SHORT REPORT

How is disease progress in Friedreich's ataxia best measured? A study of four rating scales

M C Fahey, L Corben, V Collins, A J Churchyard, M B Delatycki

J Neurol Neurosurg Psychiatry 2007;**78**:411–413. doi: 10.1136/jnnp.2006.096008

Background: Friedreich's ataxia (FRDA), the most common genetic cause of ataxia, is characterised by progressive neurodegeneration and cardiomyopathy. Initial treatments are likely to slow progression rather than reverse morbidity. An appropriate and sensitive scale to measure disease progress is critical to detect the benefit of treatments.
Objective: To compare the Friedreich Ataxia Rating Scale (FARS) with other scales proposed as outcome measures for FRDA.
Methods: 76 participants were assessed with the FARS and the International Cooperative Ataxia Rating Scale (ICARS) and 72 of these participants were also assessed with the Functional Independence Measure and the Modified Barthel Index. 43 participants had repeat measures at an interval of 12 months. Sensitivity and responsiveness were assessed using the effect size for each measure and the sample size required for a placebo-controlled clinical trial.
Results: The FARS showed a high correlation with the other three measures. A significant change in the score over 12 months was detected by the FARS, the International Cooperative Ataxia Rating Scale and the Functional Independence Measure. The FARS had the greatest effect size and requires fewer patients for an equivalently powered study.
Conclusions: Of the scales assessed, the FARS is the best to use in clinical trials of FRDA. This is based on effect size, and power calculations that show that fewer participants are required to demonstrate the same effect of an intervention. Further work is required to develop more sensitive and responsive instruments.

Friedreich's ataxia (FRDA) is characterised by progressive neurological symptoms and cardiac dysfunction. It results in a reduced life span.[1 2] As the pathogenesis of FRDA has become better understood, several potential treatments have arisen. Initial beneficial treatments will probably slow progression rather than reverse morbidity. An appropriate and responsive scale for FRDA is critical if the benefits of potential treatments are to be identified. As there is no gold standard to examine disease progress in FRDA, the evaluation of measurement scales is essential to ensure that trials are efficient and their conclusions are accurate.[3]

The Friedreich Ataxia Rating Scale (FARS) comprises a measure of ataxia, an activities of daily living (ADL) subscale and a neurological subscale. Face and content validity and inter-rater reliability are good.[4] Other aspects of validity have not been examined. The International Cooperative Ataxia Rating Scale (ICARS) was developed to assess pharmacotherapies in ataxia,[5] and has good inter-rater reliability.[6] The Functional Independence Measure (FIM)[7] and the Modified Barthel Index (MBI)[8] examine the assistance required to complete ADLs. They are validated instruments used widely for neurological disease.

We aimed to examine these ordinal scales proposed as outcome measures for FRDA to establish concurrent criterion validity and compare the change in score over time to determine the sensitivity of the scales and confirm which measure is most appropriate for use in clinical trials.

METHODS

The study was conducted after approval by the Southern Health Ethics in Human Research Committee. Participants were recruited by invitation through the local FRDA support association and from among those who attended a multidisciplinary FRDA clinic. All participants were homozygotic for an expanded GAA repeat in the *FRDA* gene. Individuals with point mutations were excluded. No active selection of participants with particular characteristics took place.

The FARS, ICARS, FIM and MBI were administered by one of three assessors (MCF, AJC and LC) according to published instructions. The FIM was assessed by team consultation following a semistructured interview and examination. Repeat measures were undertaken at an interval of approximately 12 months with the examiners being blinded to the previous results.

The FARS consists of three subscales, comprising a general score for ataxia, a score for activities of daily living (ADL) and a neurological examination. The scores can be added to make a total score ranging from 0 to 159. A higher score indicates a greater level of disability. The ICARS Scores range from 0 to 100. A higher score also indicates more disability. The FIM (possible scores 18–126) and MBI (possible scores 0–100) scores decrease with decreasing function. Clinical details including age, age at symptom onset, disease duration and drug use was recorded. Disease duration was defined as the age at initial testing minus the age of onset.

Statistical analysis

Statistical analysis was undertaken using SPSS V.12.0.1. The statistical significance of the differences in baseline characteristics between those seen once and those seen twice were assessed using Fischer's exact test for proportions, and analysis of variance for mean scores. Spearman's correlation coefficients were calculated to assess the concurrent criterion validity of the FARS and its subscales against the ICARS, FIM and MBI. The statistical significance of the mean change in scale score from baseline to 12 months was assessed using paired t tests. The sensitivity was assessed by calculating the effect size defined as the mean difference in scores divided by the standard deviation of the baseline score.[9] A larger effect size indicates a more sensitive measure—that is, a greater change has been measured over the same time period. Power calculations were undertaken using Sample Power software (SPSS, Chicago, Illinois, USA) to

Abbreviations: ADL, activities of daily living; FARS, Friedreich Ataxia Rating Scale; FIM, Functional Independence Measure; FRDA, Friedreich's ataxia; ICARS, International Cooperative Ataxia Rating Scale; MBI, Modified Barthel Index

Table 1 Information related to the 76 study participants

	Single assessment			Assessed twice 12 months apart			Total		
		SD	Range		SD	Range		SD	p Value
Number of participants	33			43			76		
Gender: male %	45.5			53.5			50		0.64
Diabetes: % taking drugs	5.9			4.9			5.3		1.0
Late-onset FRDA: % with first symptom >25 years	9.1			16.3			13.2		0.50
Antioxidant use: % taking regular treatment	54.5			41.9			47.4		0.36
GAA2 mean: repeat number	911	152.7	658–1245	952.1	155.7	548–1345	934.9	154.7	0.026
GAA1 mean: repeat number	708.9	194.3	126–1005	646.8	212.4	284–1077	672.8	206	0.20
Age at initial examination: mean (years)	29.6	13.5	9–62	32.2	12.5	10–57	31.1	12.9	0.42
Age of onset: mean (years)	13.6	7.5	4–39	15.5	6.8	4–33	14.7	7.1	0.27
Disease duration: mean (years)	16.2	11.1	1.1–39.4	16.7	10.6	2–44.8	16.5	10.8	0.83
FARS: mean score	86.6	31.7	29.0–126.5	75.1	27.6	20.8–124.5	80.1	29.8	0.10
ICARS: mean score	53.5	21.7	14–82	47.5	19.2	15–80	50.1	20.4	0.20
FIM: mean score	100.5	20.5	63–125	112.4	13	75–126	107.6	17.7	<0.01
MBI: mean score	72.7	23.4	34–100	83.6	16.4	33–100	79.2	20.1	0.02

FARS, Friedreich Ataxia Rating Scale; FIM, Functional Independence Measure; FRDA, Friedreich's ataxia; GAA1, the size of the smaller of the two GAA expansions; GAA2, the size of the larger of the two GAA expansions; ICARS, International Cooperative Ataxia Rating Scale; MBI, Modified Barthel Index.
The p value is the significant difference between those seen once and those seen twice, 12 months apart.

assess the number of patients needed for a clinical trial with a treatment and placebo group with 80% power and $\alpha = 0.05$. A p value <0.05 was considered significant.

RESULTS

In all, 76 participants had at least one measurement with the FARS and ICARS, and 72 of these individuals were assessed with concurrent FIM and MBI measures. In total, 43 participants were measured using all scales on two occasions 12 months apart. Table 1 shows the demographic and clinical information of the participants, which shows that the cohort had a broad range of the various parameters examined. Age of onset, disease duration and the use of antioxidants was similar between the group examined once and that assessed twice. There was a trend for less severe disease in those seen twice as judged by the four scales. Differences in FIM and MBI Scores were seen in the group who had repeat testing compared with those tested only once. The correlation between the age of disease onset and the size of GAA1 ($r = -0.51$; $p<0.001$) was comparable with previous studies.[10 11]

The total FARS showed high correlation with the functional measures MBI and FIM (Spearman's $r = 0.9$ and 0.94) and the ICARS (Spearman's $r>0.96$), an established measure of ataxia. When the subscales of the FARS were examined, the ataxia subscale had the highest correlation with the FIM (Spearman's $r = 0.9$). The neurological subscale correlated best with the ICARS (Spearman's $r = 0.97$) and had a high correlation with the MBI (Spearman's $r = 0.93$) and FIM (Spearman's $r = 0.9$). The ADL subscale correlated with all measures (Spearman's $r = 0.9$ for the ICARS and MBI; $r = 0.87$ for the FIM). All correlations were significant ($p<0.001$). Correlations between scores for the four scales and disease duration were significant for all measures. The greatest correlation was for the total FARS compared with disease duration (Spearman's $r = 0.79$; $p<0.001$).

There was a significant decline over 12 months detected by the FARS, ICARS and FIM but not the MBI (table 2). The FARS had the greatest effect size and required fewer patients for an equivalently powered clinical trial than any of the other measures (table 2).

DISCUSSION

Of the four scales assessed, the FARS is the best for use in clinical trials of FRDA. This is based on it having the highest effect size, and power calculations that show that fewer participants are required to demonstrate the same effect of an intervention compared with other commonly used outcome measures.

The first aim of this study was to provide further validity to the FARS as an outcome measure in FRDA. As the total FARS has components to measure not only ataxia but also ADLs, we sought to perform this through correlation with both measures of ataxia and function. Concurrent criterion validity for the FARS is evident through correlations with well-established

Table 2 Measures of change over 12 months for the four scales tested

Scale	Mean change over 12 months	SD (baseline)	SD (change)	p Value	Effect size	Power calculation: subjects per group*	Standardised ratio: subjects per group†
FARS	−9.5	27.9	9.1	<0.001	0.34	60	1
ICARS	−5	19.5	6.8	<0.001	0.26	118	2
FIM	3.1	12.9	6.0	0.02	0.24	253	4
MBI	1.9	16.3	6.2	NS	0.12	605	10

FARS, Friedreich Ataxia Rating Scale; FIM, Functional Independence Measure; ICARS, International Cooperative Ataxia Rating Scale; MBI, Modified Barthel Index.
*The number of participants needed for a clinical trial with a treatment and placebo group assuming that the treatment slows the observed disease progression over 12 months by 50%, with 80% power and $\alpha = 0.05$.
†Compared with a trial powered to detect a 50% reduction in disease progress, a 25% reduction would require four times as many patients per group. A 10% reduction in disease progress would require 25 times more patients in each group.

measures of disability, the MBI and FIM. The strong correlation between the scales suggests that these scales measure a common underlying construct. Of the subcomponents of the FARS, the neurological subscales have the highest correlation with the ICARS. The ataxia and ADL components correlate with all measures, indicating a clear overlap between functional assessment and ability.

As FRDA is a progressive neurological disease, we predict that scores indicating more severe disease would correlate with disease duration when examined in cross section. The correlation of all measures with disease duration was significant, but was greatest with the total FARS Score. The association of the score with disease duration provides face validity for these measures.

The second aim was to determine the most sensitive scale to measure clinical decline in FRDA. For a scale to be used in a clinical setting, especially for a clinical trial, it must be sensitive to change. This is especially true in a progressive, fatal condition such as FRDA, where there is no proved treatment to slow disease progression. Even a small benefit from a therapeutic intervention is likely to be clinically relevant. In this context, the most sensitive scale will be the most appropriate one to use.

Effect size calculations suggest that the FARS is a more sensitive measure than the other three scales. Power calculations indicated that the total FARS is the most responsive to change, in that fewer patients are required than if using the other three scales for equivalent power in a trial.

Our cohort was recruited by invitation through a patient support group and through a clinic that sees a broad range of individuals with FRDA. No selection took place except through exclusion of those with a point mutation. However, there may have been a selection bias with respect to less mobile individuals who were less likely to attend. Some evidence suggested that those seen twice were less severely affected at baseline than those seen once. However, there was still a broad range of disease severity in those with repeat measures, and the study compared the same individuals with all scales.

Despite the FARS performing better than the other three scales tested, the effect size over a period of 12 months is only moderate. This may reflect the sensitivity of the scale, or alternatively that the true amount of change seen in 12 months is modest. A longer period of assessment with greater numbers of patients will help to determine this. Four scales were used for comparison. How newly developed measures of ataxia, such as the Scale for the Assessment and Rating of Ataxia, will perform against a specific FRDA ataxia scale is yet to be determined.[12] Furthermore, being an ordinal scale, the true relationship between scale points is not defined for the FARS. Therefore, work is required to find improved measures for FRDA. Other methods may help in refining measures to behave in a more linear or continuous manner. These include techniques such as Rasch analysis of the FARS and the development of a functional composite.[13]

ACKNOWLEDGEMENTS

We thank the participants, who willingly gave their time, and continue to support our research.

Authors' affiliations

M C Fahey, L Corben, V Collins, M B Delatycki, Bruce Lefroy Centre for Genetic Health Research, Murdoch Childrens Research Institute, Parkville, Victoria, Australia
M C Fahey, A J Churchyard, Monash Neurology, Monash Medical Centre, Melbourne, Victoria, Australia

Funding: Funding was received through the National Health and Medical Research Council, Australia (MCF and MBD), the Friedreich Ataxia Research Association (Australasia) and the Friedreich Ataxia Research Alliance (USA).

Competing interests: None.

Correspondence to: Associate Professor Martin Delatycki, Bruce Lefroy Centre for Genetic Health Research, Murdoch Childrens Research Institute, Flemington Road, Parkville, Victoria 3052, Australia; martin.delatycki@ghsv.org.au

Received 17 April 2006
Revised 21 July 2006
Accepted 17 October 2006
Published Online First 20 October 2006

REFERENCES

1 **Durr A**, Cossee M, Agid Y, *et al*. Clinical and genetic abnormalities in patients with Friedreich's ataxia. *New Engl J Med* 1996;**335**:1169–75.
2 **Voncken M**, Ioannou P, Delatycki MB. Friedreich ataxia—update on pathogenesis and possible therapies. *Neurogenetics* 2004;**5**:1–8.
3 **Hobart JC**, Lamping DL, Freeman JA, *et al*. Evidence-based measurement: which disability scale for neurologic rehabilitation? *Neurology* 2001;**57**:639–44.
4 **Subramony SH**, May W, Lynch D, *et al*. Measuring Friedreich ataxia: interrater reliability of a neurologic rating scale. *Neurology* 2005;**64**:1261–2.
5 **Trouillas P**, Takayanagi T, Hallett M, *et al*. International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. The Ataxia Neuropharmacology Committee of the World Federation of Neurology. *J Neurol Sci* 1997;**145**:205–11.
6 **Storey E**, Tuck K, Hester R, *et al*. Inter-rater reliability of the International Cooperative Ataxia Rating Scale (ICARS). *Mov Disord* 2004;**19**:190–2.
7 **Keith RA**, Granger CV, Hamilton BB, *et al*. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil* 1987;**1**:6–18.
8 **Shah S**, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol* 1989;**42**:703–9.
9 **Kazis LE**, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;**27**:S178–89.
10 **Filla A**, De Michele G, Cavalcanti F, *et al*. The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am J Hum Genet* 1996;**59**:554–60.
11 **Delatycki MB**, Paris DB, Gardner RJ, *et al*. Clinical and genetic study of Friedreich ataxia in an Australian population. *Am J Med Genet* 1999;**87**:168–74.
12 **Schmitz-Hubsch T**, du Montcel ST, Baliko L, *et al*. Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* 2006;**66**:1717–20.
13 **Lynch DR**, Farmer JM, Wilson RL, *et al*. Performance measures in Friedreich ataxia: potential utility as clinical outcome tools. *Mov Disord* 2005;**20**:777–82.