Poster 7.018

biohaven

Psychometric Validation of the Modified Functional Scale for the Assessment and Rating of Ataxia

Michele H. Potashman, PhD[1]; Evan Popoff, MSc[2]; Lauren Powell, BS, MPH[2]; Melissa Wolfe Beiner, MD[1]; Ainsley Mackenzie, BS[1]; Vlad Coric, MD[1]; Jeremy D. Schmahmann, MD[3]; Gil L'Italien, PhD[1]

[1]Biohaven Pharmaceuticals, Inc, New Haven, CT, USA; [2]Broadstreet HEOR, Vancouver, BC, Canada; [3]Department of Neurology, Ataxia Center, Laboratory for Neuroanatomy and Cerebellar Neurobiology, Massachusetts General Hospital, Boston, MA, USA

CONCLUSIONS

1. The psychometric evaluation showed that the modified functional Scale for the Assessment and Rating of Ataxia (f-SARA) performed well on a range of analyses examining the reliability and validity of the measure in subjects with spinocerebellar ataxia (SCA)
2. Furthermore, f-SARA demonstrated the ability to serve as a reliable assessment of disease progression over a period of ≥1 year
3. A threshold of 1 point as the intra-individual meaningful change was supported by multiple analyses

BACKGROUND

- SCAs are a group of rare, dominantly inherited, heterogenous disorders that cause progressive neurodegeneration of the cerebellum and spinal cord[1,2]
  - More than 50 distinct genetic subtypes have been identified, with the most prevalent worldwide being SCA3 (25%–50%), SCA2 (13%–18%), and SCA6 (13%–15%)[3–5]
  - Health-related quality of life is severely impacted by SCA, and patients experience a high clinical burden due to limited independence, reliance on caregivers, and impacts on social and physical function[6,7]
  - There is interest in improving the measurement of clinically meaningful ataxia symptoms for use in clinical trial settings[8,9]
- The f-SARA, a derivative of the SARA, was developed with input provided by the US Food and Drug Administration (FDA) and analysis of US natural history data from the clinical research consortium for the study of cerebellar ataxia and phase 2 troriluzole data[10]
  - The f-SARA is comprised of the axial items of the SARA (items 1 through 4: gait, stance, sitting, and speech), with response options collapsed to a uniform 5-point ordinal scale
  - Rating options on the f-SARA reflect normal function (0), mild impairment (1), moderate impairment (2), severe impairment (3), and inability to complete the task (4)
  - The maximum total score is 16 with higher scores indicating more severe impairment
- The development of a clinical outcome assessment importantly needs psychometric validity of the measure to be established, with inclusion of a clinimetric assessment, to ensure adequate measurement properties have been achieved

References: **1.** Klockgether T, et al. *Nat Rev Dis Primers*. 2019;5(1):24. **2.** Diallo A, et al. *J Neurol*. 2021;268(8):2749-2756. **3.** Bhandari J, et al. Spinocerebellar ataxia. Accessed February 2, 2024. https://www.ncbi.nlm.nih.gov/pubmed/32491748 **4.** Teive HAG, et al. *Mov Disord Clin Pract*. 2019;6(7):531-540. **5.** Alshammeri S, et al. *Mov Disord Clin Pract* 2023;10:533–44. **6.** Schmitz-Hübsch T, et al. *Movement Dis*. 2010;25(5):587-595. **7.** Sánchez-Lopez CR, et al. *Neurologia*. 2017;32(3):143-151. **8.** Chen ML, et al. *J Neurol Sci*. 2021;424:117417. **9.** Maas R, van de Warrenburg BPC. *Parkinsonism Relat Disord*. 2021;91:37-41. **10.** Beiner M, et al. *Neurology* 2021: 96:2837. **11.** Moulaire P, et al. *Mov Disord* 2023: 38:35-44 **12.** ClinicalTrials.gov. Troriluzole in adult subjects with spinocerebellar ataxia. Accessed February 2, 2024. https://clinicaltrials.gov/study/NCT03701399.

Disclosures: MHP, AM, MWB, VC, and GL are employees of Biohaven Pharmaceuticals, Inc. EP and LP are employees of Broadstreet HEOR, which received funds from Biohaven Pharmaceuticals, Inc, for this work. JDS is a consultant for Biohaven Pharmaceuticals and site PI for Biohaven Pharmaceuticals clinical trials (NCT03701399, NCT02960893, and NCT03952806); receives royalties from Oxford University Press, Elsevier, Mac Keith Press, and Springer; and is the inventor of the Brief Ataxia Rating Scale, Cerebellar Cognitive Affective/Schmahmann Syndrome Scale, Patient-Reported Outcome Measure of Ataxia, and Cerebellar Neuropsychiatry Rating Scale, which are licensed to the General Hospital Corporation.

Acknowledgments: Medical writing support was provided by James Glen, of Nucleus Global and funded by Biohaven Pharmaceuticals, Inc. This presentation was developed in accordance with Good Publication Practice (GPP2022) guidelines. Authors had full control of the content and made the final decision on all aspects of this presentation.

To download a copy of this poster, scan QR code.

American Academy of Neurology (AAN) 2024 Annual Meeting; April 13-18, 2024 · Denver, Colorado & Virtual

OBJECTIVE and METHODS

Objective

- To examine the psychometric and clinimetric measurement properties of f-SARA in subjects with SCA using two SCA datasets

Methods

- Psychometric measurement properties of the f-SARA were evaluated using data from a cohort of SCA subjects enrolled in the PROM-Ataxia validation study from Massachusetts General Hospital (cross-sectional; MGH cohort) and subjects enrolled in BHV4157-206 (NCT03701399; 48-week study; Study BHV4157-206 cohort)[12]
  - The MGH psychometric cohort (n=33) represented patients in a real-world clinical setting
    - Key inclusion criteria for the subjects (MGH) were age of ≥18 years and diagnosis of SCA (SCA types 1, 2, 3, 6, 7, 8, or 10)
    - Subjects from BHV4157-206 were not included in the MGH cohort
  - The Study BHV4157-206 cohort (n=217) further allowed evaluation of the SCA3 genotype separately due to its large size, and for psychometric and clinimetric assessments that required longitudinal data
- Psychometric properties evaluated included: data acceptability (ceiling and floor effects), internal consistency (Cronbach α), test-retest reliability, convergent and divergent validity, known-groups validity, and responsiveness
  - **Data acceptability** was determined by examining the distributions (minimum and maximum values and IQR) of the total score and scores for each item
    - Acceptability is supported when observed scores are well distributed, mean scores are near the scale midpoint, and floor and/or ceiling effects are minimized
    - A threshold of >15% of subjects with scores at either the minimum or maximum item value was used to indicate floor and ceiling effects, respectively
  - **Internal consistency reliability** was assessed by the Cronbach α coefficient (raw and standardized) and item-to-total correlations (Spearman r). An α of ≥0.70 and item-to-total correlation of ≥0.30 served as thresholds of acceptable internal consistency
  - **Test-retest reliability** was assessed by calculating the intraclass correlation coefficient (ICC) for each item and the total score between screening and baseline, with a threshold of >0.60 being acceptable
  - **Convergent and divergent validity** was assessed through a correlation matrix of the f-SARA scores with several clinician-assessed measures and patient-reported outcomes measures
    - Domains examined included upper limb mobility, lower limb mobility, speech, overall ataxia symptoms, fatigue, overall physical abilities, ADL, anxiety, depression, and overall mental or emotional state
  - **Known-groups validity** was examined by comparing mean values (using an independent *t* test on f-SARA items and total scores) between two groups with differing disease severity:
    - Least severe vs most severe (by tertile), as defined by staging measures of Klockgether severity score and FARS-FUNC score
  - **Responsiveness** data were derived primarily through anchor-based analytics; however, known-groups findings (by quartile) were considered as supportive
- Intra-individual **meaningful change thresholds** were examined in the Study BHV4157-206 cohort, with minimum important change (MIC) values derived using distribution-based and anchor-based methods
  - Distribution-based methods: 0.5×SD and standard error of measurement (SEM)
  - Anchor-based methods, with Clinical Global Impression–Global Improvement Scale (CGI-I) as anchor: empirical cumulative distribution function (eCDF) and probability density function (PDF) curves plotted

RESULTS

Demographics and Clinical Characteristics

- A total of 33 subjects comprised the MGH psychometric cohort, representing SCA genotypes SCA1, SCA2, SCA3, SCA6, SCA6/8, and SCA8, with the most common ataxias being SCA3 (54.5%), SCA2 (15.2%), and SCA6 (15.2%) (**Table 1**)
- The mean (SD) total f-SARA score was 5.8 (3.6), with a range of 0.0 to 12.0
- Study BHV4157-206 enrolled 217 subjects with SCA; mean (SD) age was 47.6 (12.8) years, 51.2% were female, mean (SD) age at symptom onset was 38.3 (12.3) years, and mean (SD) total f-SARA score was 4.9 (1.8), with a range of 2.0 to 11.0
  - In a subgroup of 89 patients with an SCA3 genotype, baseline characteristics were as follows: mean (SD) age was 46.7 (12.1) years, 51.7% were female, mean (SD) age at symptom onset was 39.1 (11.8) years, and mean (SD) total f-SARA score was 4.9 (1.8), with a range of 2.0 to 10.0

Table 1. Demographics and Clinical Characteristics (MGH Psychometric Cohort)

Characteristics	MGH psychometric cohort (n=33)
Known genotype, n (%)	
SCA1	2 (6.1)
SCA2	5 (15.2)
SCA3	18 (54.5)
SCA6	5 (15.2)
SCA6/8	2 (6.1)
SCA7	0
Klockgether severity, n (%)	
Presymptomatic (stage 0)	2 (6.1)
Mild (stage 1)	13 (39.4)
Moderate (stage 2)	7 (21.2)
Severe (stage 3)	11 (33.3)
Baseline total f-SARA score	
Mean (SD)	5.8 (3.6)
Median (range)	6.0 (0.0–12.0)

f-SARA, modified functional Scale for the Assessment and Rating of Ataxia; MGH, Massachusetts General Hospital; SCA, spinocerebellar ataxia.

Psychometric Properties (MGH Cohort, With Study BHV4157-206 as Supportive)

- Among subjects enrolled in the MGH cohort (n=33), ceiling effects were absent while floor effects were observed for 3 of 4 items (floor effects were not observed for the gait item). IQRs were skewed toward the lower end of response options (**Table 2**)

Table 2. f-SARA Data Acceptability (MGH Psychometric Cohort)

f-SARA domain (item statistic)	MGH psychometric cohort (n=33)
Gait (#1 gait)	
Mean (SD)	2.0 (1.2)
Median (IQR)	2.0 (1.0–3.0)
Balance (#2 stance)	
Mean (SD)	1.5 (1.1)
Median (IQR)	2.0 (1.0–2.0)
Sitting (#3 sitting)	
Mean (SD)	1.0 (0.8)
Median (IQR)	2.0 (0.0–2.0)
Speech (#4 speech disturbance)	
Mean (SD)	1.3 (0.9)
Median (IQR)	1.0 (1.0–2.0)

f-SARA, modified functional Scale for the Assessment and Rating of Ataxia; MGH, Massachusetts General Hospital.

- Excellent internal consistency was demonstrated, with an overall raw Cronbach α of 0.88 (α_{total}=0.90; $\alpha_{items\text{-}removed}$=0.86-0.90), and item-to-total correlations were acceptable (r=0.82-0.91 per item) (**Table 3**)

Table 3. f-SARA Internal Consistency Reliability (MGH Psychometric Cohort)

f-SARA domain (item statistic)	Cronbach α, standardized (raw)[a]	Item-to-total correlation[b]
Gait (#1 gait)	0.86 (0.85)	0.91
Balance (#2 stance)	0.87 (0.85)	0.89
Sitting (#3 sitting)	0.86 (0.85)	0.90
Speech (#4 speech disturbance)	0.90 (0.88)	0.82
f-SARA total score	**0.90 (0.89)**	—

f-SARA, modified functional Scale for the Assessment and Rating of Ataxia; MGH, Massachusetts General Hospital.
[a] Cronbach α overall and per item if item deleted, raw and standardized.
[b] Spearman r.

- Convergent and divergent validity were supported, with stronger correlations observed between f-SARA and scales of similar constructs (*P*<0.001) (e.g., FARS-FUNC, r=0.92; BARS, r=0.88; PROM-ADL, r=0.83), a moderate correlation with the FARS-ADL (r=0.69, [*P*<0.001]), and weaker correlations among measures of differing constructs (e.g., PIFAS-FATIGUE, r=0.41 [*P*=0.017]; PIFAS-SPEECH, r=0.45 [*P*=0.009]; and BAI, r=0.45, [*P*=0.008]) (**Table 4**)
  - Relationships hypothesized to have low correlation but determined to be higher included the Neuro-QOL [upper], r=−0.82 [*P*<0.001]; PIFAS-EMOTION r=0.57, [*P*=0.001], and PROM-MENTAL, r=0.51; *P*=0.002)

Table 4. f-SARA Construct Validity—Convergent Validity (MGH Psychometric Cohort)

Instrument	Spearman correlation with f-SARA total score	*P* value
Total PIFAS score	0.67	<0.001
PIFAS-FATIGUE score	0.41	0.017
PIFAS-GAIT score	0.70	<0.001
PIFAS-ADL score	0.71	<0.001
PIFAS-SPEECH score	0.45	0.009
PIFAS-EMOTION score	0.57	0.001
FARS-ADL total score	0.69	<0.001
FARS-FUNC total score	0.92	<0.001
Neuro-QOL (upper)	−0.82	<0.001
Neuro-QOL (lower)	−0.76	<0.001
Neuro-QOL (fatigue)	0.64	<0.001
BARS total score	0.88	<0.001
BDI total score	0.23	0.195
BAI total score	0.45	0.008
PROM-PHYS score	0.72	<0.001
PROM-ADL score	0.83	<0.001
PROM-MENTAL score	0.51	0.002

ADL, activities of daily living; BAI, Beck anxiety inventory; BARS, brief ataxia rating scale; BDI, Beck depression inventory; FARS-ADL, Friedreich's Ataxia Rating Scale–Activities of Daily Living; FARS-FUNC, Friedreich's Ataxia Rating Scale–Function; f-SARA, modified functional Scale for the Assessment and Rating of Ataxia; Neuro-QOL (fatigue), Neurology Quality of Life Fatigue Scale; Neuro-QOL (lower), Neurology Quality of Life Lower Extremity Scale; Neuro-QOL (upper), Neurology Quality of Life Upper Extremity Scale; PIFAS, Patient Impression of Function and Activities of Daily Living Scale; PROM, Patient-Reported Outcome Measure.

- Known-groups validity was supported as subjects with worse disease severity indicators had significantly higher f-SARA scores in multiple analyses
  - The mean (SD) f-SARA score (as examined by Klockgether severity) was 3.1 (2.6) for those in the lowest group and 9.5 (1.6) in the highest group (*P*<0.001) (**Table 5**)
  - This was similarly observed for the FARS-FUNC severity indicator, with the mean (SD) f-SARA score being 1.7 (1.3) and 9.1 (1.7) in the lowest and highest FARS-FUNC score groups, respectively (**Table 6**)
    - Examination of f-SARA scores by FARS-FUNC quartiles by item and total score (data not shown) further discriminated between disease severity categories
    - This level of discrimination often signals that a measure may be responsive to change in longitudinal testing
- Similar trends were observed between the MGH cohort and Study BHV4157-206, in both the all-SCA and SCA3 cohorts

Table 5. f-SARA Construct Validity—Known Groups Klockgether Severity Score (MGH Psychometric Cohort)

	Klockgether severity			
	Group 1 pre-symptomatic or mild (n=15)	Group 2 Moderate (n=7)	Group 3 severe (n=11)	*t*-test independent sample (*P* value)
Mean (SD) f-SARA score	3.1 (2.6)	6.0 (1.7)	9.5 (1.6)	<0.001

f-SARA, modified functional Scale for the Assessment and Rating of Ataxia.

Table 6. f-SARA Construct Validity—Known Groups FARS-FUNC Score (MGH Psychometric Cohort)

	FARS-FUNC score			
	Group 1 FARS-FUNC score 0 to 2 (n=10)	Group 2 FARS-FUNC score 2.5 to 3.5 (n=8)	Group 3 FARS-FUNC score 4 to 6 (n=15)	*t*-test independent sample (*P* value)
Mean (SD) f-SARA score	1.7 (1.3)	5.0 (1.6)	9.1 (1.7)	<0.001
Mean (SD) FARS-FUNC score	1.4 (0.8)	3.2 (0.3)	4.4 (0.6)	—

FARS-FUNC, Friedreich's Ataxia Rating Scale–Function; f-SARA, modified functional Scale for the Assessment and Rating of Ataxia.

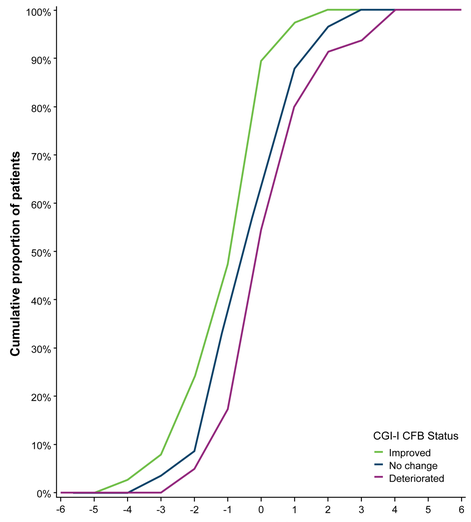
Psychometric properties (Study BHV4157-206 Cohort, All SCA)

- Test-retest reliability was supported with ICC (95% CI) values of 0.91 (0.88−0.93) for the f-SARA total score and was 0.92 (0.89–0.94) for gait, 0.77 (0.70–0.82) for stance, 0.73 (0.65–0.79) for sitting, and 0.77 (0.71–0.82) for speech

Anchor-Based Analysis by CGI-I Status

- Using Study BHV4157-206 all-SCA data, the eCDF curve by CGI-I anchor category showed that the f-SARA was able to capture both meaningful improvements and deterioration over 48 weeks (**Figure 1**)
  - The mean 48-week change scores in subjects divided by anchor category was −0.68 points in subjects with improvement, 0.02 points in those with no change, and 0.58 points in those with deterioration

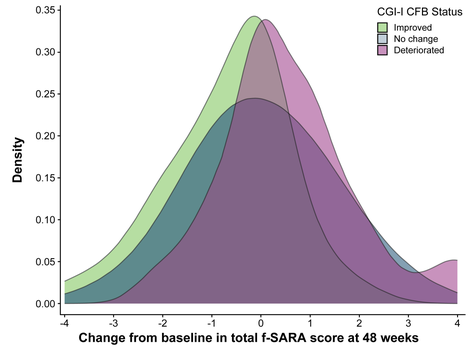
Figure 1. f-SARA eCDF by CGI-I (Study BHV4157-206)[a]



CFB, change from baseline; CGI, Clinical Global Impression-Global Improvement Scale; eCDF, empirical cumulative distribution function; f-SARA, modified functional Scale for the Assessment and Rating of Ataxia.
[a] All SCA subjects examined in this analysis.

- The anchor-based PDF curves are supportive of the findings in the eCDF curve analysis (**Figure 2**)
- Distribution-based findings were 0.5×SD=0.89 and SEM=1.12

Figure 2. f-SARA PDF by CGI-I Anchor Status



CFB, change from baseline; CGI, Clinical Global Impression-Global Improvement Scale; f-SARA, modified functional Scale for the Assessment and Rating of Ataxia; PDF, probability density function.