consistency, criterion, convergent, construct, discriminant validity, and interpretability (Minimal Detectable Change- MDC) of four measures of balance. We hypothesise the measures of balance will have good reliability, moderate to good validity and be able to discriminate between assistive device users and non users. The findings of this psychometric analysis are expected to help strengthen recommendations for a core set of measures of balance in people with CA secondary to MS.

Methods

Participants

Sixty participants with CA secondary to MS were recruited from Dunedin Public Hospital, (n=36), Southland Hospital, Invercargill (n=6), Dustan Hospital, Clyde (n=2), NZ and Centres for Rehabilitation Services outpatient clinics, the University of Pittsburgh, Medical Center, Pennsylvania, USA (n=16). Participants gave written consent to be involved in this study.

Ethical approval was obtained from the University of Otago Human Ethics Committee, Dunedin, NZ (Ref: 13/041) and the University of Pittsburgh Biomedical Institutional Review Board (IRB), USA (Ref: PRO13080051). Detailed information on the methodology is published elsewhere.[14] In summary, the investigators assessed all participants on one occasion that lasted for approximately 60 minutes. Included participants had a definite diagnosis of MS presenting with at least one clinical cerebellar symptom (gait ataxia, limb ataxia, dysarthria or nystagmus), aged between 18 and 65 and were able to walk with or without assistive walking device for 10 metres. Those with severe visual impairment, Expanded Disability Status Scale (EDSS) score of > 6.5 and participants who did not give permission for the research team to access their medical records were excluded. To optimise accurate validity estimates, care was taken to standardise the assessment venues with regards to room dimension, lighting, equipment used, and texture of the testing surface. Participants balance performance was assessed using the four measures of

balance (BBS, TUG, PG-ICARS and SARABal), and was simultaneously video recorded. Repeat assessments to estimate reliability were completed through observation of the video-recording. Video-recording was performed using a wide angle digital video recorder in order to enhance a wide angle capture. In addition, the disease duration was recorded and the Barthel Index, the EDSS, the International Co-operative Ataxia Rating Scale (ICARS) and the Scale for the assessment and Rating of Ataxia (SARA) were scored to derive the constructs of validity.

Gait and balance performance were scored three times. Time 1 was the 'live assessment', time 2 was 'video assessment 1' and time 3 was the second scoring of the video 'video assessment 2'. Live assessments were done by the primary investigator in NZ and the research assistant at the USA, a video recording was done during the assessment. After recording a test session, the video data were then transferred onto a DVD in NZ and onto a password protected memory stick in USA to enable data transfer between the study centres. Video assessment 1 was done by the same investigator by observing the video after 7 to 10 days. For video assessment 2, data from NZ was distributed among the three members (CS, LH and LC) of the research team and the data from the USA was assessed by the primary investigator. Since the second assessors' assessment was done looking at the video, the video assessment 2 was not done on the same day and did not seem necessary.

Measures

The BBS is a generic measure for the assessment of balance.[15] It is a five-point ordinal scale scored between 0 and 4 for each task and has 14 tasks to be tested. The highest total score a participant could obtain is 56. Higher the score better the balance. The BBS has good

reliability[9,11] and acceptable validity in assessing balance among people with MS.[10] The TUG is a measure of dynamic stability of the individual that can predict the risk of falls. This timed measure records the time taken to arise from a chair, walk 3 meters, turn 180 degrees and walk back to the seat as fast as possible.[16] The longer the time taken to complete the TUG, poorer the balance and higher the persons' risk of falling. The TUG has good reliability,[11] and acceptable validity in assessing balance in MS.[10] The SARA is an ataxia rating measure used among different health conditions resulting in CA.[7] This measure is scored across eight items among which the gait, sitting and standing sub-components are relevant to balance assessment. The scale is scored out of 40. The balance sub-components (gait, stance and sit) are scored out of 18 and called SARABal. The scoring of the eight sub-components does not have equal weightage, scores range between 8 for 'gait' sub-component and 4 for 'heel-shin glide'. The higher the score, the greater the severity of ataxia. The SARA has been tested for psychometric properties among genetic and acquired forms of cerebellar health conditions and reported to be reliable[7,17,18] and valid.[19,20] The ICARS is a measure of ataxia severity.[6] Though comprehensive in assessing ataxia severity, this scale has been criticised for the time required to complete, taking over 20 minutes.[7] The ICARS has 19 items which are categorised as (i) posture and gait disturbances; (ii) kinetic function; (iii) speech disorders; and (iv) oculomotor disorders. The full scale is scored out of 100. The posture and gait sub-component is relevant to balance assessment and is scored out of 34. Similar to the SARA, scoring across each sub-component does not have equal weightage and ranges between 6 for 'oculomotor disorders' and 52 for 'kinetic score'. A high score on ICARS denotes severe ataxia. The ICARS has excellent reliability,[7] adequate validity[21,22] and good responsiveness.[23] The Barthel Index measures the performance of activities of daily living (ADL).[24] This scale has been commonly used for the functional assessment for people with

musculoskeletal and neuromuscular disorders.[25] The performance of ten items relating to ADL and mobility are scored between 0 and 15, scoring is not even across the items. The lower the score obtained, the poorer the functional independence. The scale has a maximum of 100 and a minimum of 0. The scale has moderate to excellent reliability,[26] and validity[27] in assessing ADL among people with MS, stroke, and traumatic brain injury. The EDSS is a measure to rate disability due to MS.[28] Eight functional systems that are scored using the Functional System Score (FSS) and based on the FSS scores, the EDSS is scored between 0 to 10. The higher the score, the greater the disability due to MS. The EDSS demonstrates good reliability and validity for rating disability among people with MS.[29]

Statistical analysis

Data analysis was done using Statistical Package for the Social Sciences (SPSS) statistics version 20. Intraclass correlation coefficient (ICC) with one-way random model and absolute agreement was used to determine intra-rater and inter-rater reliability. Cronbach's alpha (α) was used to estimate internal consistency. There are no universal guidelines for interpreting the ICC and α, in general an ICC over 0.75 is indicative of good reliability and the higher the value towards 1.00, the greater the reliability.[30] In this study the ICC and α were interpreted as: <0.50 as weak, those between 0.5 and 0.79 as moderate, and those > 0.8 as good. The measures of balance were correlated between each other for criterion validity and with the EDSS, disease duration and Barthel Index for construct validity. The discriminant validity was determined by assessing the ability of the balance measures to differentiate between two known groups within the study sample. The participants were sub-divided into assistive (walking) device users and non-users.

The group difference across the four measures was observed using the Mann Whitney U test. The cut-off score, sensitivity, and specificity of the measures of balance to predict the use of assistive device were identified by constructing a receiver operating characteristics (ROC) curve. Further, to determine which measure had a best predictive ability, the 'Area Under the Curve' (AUC) was used. Spearman correlation coefficient (ρS), bivariate analysis of a non-parametric sample was used to establish criterion validity and hypothesis testing (convergent and construct validity). Interpretation of validity estimates were similar to that of reliability. The MDC was estimated using a data driven method proposed by Wyrwich et.al[31] using the Standard Error of Measurement (SEM).

Results

Demographic characteristics of the included participants are reported in Table 1. Fifty of the 60 participants were reassessed using video-recording to estimate inter-rater and intra-rater reliability. Thirty-eight participants (63%) did not use an assistive device for walking and the remaining 22 (37%) either used rollator (rolling walker), (n=5), one quadripod (four-legged cane) (n=2), two elbow crutches (n=2), one elbow crutch (n=4), or one cane (n=7).

Insert table 1 about here.

Reliability

Table 2 reports the reliability estimates and MDC. The intra and inter-rater reliability were good with ICCs above 0.9 for all four measures. The BBS and PG-ICARS had good internal consistency with α values of 0.94 and 0.87, respectively. The SARABal had moderate internal consistency as indicated by a α of 0.72. With regards to the individual test items of the measures of balance, deletion of item 3 (the sit item of the SARABal) increased the internal consistency of SARABal from 0.72 to 0.87. Similarly, deletion of item 7 of the PG of ICARS (the quality of the sitting position) increased the measure's internal consistency from 0.87 to 0.9.

Insert table 2 about here

Validity

Table 3 illustrates the estimated validity of the four balance measures. With the exception of the TUG, all measures demonstrated good correlation with each other as indicated by Spearman's correlation coefficients (ρS) ranging between -0.89 and 0.92. The ataxia rating scales correlated moderate to good with all measures of balance (ρS between -0.75 and 0.83). The TUG had moderate criterion and convergent validity and the other three measures were good as indicated by a high correlation co-efficient at a significance level of $p< 0.01$. Disease severity (EDSS), disease duration and functional independence (Barthel Index) correlated weak to moderate (ρS between -0.39 and 0.58).

All four measures demonstrated significant ($p<0.01$) score differences across assistive device users and non-users indicating strong discriminant validity (Table 4). The ability of all balance measures to correctly categorise participants as users and non-users of an assistive device were

good as indicated by the AUC of more than 0.92. The sensitivity to identify assistive device user was 90% and the specificity ranged between 81% and 100%.

Insert table 3 and 4 about here

Discussion

This study demonstrated good reliability for all four balance measures, acceptable validity for the BBS, PG-ICARS and SARABal, and moderate validity for the TUG for participants with CA secondary to MS. The good inter and intra-rater reliability demonstrated for the four measures is consistent with that reported previously.[7,9,11,17-20] The BBS and TUG had the highest reliability score followed by the SARABal and thereafter the PG-ICARS.

The use of video-recording for estimating reliability had both advantages and disadvantages. The advantages being: variability in scores due to worsening of the condition by the follow up was eliminated, loss of data from participants due to inability to attend follow up assessment was removed, and it is a time and cost efficient method of reliability estimation both for the researcher and participant. However, video-recording was disadvantageous due to its inability to capture the variation in score that occurs during the follow up assessment that is not related to the disease progress. Considering variation in the score with follow up assessments is an important factor for reliability estimates, and we recommend future studies investigate this

parameter. Our results most likely over estimate reliability as there was no participant variation between trials.

Of the three measures, the BBS demonstrated the best internal consistency among test items (α=0.94) (internal consistency assessment was not applicable for the TUG as it is a single item measure). Among the cerebellar-specific measures of balance, internal consistency was good; however, was less than the BBS. The reduction in the Cronbach's α value for the cerebellar-specific measures is attributed to the sitting items of the measures (Item 7 of PG of ICARS and sub-component 3 of SARA). Likewise the intra and inter-rater reliability for these single sitting items were poor to moderate (ICC range 0.39-0.58). Clinical judgment based on observation to quantify the magnitude of postural sway while sitting may be considered subjective accounting for reduced reproducibility. Among the assistive device non-users group, 93% received the optimal (best) score. The inconsistency in scoring was more of an issue among the assistive device users who are likely to have greater sitting postural sway. Therefore, caution is recommended when assessing postural sway in sitting using the cerebellar-specific measures. To reduce subjectivity, the assessment may be carried out by positioning the participant against (but not touching) a wall-mounted ruler or a postural sway grid in an attempt to record more accurate postural sway measurements.

Disease duration demonstrated a weak to moderate correlation with the four balance measures. The heterogeneity with regards to participant's disease course might account for this moderate correlation. The study sample included individuals with a disease duration ranging between two and 26 years, and there was a mixture of sub-types of MS disease course that included relapsing remitting, secondary progressive, primary progressive, and progressive relapsing. Each of these sub-types has its own unique progression and symptomatology presentation.[32] Disease severity

may not be proportional to the duration of disease among the sub-types and therefore a moderate correlation is acceptable.

Moderate correlation between the Barthel Index and the four measures of balance are consistent with previous findings that tested the construct validity of the ICARS (γ= -0.70)[7] and SARA (γ= -0.63)[17] among participants with spino-cerebellar ataxia. Since cerebellar-specific measures for functional independence were not available, the Barthel Index was used in this study. Twenty-one of the 60 participants (35%) included in this study scored maximum (100) on this scale. Considering the chronicity of the condition and the limits of the ceiling of the scale, the Barthel Index may not be the best choice in assessing functional independence among ambulant, community dwelling participants with MS and CA. Instead, disease specific measures for MS such as the Functional Assessment of Multiple Sclerosis (FAMS)[33] or the Functional Independence Measure (FIM),[34] as highly recommended tools[35] may be considered for future studies.

Correlation between the sub-components of balance (PG-ICARS and SARABal) and the full ataxia rating scales (ICARS and SARA) were good and in line with previous studies.[19,22,23,36] Unlike previous studies recommending balance scales for people with MS[9-11] the current study highlights the correlation between ataxia rating scales and generic measures of balance (BBS and TUG), demonstrating their usability among people with CA secondary to MS. The BBS had high correlation (-0.75 and -0.79) with the ataxia rating scales, but the TUG only correlated moderately (0.54 and 0.58). To reduce the influence of fatigue on the performance of the TUG, the order of assessment was reversed among half of the participants. However, whether tested in reverse order or not, in both cases the TUG fell in the middle of the order of testing. As fatigue is a major issue for many participants with MS,[37] future studies should consider fatigue when

multiple assessments are utilized. Though fatigue may have influenced the performance of the TUG, moderate correlation obtained for ataxia rating scales and other measures of balance deems reconsideration of including this measure in the core set. The TUG is a test of dynamic balance based on gait speed and functional mobility.[16] The moderate correlation observed in this study is an indication that gait speed and functional mobility may not necessarily correlate with balance deficits resulting from ataxia, therefore the TUG may not be considered one of the core set of balance measures in CA. On removing the TUG from the core set, the balance assessment is deprived of estimating timed walking ability. However, the SARABal and PG-ICARS have items to assess walking ability.

Among the assistive device non-users, item 1 and item 3 of the BBS scored maximally. Therefore, it is suggested that these items may be omitted and full score given to those individuals who are able to walk without assistive devices in order to save time and conserve energy. With regards to the walking item of the cerebellar-specific measures of balance, PG-ICARS lacks a smooth transition for grading of severity of walking between stage 0 and stage 4. Stage 0 is graded 'normal walking' and stage 1 is interpreted as 'almost normal naturally, but unable to walk with feet in tandem position'. In our observation, most participants were able to walk in tandem; however they had difficulty in completing the task. An additional grade between stage 0 and stage 1 may provide a more refined grading of walking ability. In addition, stage 2 (walking without support, but with a clearly abnormal and irregular gait) and stage 3 (walking without support but with considerable staggering, difficulties in half turn) indicated redundancy as it was difficult to differentiate between the two grades. In this observation grading those who do not use assistive walking devices as either stage 2 or stage 3 was not clear as they had a mixture of presentation explained by these stages and they appeared to be arguably similar. The

gait sub-component of the SARA was found to have a smooth transition and clear demarcation between the stages. It would therefore appear that the SARABal is more useful to PG-ICARS in the assessment of balance in participants with CA, which reiterates a previous observation.[38] All items of the PG-ICARS except the 'Spread of feet' can be estimated using the BBS and SARABal.

Nearly 10% of participants obtained best scores of balance on all four measures (6 on BBS and TUG, 5 on SARABal and 1 on PG-ICARS) indicating no balance deficits. They were recruited due to the presence of other cerebellar signs such as limb ataxia and/or nystagmus and/or scanning speech. Though the clinical measures indicate no balance deficits it can be argued that the measures were not sensitive enough to pick minute changes. In clinical practice these best scores of balance may be used to document the baseline readings and revisited to track the disease progress.

This is the first study to report MDC and cut-off values for measures of balance to differentiate between assistive device users and non-users in the population of interest. Of the four measures, the SARABal had the best predictive cut-off score of >5 with 90% sensitivity and 100% specificity. The derived MDC carries meaningful information about the expected change in score that may be a result of a true change in health status following an intervention.

Strengths and limitations:

The heterogeneity of participants in terms of disease course, and homogeneity in terms of the focused group of CA secondary to MS added strength to the study. The heterogeneity among the

sample enables a wider generalizability of the findings among participants with MS. The homogeneity enhances the appropriateness of the tested samples to examine the study objectives.

The use of video analysis limits the scope of the reliability. Fatigue may have influenced the results. Although it was minimized by changing the order of assessment among participants and allowing rest periods when required, this factor could not be completely eliminated. Randomizing the order of assessment among participants may have yielded greater accuracy. In addition, the time of day that the assessment took place was not standardized.[39] MS affects multiple systems, and although participants were recruited based on the presence of ataxia, it is very likely that other systems may have contributed to the balance dysfunction. Given the heterogeneity of the samples and lack of responsiveness estimation we hesitate to make a strong recommendation on the core set of measures of balance for CA.

Conclusion

We recommend the BBS and SARABal for the assessment of balance in people with CA secondary to MS. The PG-ICARS involves more time spent on testing repeated items and, the TUG demonstrates moderate construct, convergent and criterion validity estimates making them unsuitable. Future studies are warranted to examine the responsiveness of this core set of measures to strengthen this recommendation.

Reference

1. Compston A. McAlpine's Multiple Sclerosis. In: Ian McDonald A, Compston A. The symptoms and signs of multiple sclerosis. Philadelphia, PA: Elsevier Inc.; 2006..
2. Alla S, Mason DF. Multiple sclerosis in New Zealand. Journal of Clinical Neuroscience. 2013;21(8):1288-91.
3. Morton SM, Bastian AJ. Cerebellar control of balance and locomotion. The Neuroscientist. 2004;10:247-59.
4. Nelson EC, Berwick DM. The measurement of health status in clinical practice. Med Care. 1989; 27:S77-S90.
5. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess. 1998;2: 1-74.
6. Trouillas P, Takayanagi T, Hallett M, Currier R, Subramony S, Wessel K, Bryer A, Diener HC, Massaquoi S, Gomez CM, Coutinho P, Ben Hamida M Campanella G, Filla A, Schut L, Timann D, Honnorat J, Nighoghossian N. International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. J Neurol Sci. 321 1997;145:205-11.
7. Schmitz-Hübsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, Giunti P, Globas C, Infante J, Kang JS, Kremer B, Mariotti C, Melegh B, Pandolfo M, Rakowicz M, Ribai P, Rola R, Schols L, Szymanski S, vande Warrenburg BP, Durr A, Klockgether T, Fancellu R. Scale for the assessment and rating of ataxia Development of a new clinical scale. Neurology. 2006;66:1717-20.

8. Fonteyn EM, Keus SH, Verstappen CC, van de Warrenburg BP. Physiotherapy in degenerative cerebellar ataxias: utilisation, patient satisfaction, and professional expertise. Cerebellum. 2013;12:841-7.
9. Cattaneo D, Johanna J, Stefania R. Reliability of four scales on balance disorders in persons with multiple sclerosis. Disabil Rehabil. 2007;29:1920-5.
10. Cattaneo D, Regola A, Matteo M. Validity of six balance disorders scales in persons with multiple sclerosis. Disability & Rehabilitation. 2006;28(12):789-95.
11. Learmonth YC, Paul L, McFadyen AK, Mattison P, Miller L. Reliability and clinical significance of mobility and balance assessments in multiple sclerosis. Int J Rehabil Res. 2012;35:69-74.
12. Winser SJ, Smith CM, Hale LA, Claydon LS, Whitney SL, Mehta P. Systematic review of the psychometric properties of balance measures for cerebellar ataxia. Clin Rehabil. 2015;29:69-79.
13. Winser SJ, Smith C, Hale LA, Claydon LS, Whitney SL. Balance outcome measures in cerebellar ataxia: a Delphi survey. Disabil Rehabil. 2015;37:165-70.
14. Winser SJ, Smith CM, Hale LA, Claydon LS, Whitney SL. Balance Assessment in Multiple Sclerosis and Cerebellar Ataxia: Rationale, Protocol and Demographic Data. Phys Med Rehabil Int. 2014;1:6.
15. Berg K, Wood-Dauphinee S, Williams JI. The Balance Scale: reliability assessment with elderly residents and patients with an acute stroke. Scand J Rehabil Med. 1995;27(1):27-36.
16. Podsiadlo D, Richardson S. The timed" Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142-8.
17. Weyer A, Abele M, Schmitz-Hübsch T, Schoch B, Frings M, Timmann D, Klockgether T. Reliability and validity of the scale for the assessment and rating of ataxia: a study in

64 ataxia patients. Mov Disord. 2007;22:1633-7.

18. Yabe I, Matsushima M, Soma H, Basri R, Sasaki H. Usefulness of the Scale for Assessment and Rating of Ataxia (SARA). J Neurol Sci. 2008;266:164-6.

19. Bürk K, Mälzig U, Wolf S, Heck S, Dimitriadis K, Schmitz-Hübsch T, Hering S, Lindig TM, Haug V, Timmann D, Degen I, Kruse B, Dorr JM, Ratza S, Ivo A, Schols L, Boesch S, Klockgether T, Klopstock T, Scholz JB. Comparison of three clinical rating scales in Friedreich ataxia (FRDA). Mov Disord. 2009;24:1779-84.

20. Lee Yc, Liao Yc, Wang Ps, Lee IH, Lin Kp, Soong Bw. Comparison of cerebellar ataxias: A three-year prospective longitudinal assessment. Mov Disord. 2011;26:2081-7.

21. Schoch B, Regel JP, Frings M, Gerwig M, Maschke M, Neuhäuser M, et al. Reliability and validity of ICARS in focal cerebellar lesions. Movement Disorders. 2007;22(15):2162-9.

22. Cano SJ, Hobart JC, Hart PE, Korlipara L, Schapira AH, Cooper JM. International Cooperative Ataxia Rating Scale (ICARS): appropriate for studies of Friedreich's ataxia? Mov Disord. 2005;20:1585-91.

23. Morton SM, Tseng Ya-Weng, Zackowski Ka 381 thleen M, Daline Jaclyn R, J. Bastian A. Longitudinal tracking of gait and balance impairments in cerebellar disease. Mov Disord. 2010;25:1944-52..

24. Mahoney FI. Functional evaluation: the Barthel index. Maryland state Med J. 1965;14:61-5.

25. Carole L, Shaw K. The (Original) Barthel Index of ADLs. Geriatric Function. 2006;17(21):8.

26. Hsueh I-P, Lee M-M, Hsieh C-L. Psychometric characteristics of the Barthel activities of daily living index in stroke patients. J Formos Med Assoc. 2001;100(8):526-32.

27. Hobart J, Thompson A. The five item Barthel index. J Neurol Neurosurg Psychiatry. 2001;71(2):225-30.

413 28. Kurtzke JF. Rating neurologic impairment in multiple sclerosis an expanded disability status scale
414 (EDSS). Neurology. 1983;33(11):1444-.
415 29. Hobart J, Freeman J, Alan T. Kurtzke scales revisited: the application of psychometric methods to
416 clinical intuition. Brain. 2000;123(5):1027-40.
417 30. Portney LG, Watkins M. Foundations of Clinical Research: Applications to Practice. In:
418 Statistical measures of reliabillity. 3rd ed. New Jersey, USA: Pearson Education; 2009.
419 31. Wyrwich KW, Tierney WM, Wolinsky FD. 340 Further evidence supporting an SEM-based
420 criterion for identifying meaningful intra-individual changes in health-related quality of
421 life. J Clin Epidem. 1999;52:861-73.
422 32. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis results of an
423 international survey. Neurology. 1996;46:907-11.
424 33. Cella D, Dineen K, Arnason B, Reder A, Webster K, Karabatsos G, Chang C, Lloyd S,
425 Steward J, Stefoski D. Validation of the functional assessment of multiple sclerosis
426 quality of life instrument. Neurology. 1996;47:129-39..
427 34. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence
428 measure: a new tool for rehabilitation. Adv Clin Rehabil. 1987;1:6-18..
429 35. Kirsten P, Diane A, Amy M Y, Gail W, Susan B, Evan ,et al. Multiple Sclerosis Outcome Measures
430 Taskforce Neurology Section OutcomeMeasures Recommendations [Internet]. 2010:[162-72
431 pp.]. URL:
432 http://www.neuropt.org/docs/ms-edge-documents/final-ms-edge-document.pdf?sfvrsn=4.
433 36. D'Abreu A, Franca M, Lopes-Cendes I, Cendes F. The international cooperative ataxia
434 rating scale in Machado-Joseph disease. Comparison with the unified multiple system
435 atrophy rating scale. Mov Disord. 2007;22:1976-9.
436 37. Aronson K, Goldenberg E, Cleghorn G. Sociodemographic characteristics and health

status of persons with multiple sclerosis and their caregivers. MS Management. 1996;3:5-15.

38. Schmitz-Hübsch T, Tezenas du Montcel S, Baliko L, Boesch S, Bonato S, Fancellu R, Guiti P, Globas C, Kang JS, Kremer B, Mariotti C, Melegh B, Rakowicz M, Rola R, Romano S, Schols L, Szymanski S, van de Warrenberg BP, Zdzienicka E, Durr A, Klockgether T. Reliability and validity of the International Cooperative Ataxia Rating Scale: a study in 156 spinocerebellar ataxia patients. Mov Disord. 2006;21(5):699-704.

39. Freal J, Kraft G, Coryell J. Symptomatic fatigue in multiple sclerosis. Arch Phys Med Rehabil 1984;65:135-8.