BRIEF COMMUNICATION

Test–retest reliability of the Friedreich's ataxia rating scale

Christian Rummey[1] (iD), Theresa A. Zesiewicz[2], Santiago Perez-Lloret[3,4], Jennifer M. Farmer[5], Massimo Pandolfo[6] & David R. Lynch[7] (iD)

[1]Clinical Data Science GmbH, Basel, Switzerland
[2]University of South Florida Ataxia Research Center, Tampa, Florida
[3]Institute of Cardiological Research, University of Buenos Aires, National Research Council (ININCA-UBA-CONICET), Marcelo T. de Alvear 2270, Buenos Aires, C1122, Argentina
[4]Department of Physiology, School of Medicine, University of Buenos Aires (UBA), Buenos Aires, Argentina
[5]Friedreich's Ataxia Research Alliance, Downingtown, Pennsylvania
[6]Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium
[7]Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

Correspondence

David R. Lynch, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. Tel: 2155902242; Fax: 2155903779; E-mail: lynchd@mail.med.upenn.edu

Abstract

The modified Friedreich Ataxia Rating Scale (mFARS) is a disease specific, exam-based neurological rating scale commonly used as a outcome measure in clinical trials. While extensive clinimetric testing indicates it's validity in measuring disease progression, formal test–retest reliability was lacking. To fill this gap, we acquired results from screening and baseline visits of several large clinical trials and calculated intraclass correlation coefficients, coefficients of variance, standard error, and the minimally detectable changes. This study demonstrated excellent test–retest reliability of the mFARS, and it's upright stability subscore.

Introduction

Friedreich's ataxia (FRDA) is a progressive, neurodegenerative disease that affects children and young adults with gait and limb ataxia, dysarthria, loss of reflexes, proprioceptive dysfunction, and muscle weakness, as well as non-neurological features of cardiomyopathy, diabetes, and scoliosis.[1] FRDA is caused by a GAA repeat expansion in the *FXN* gene,[2] leading to a marked reduction in frataxin, a mitochondrial protein that plays a vital role in energy production and iron homeostasis.

While FRDA currently has no treatment or cure, novel prospective therapeutic strategies are actively being studied in clinical trials, requiring the use of validated and clinically meaningful outcome measures. One commonly used disease specific measure Friedreich's Ataxia Rating Scale (FARS), an exam-based neurological measure. The FARS was initially introduced in 2005.[3] In addition to a standardized neurological exam, its initial validation included in parallel a patient reported activities of daily living scale (ADL), a functional disability staging (FDS), and time-based performance measures. The initial version

of the neurological exam component of the FARS was comprised of five subscores: bulbar (11 points), upper limbs (36), lower limbs (16), peripheral nervous system (26), and upright stability (28), summed for a total maximum score of 117 points.[3] Later, two-timed stance items without visual aid (standing with feet apart, eyes closed and standing with feet together, eyes closed) were added to the upright stability section, resulting in the FARS-neuro (FARSn) with a total of 125 points.[4] This version has been used in multiple clinical trials as well as a prospective, longitudinal natural history, and clinical outcome measures study (FA-COMS, NCT03090789) that started in 2003.[4] The FARSn has undergone further refinement as a clinical outcome measure, omitting items that do not directly assess functional abilities. Specifically, the peripheral nervous system subscore and two bulbar components were removed, leading to the modified FARS, or "mFARS" score (maximum total of 93 points[5]). Psychometric properties of the FARSn and the mFARS were recently summarized,[5] but there remained a lack of published information specifically on test–retest reliability in these scales. In this study, we evaluated the test–retest

reliability in the mFARS and the FARSn scales using data from recent clinical research in FRDA.

Material and Methods

Data source/ clinical studies

FARSn and mFARS scores were acquired from the FRDA integrated clinical database (FA-ICD) maintained by the Critical Path Institute (c-PATH). As of May 2020, these data included screening and baseline visits from the following clinical trials in FRDA: MICONOS (NCT00905268), IONIA[6] (NCT00537680), LA-29[7] (NCT00530127), and EPI-743[8] (NCT01728064). For the EPI-743 study and LA-29, the FA-ICD only included data from the placebo arm. Additional data were provided directly by sponsors: Reata Pharmaceuticals (MOXIe, NCT02255435) and Chiesi Canada (additional predose data from LA-29, NCT00530127). The scales analysed in this study were the mFARS, FARSn, and FARSn-117. For the test–retest analyses, we used all available data, as long as full exam results from two visits (screening and baseline) were available.

Statistical analysis/ Test–Retest-Reliability

Testing took place during screening and baseline visits of clinical trials, with the time between those visits generally not considered to lead to a relevant clinical decline of the condition. This was assessed by comparing means between the two visits using paired t-test and calculating the mean in-between change. In this specific situation, every patient was administered two exams by the same rater, allowing for assessment of the consistency between the ratings, but not between rater-variability. We, therefore, used the one-way analysis of variance ANOVA-based intraclass correlation coefficient (ICC2,1 by standard nomenclature [9,10]) to assess the test–retest reliability of the FARS scales. Also, we calculated coefficients of variance (CV), the standard error of the measurement (SEM), and the minimally detectable change at 90% confidence (MDC), reflecting the magnitude of change necessary in an individual to ensure that a change is not the result of random variation or measurement error. Group level MDC is usually of less interest due to sample size calculations in clinical trials.[11] The MDC was then divided by the range of the individual measure to yield the relative amount of the random measurement error (MDC%). Bland-Altman plots were used to visualize the difference and mean score of each pair of measurements. All data derivation and analyses were conducted in R (www.r-project.org) utilizing the psych-package[12] for calculation of the ICCs.

Results

mFARS and FARSn test–retest data were available from 172 patients from the IONIA and MOXIE studies (Table 1). Only in these studies did the neurological exams include the two stance items performed with eyes closed,[5] necessary to calculate mFARS and FARSn. For the remaining studies (EPI-743, LA-29, and MICONOS), only FARSn-117 data were available. In addition, the EPI-743 study only included one predose visit. Therefore, (excluding EPI-743) data from a total of 405 patients could be evaluated for the test–retest analysis of FARSn-117. Visit timing indicates values foreseen in study protocols as exact dates were only on hand for the LA-29 study (median time between visits was 42 d, range 26 d to 97d, Table 1).

ICC values above 0.90 usually indicate excellent test–retest reliability;[13] the present results overall demonstrate this for the mFARS scale with an ICC of 0.95 (95%CI 0.94–0.96). Mean values for both visits were 42.3 (SD 10.8) and 42.3 (SD 10.7), and the mean change was −0.1 (SD 3.3), showing that no relevant change in the scores has occurred between the visits. ICC and other corresponding values for FARSn and the FARSn-117 confirm these results (Table 2). All P-values from paired t-tests were nonsignificant (i.e. larger than 0.05). The minimally detectable change for mFARS was 5.51 points, which corresponds to a percentage MDC of 6%, likewise an excellent value.[11,13]

The mFARS subscores and, in particular, upright stability showed similar preferable results with ICCs of 0.95 (upright stability), 0.89 (upper limbs), 0.83 (lower limbs), and 0.73 (bulbar). Among the individual items in the scale, this ensures that the essential stance items (E2A,

Table 1. Summary of available data at screening and baseline visits.

Study	Scales Evaluated	Time from Screening to Baseline	N[1]	Comments
MOXIE	mFARS FARSn FARSn-117	up to 1 month	103	
IONIA	mFARS FARSn FARSn-117	within 8 weeks	69	
EPI-743	FARSn-117	–	20	Study included only 1 predose visit.
MICONOS	FARSn-117	within 8 weeks	161	
LA-29	FARSn-117	less than 3 months	72	

[1]Participants with full scale data from two visits available.

Table 2. Test–Retest Reliability indices of mFARS scale.

Parameter	max. score	n	Mean (SD) D1	Mean (SD) D2	d	P*	ICC	95%CI	CV	SEM	MDC	MDC (%)
mFARS	93	172	42.3 (10.8)	42.3 (10.7)	−0.1 (3.3)	0.956	0.95	(0.94–0.96)	5.57	2.36	5.51	6
FARSn	125	172	52.2 (12.2)	52.2 (12)	0 (4.3)	0.995	0.94	(0.92–0.95)	5.75	3.02	7.05	6
FARSn-117	117	405	48.1 (18.1)	48.1 (17.8)	0 (4.8)	0.969	0.96	(0.96–0.97)	6.99	3.39	7.91	7
Subscores												
Bulbar	5	172	0.7 (0.6)	0.8 (0.6)	0.1 (0.5)	0.251	0.73	(0.65–0.79)	42.89	0.33	0.77	15
Upper Limbs	36	172	12.6 (4.7)	12.5 (4.7)	−0.1 (2.2)	0.806	0.89	(0.85–0.92)	12.57	1.58	3.68	10
Lower Limbs	16	172	7.1 (2.6)	7.2 (2.6)	0.1 (1.5)	0.790	0.83	(0.78–0.87)	15.07	1.07	2.49	16
Upright Stability	36	172	21.9 (6.3)	21.8 (6.4)	−0.1 (2)	0.892	0.95	(0.93–0.96)	6.48	1.40	3.26	9
Items (mFARS only)												
A4 cough	2	172	0.1 (0.3)	0.1 (0.3)	0 (0.2)	0.705	0.75	(0.68–0.81)	129.30	0.14	0.33	16
A4 speech	3	172	0.6 (0.5)	0.7 (0.5)	0.1 (0.4)	0.226	0.65	(0.56–0.73)	45.71	0.29	0.68	23
B1 finger to finger	6	172	1.5 (1)	1.6 (1.1)	0.1 (0.8)	0.439	0.72	(0.63–0.78)	36.97	0.56	1.30	22
B2 nose to finger	8	172	2.3 (1)	2.2 (1.1)	0 (0.8)	0.782	0.70	(0.61–0.77)	26.17	0.57	1.32	17
B3 dysmetria	8	172	2.8 (1.3)	2.7 (1.3)	−0.1 (1)	0.447	0.68	(0.59–0.75)	26.81	0.76	1.78	22
B4 rapid movements	6	172	2.9 (1.3)	3 (1.3)	0 (0.9)	0.837	0.76	(0.68–0.81)	21.99	0.66	1.55	26
B5 finger taps	8	172	3.1 (1.5)	3 (1.5)	−0.1 (1.1)	0.519	0.72	(0.64–0.79)	25.92	0.80	1.88	23
C1 heel shin slide	8	172	4.2 (1.4)	4.1 (1.4)	−0.1 (1.1)	0.677	0.72	(0.64–0.79)	18.19	0.76	1.78	22
C2 heel shin tap	8	172	2.9 (1.6)	3.1 (1.6)	0.1 (1)	0.419	0.80	(0.73–0.84)	24.09	0.73	1.71	21
E1 sitting position	4	172	0.7 (0.5)	0.8 (0.5)	0.1 (0.4)	0.362	0.64	(0.55–0.72)	41.88	0.32	0.75	19
E2A stance feet apart	4	172	0.6 (1.4)	0.6 (1.3)	0 (0.6)	0.989	0.90	(0.87–0.92)	71.62	0.43	1.00	25
E2B (eyes closed)	4	172	2.8 (1.7)	2.8 (1.7)	0 (0.7)	0.924	0.91	(0.88–0.93)	18.61	0.52	1.22	30
E3A stance feet together	4	172	1.9 (1.8)	1.9 (1.8)	0.1 (0.9)	0.797	0.87	(0.83–0.9)	34.55	0.65	1.52	38
E3B (eyes closed)	4	172	3.9 (0.4)	3.9 (0.4)	0 (0.2)	0.734	0.89	(0.85–0.92)	3.61	0.13	0.31	8
E4 tandem stance	4	172	3.6 (0.9)	3.4 (1.3)	−0.2 (1)	0.104	0.60	(0.49–0.69)	20.15	0.60	1.39	35
E5 stance dominant foot	4	172	3.9 (0.4)	3.9 (0.4)	0 (0.3)	0.772	0.82	(0.77–0.87)	4.65	0.18	0.41	10
E6 tandem walk	3	172	2.4 (0.7)	2.4 (0.7)	0 (0.5)	0.708	0.75	(0.67–0.81)	15.21	0.37	0.87	29
E7 gait	5	172	2 (1.2)	2 (1.2)	0 (0.5)	0.895	0.92	(0.9–0.94)	16.61	0.34	0.80	16

*p-value for paired t-test comparing the means between both visits

E2B, and E3A), as well as the gait item, have excellent to good ICCs.

A Bland-Altman plot for the assessments of mFARS (Fig. 1) shows no clear trends for less reliability at any mean scores. Outliers occurred both at low and at average overall scores.

Discussion

The present study shows that the mFARS and FARS exams have excellent test–retest properties as need for use in therapeutic trials and other clinical studies. The FARS was first introduced in 2005[3] as a disease-specific clinical rating instrument to capture functional abilities related to neurological aspects of FRDA. Further optimization and focus on functional, patient relevant items, together with psychometric analyses of the FARSn[5] has further refined this instrument to the "mFARS" score, which has been used both as primary and secondary endpoint in contemporary clinical trials of FRDA. Compared to the FARSn, the mFARS appears less prone particularly to floor effects, and shows a better dimensional structure, while retaining an adequate level of internal consistency.[5] While mild ceiling effects are retained, particularly the gait subscore of the mFARS was shown to captures well the progressive function loss associated with loss of ambulation in FRDA.[14] The results in the present work complement these features.

Validity of the FARS scales has been assessed in many observational studies, demonstrating its high correlation with age of onset, genetic burden of disease (GAA repeat length), and patient reported outcome measures such as the ADL scale.[4,15,16] It also correlates with other rating scales such as the Scale for the Assessment and Rating of Ataxia (SARA) score, a modified version of Barthel Index, and the Functional Independence measure.[17,18] The mFARS/FARS associates not only with walking speed, cadence, and stability indexes,[19] but also with physiopathological aspects of the disease, including iron clustering at the cerebellum,[20] atrophy in cerebellar peduncles,[21,22] degeneration of the spinal cord,[23] and *FXN* expression.[24] The FARSn and mFARS scales have shown adequate internal consistency, and inter-rater reliability.[3,5]
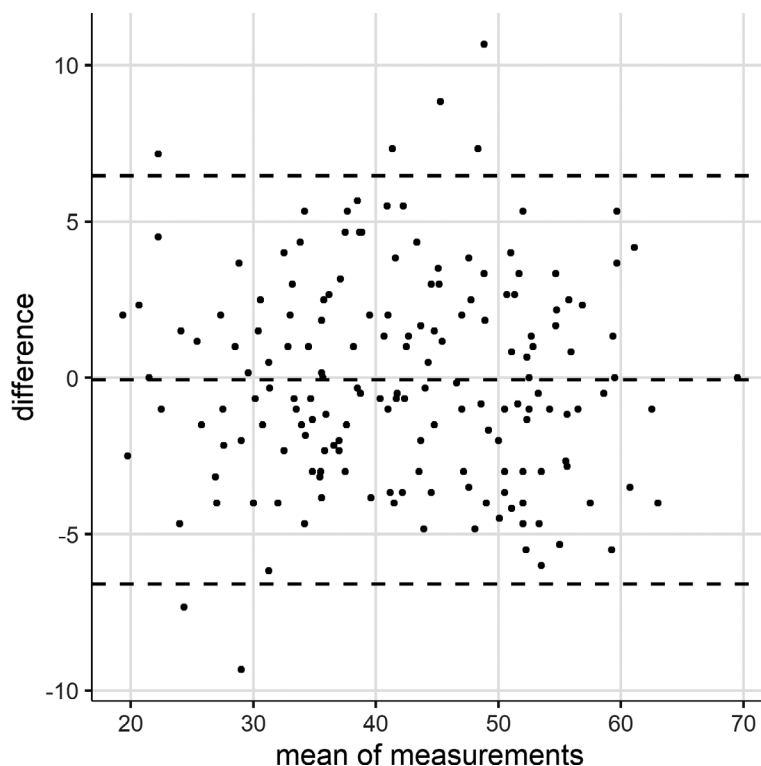
Figure 1. Bland–Altman Plot for differences between screening and baseline visits in the mFARS scale.

Test–retest reliability however requires short-term follow-up data, which requires nontrivial efforts to obtain. In this study, we filled this gap using data from recent clinical studies and successfully proved the test–retest reliability of the overall scale. This study demonstrated excellent test–retest reliability of the mFARS, the FARSn, and the FARSn-117 scores as well as the mFARS upright stability score. In isolation, upper - and lowerlimb subscores still showed good ICCs, while the bulbar subsection had a moderate ICC. A potential limitation of the current study is the lack of intraday retesting, although such testing could conceivably be associated with practice effects. In addition, the fatigability of FRDA patients could confound same day testing. Also, in specific studies (e.g. LA-29) the between-test interval was fairly long, but the impact is considered low, given the (relatively) slow progressive nature of the disease. Due to the inherent variability of all FRDA patients, future clinical trials will likely focus on targeted subpopulations, and further work hopefully will provide evidence that the excellent overall qualities of the mFARS scale apply also to these dedicated patient subgroups.

Acknowledgments

Author Contributions

TAZ, CR, and DRL contributed to conception and design of the study. TAZ, CR, DRL, SPL, JMF, and MP contributed to the acquisition and analysis of data and review of the manuscript. CR and TAZ drafted text, tables, and the figure.

Conflict of Interest

None declared.

References

1. Dürr A, Cossee M, Agid Y, et al. Clinical and genetic abnormalities in patients with Friedreich's ataxia. N Engl J Med 1996;335:1169–1175.
2. Campuzano V, Montermini L, Moltò MD, et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 1996;271:1423–1427.
3. Subramony SH, May W, Lynch D, et al. Measuring Friedreich ataxia: interrater reliability of a neurologic rating scale. Neurology 2005;64:1261–1262.

4. Lynch DR, Farmer JM, Tsou AY, et al. Measuring Friedreich ataxia: Complementary features of examination and performance measures. Neurology 2006;66:1711–1716.

5. Rummey C, Corben LA, Delatycki MB, et al. Psychometric properties of the friedreich ataxia rating scale. Neurol Genet 2019;5:371.

6. Lynch DR, Perlman SL, Meier T. A phase 3, double-blind, placebo-controlled trial of idebenone in Friedreich ataxia. Arch Neurol 2010;67:941–947.

7. Pandolfo M, Arpa J, Delatycki MB, et al. Deferiprone in Friedreich ataxia: A 6-Month randomized controlled trial: Deferiprone in FRDA. Ann Neurol 2014;76:509–521.

8. Zesiewicz Theresa, Salemi Jason L, Perlman Susan, Sullivan Kelly L, Shaw Jessica D, Huang Yangxin, Isaacs Charles, Gooch Clifton, Lynch David R, Klein Matthew B Double-blind, randomized and controlled trial of EPI-743 in Friedreich's ataxia. Neurodegenerative Disease Management 2018;8(4):233–242. http://dx.doi.org/10.2217/nmt-2018-0013

9. Shrout PE, Fleiss JL. Intraclass Correlations: Uses in Assessing Rater Reliability. Psychol Bull 1979;86:420.

10. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30–46.

11. Lee P, Liu C-H, Fan C-W, et al. The test–retest reliability and the minimal detectable change of the Purdue pegboard test in schizophrenia. J Formos Med Assoc 2013;112:332–337.

12. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research [Internet]. Evanston, Illinois: Northwestern University, 2018. https://cran.r-project.org/web/packages/psych/citation.html.

13. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–163.

14. Rummey C, Farmer JM, Lynch DR. Predictors of loss of ambulation in Friedreich's ataxia. EClinicalMed 2020;18:100213.

15. Friedman LS, Farmer JM, Perlman S, et al. Measuring the rate of progression in Friedreich ataxia: Implications for clinical trial design. Mov Disord 2010;25:426–432.

16. Patel M, Isaacs CJ, Seyer L, et al. Progression of Friedreich ataxia: quantitative characterization over 5 years. Ann Clin Transl Neurol 2016;3:684–694.

17. Bürk K, Mälzig U, Wolf S, et al. Comparison of three clinical rating scales in Friedreich ataxia (FRDA). Mov Disord 2009;24:1779–1784.

18. Fahey MC, Corben L, Collins V, et al. How is disease progress in Friedreich's ataxia best measured? A study of four rating scales. J Neurol Neurosurg Psychiatry 2007;78:411–413.

19. Zesiewicz TA, Stephenson JB, Kim SH, et al. Longitudinal gait and balance decline in Friedreich's Ataxia: A pilot study. Gait Posture 2017;55:25–30.

20. Bonilha da Silva C, Bergo FPG, D'Abreu A, et al. Dentate nuclei T2 relaxometry is a reliable neuroimaging marker in Friedreich's ataxia. Eur J Neurol 2014;21:1131–1136.

21. Clemm von Hohenberg C, Schocke MF, Wigand MC, et al. Radial diffusivity in the cerebellar peduncles correlates with clinical severity in Friedreich ataxia. Neurol Sci 2013;34:1459–1462.

22. Akhlaghi H, Corben L, Georgiou-Karistianis N, et al. Superior cerebellar peduncle atrophy in Friedreich's Ataxia correlates with disease symptoms. Cerebellum 2011;10:81–87.

23. Chevis CF, da Silva CB, D'Abreu A, et al. Spinal Cord Atrophy Correlates with Disability in Friedreich's Ataxia. Cerebellum 2013;12:43–47.

24. Evans-Galea MV, Carrodus N, Rowley SM, et al. FXN methylation predicts expression and clinical outcome in Friedreich ataxia. Ann Neurol 2012;71:487–497.