Assessment of impaired coordination in children
Lawerman, Tjitske Fenna

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 7

Reliability and discriminant validity of ataxia rating scales in early onset ataxia

R. Brandsma
T.F. Lawerman
M.J. Kuiper
R.J. Lunsing
H. Burger
D.A. Sival

Developmental Medicine & Child Neurology 2017; 59(4):427-432
ABSTRACT

**Aim** To determine whether ataxia rating scales are reliable disease biomarkers for early onset ataxia (EOA).

**Method** In 40 patients clinically identified with EOA (28 males, 12 females; mean age 15y 3mo [range 5–34y]), we determined interobserver and intraobserver agreement (interclass correlation coefficient [ICC]) and discriminant validity of ataxia rating scales (International Cooperative Ataxia Rating Scale [ICARS], Scale for Assessment and Rating of Ataxia [SARA], and Brief Ataxia Rating Scale [BARS]). Three paediatric neurologists independently scored ICARS, SARA and BARS performances recorded on video, and also phenotyped the primary and secondary movement disorder features. When ataxia was the primary movement disorder feature, we assigned patients to the subgroup ‘EOA with core ataxia’ (n=26). When ataxia concurred with other prevailing movement disorders (such as dystonia, myoclonus, and chorea), we assigned patients to the subgroup ‘EOA with comorbid ataxia’ (n=12).

**Results** ICC values were similar in both EOA subgroups of ‘core’ and ‘comorbid’ ataxia (0.92–0.99; ICARS, SARA, and BARS). Independent of the phenotype, the severity of the prevailing movement disorder predicted the ataxia rating scale scores (β=0.83–0.88; p<0.05).

**Interpretation** In patients with EOA, the reliability of ataxia rating scales is high. However, the discriminative validity for ‘ataxia’ is low. For adequate interpretation of ataxia rating scale scores, application in uniform movement disorder phenotypes is essential.

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARS</td>
<td>Ataxia rating scales</td>
</tr>
<tr>
<td>BARS</td>
<td>Brief Ataxia Rating Scale</td>
</tr>
<tr>
<td>EOA</td>
<td>Early onset ataxia</td>
</tr>
<tr>
<td>ICARS</td>
<td>International Cooperative Ataxia Rating Scale</td>
</tr>
<tr>
<td>ICC</td>
<td>Intraclass correlation coefficient</td>
</tr>
<tr>
<td>SARA</td>
<td>Scale for Assessment and Rating of Ataxia</td>
</tr>
</tbody>
</table>
INTRODUCTION

Early onset ataxia (EOA) concerns a group of rare, predominantly genetic and metabolic ataxic disorders, manifesting before the 25th year of life.\(^1\)\(^-\)\(^3\) This diagnostic group covers a wide heterogeneity of disorders regarding age at onset, inheritance, and underlying pathogenesis. Consequently, the phenotype is also heterogeneous, involving both EOA with core ataxia (i.e. EOA with ataxia as the core symptom) and EOA with comorbid ataxia (i.e. EOA with other movement disorder features that may prevail over ataxia).\(^4\) Especially in children, EOA is often prevalent as a combined phenotype, with concurrent features of dystonia, myoclonus, chorea, and spasticity, which may prevail over ataxia.\(^4\)\(^,\)\(^5\) This heterogeneity makes uniform phenotypic EOA assessment a challenging task for paediatric movement disorder specialists. Therefore, quantitative ataxia rating scales (ARS) are often used as additionally reproducible ‘surrogate’ biomarkers for ataxia.\(^5\)\(^-\)\(^11\) The International Cooperative Ataxia Rating Scale (ICARS),\(^6\) the Scale for Assessment and Rating of Ataxia (SARA),\(^7\) and the Brief Ataxia Rating Scale (BARS)\(^8\) are the most commonly applied ARS in children and adults.

These ARS quantify ataxia in four domains: (1) posture and gait; (2) kinetic function; (3) speech; and (4) oculomotor function (exclusively BARS and ICARS).\(^6\)\(^-\)\(^8\) ICARS is considered to be the most detailed, BARS the briefest, and SARA the most uniformly reproducible scale.\(^6\)\(^-\)\(^8\),\(^10\),\(^11\) In children, we have shown that ARS are not only influenced by ataxia, but also by age\(^12\) and by muscle weakness (in Friedreich’s ataxia).\(^13\) In EOA, this implies that other influences, such as concurrent movement disorders, could influence the scores. For reliable interpretation of ARS as ‘ataxia’ biomarkers, we therefore reasoned that, first, clarification was required of the paediatric ARS test construct. In the present EOA study, we thus aimed to elucidate ARS for: (1) observer agreement and (2) discriminant validity (i.e. the potential to determine ‘ataxia severity’ and not the severity of other, with ataxia concurrent movement disorders). Such information may support reliable data entry in international EOA databases, and may also support the interpretation of ARS outcomes in therapeutic trials. Especially in regards to ongoing, innovative trials in heterogeneous patients with EOA, we reasoned that confounding effects should be identified before small fluctuations in ARS scores are over-interpreted as therapeutic ‘ataxic’ improvement.\(^12\)\(^-\)\(^17\) Thus, we aimed to investigate the observer agreement and discriminant validity of ARS in patients with EOA.

METHOD

Patients
The medical ethical committee of the University Medical Center Groningen, the Netherlands, approved the study. We based our sample size calculation on previously published interobserver agreement (intraclass correlation coefficient [ICC]) data in adults with ataxia,\(^7\) because data on
quantitative ARS scores in children with ataxia are still lacking. In adults with ataxia, a sample size of 36 patients scored by three observers achieved a 90% power ($b=0.1$) to detect an ICC of 0.8 or over the null hypothesis of a moderate ICC of 0.6, using a significance level of 0.05. Based on the clinical diagnosis (patient record descriptions from 1998 to 2012), we approached 40 patients (28 males, 12 females; mean age 15y 3mo [range 5–34y]). All patients were clinically identified with ataxic features before the 25th year of life and fulfilled the ‘classical’ definition of EOA.1 We excluded patients with postnatally acquired focal cerebellar lesions (such as by infections, trauma, inflammation, or cerebrovascular attacks). In accordance with Dutch medical ethical law, legal representatives (when younger than 18y of age) and patients (when older than 12y of age) consented to participate. The response rate was 100%. For clinical description of the included patients, see Table I.

**Procedure**

We video recorded ARS performances in all 40 patients.12 Three paediatric neurologists (RB, RJL, DAS) quantitatively scored the videotaped test-performances according to the guidelines of ICARS, SARA, and BARS. We determined interobserver agreement by comparing the total and subscale scores of the three assessors. After a latent time interval of 5 (3–7) weeks, the three assessors repeated their ARS assessments in the first 10 videotaped patients, without permission to review their previous scores. We determined intraobserver agreement by comparing the first and second scores. After a latent time interval of 6 months, the same assessors phenotyped the video-taped testperformances for the presence of ataxia and/or other movement disorders (i.e. ataxia, dystonia, chorea, myoclonus, tremor, spasticity, and ‘sloppiness’), either as the primary or as the secondary feature. We subsequently assigned patients to an EOA subgroup with ‘core ataxia’ when (1) all three assessors independently recognized ataxia as the primary movement disorder, or when (2) all three assessors had independently confirmed the presence of ataxia and when the underlying diagnosis (genetically and/or metabolically) confirmed an ataxic phenotype. We assigned patients to an EOA subgroup with ‘comorbid ataxia’ when the criteria for the EOA subgroup ‘core ataxia’ were not met and when ataxia was observed (by at least one observer) as a concurrent feature with other movement disorders. The assessors indicated the perceived severity of the movement disorder (i.e. mild [1], moderate [2], or severe [3]). To check for the reliability of these assessments, we compared the perceived severity between the participating assessors and four other members of the movement disorder team of the University Medical Center Groningen (who had not rated the ARS), revealing a significant association ($\chi^2$ test; $p<0.001$).5

For global data interpretation of phenotypic ataxia severity assessment, we also compared phenotypic ataxia severity outcomes with the ataxia severity grading system, proposed by Klockgether et al.18 (i.e. stage 0, no gait difficulties; stage 1, gait difficulties; stage 2, loss of independent gait with permanent use of a walking aid; stage 3, confinement to a wheelchair; stage 4, dead).
Statistical analysis

We performed statistical analysis by PASW Statistics 20 for Windows (SPSS, Hong Kong). We determined mean ICARS, SARA, and BARS total scores from the quantitative assessments by the three assessors. We also determined ARS total scores per primary movement disorder and median phenotypic severity of the primary movement disorder by the three assessors. We assessed normality of age, disease duration, and ARS total scores, by probability plots (Q–Q plots). We compared the ARS scores between patients with EOA with core ataxia and patients with EOA with comorbid ataxia by Student’s t-test (in case of non-normality by Mann–Whitney U test; $= Mann-Whitney U test; $ = Student T-test; # = Chi square test; SD = Standard deviation; ICARS = International Cooperative Ataxia Rating Scale; SARA = Scale for Assessment and Rating of Ataxia; BARS = Brief Ataxia rating scale. Underlying diagnoses in the ‘primary-EOA’ subgroup were: Friedreich’s ataxia (FRDA) (n= 7), Niemann Pick type C (n=1), Ataxia with vitamin E deficiency (AVED) (n=3), NARPmutation (n=1), Ataxia Telangiectasia (n=1), Kearns Sayre syndrome (n=1), North Sea Myoclonus (GOSR-2 mutation) (n=4), 2-methyl-3-hydroxybutyryl-CoA-hydrogenase deficiency (MHBD) (n=1), Joubert syndrome (KIAA0586 mutation) (n=1), CACNA1A mutation (n=1) and unknown causes (n=5). Underlying diagnoses in the ‘combined-EOA’ subgroup were: Benign hereditary chorea (TIFF1-mutation) (n=1), Huntington disease (n=1), cerebellar malformation (n=1), Chediak Higashi syndrome (n=1), Spastic paraplegia type 11(SPG-11 mutation) (n =1), CTNNB1 mutation (n =1), ataxic cerebral palsy (n=1), congenital CMV infection (n=1), functional disorder (n=1) and unknown causes (n=3). *Disease severity grading by Klockgether*
U test). We calculated the percentage of the subscale score compared with the total score by: 
(subscale score/total ARS score) × 100%. We compared outcomes between both EOA subgroups.
We determined interobserver and intraobserver agreement by ICC. We used the two-way random 
single measurement variant for the interobserver agreement and the one-way single measurement 
variant for the intraobserver agreement. 19 According to Cicchetti, 20 official cut-offs for qualitative 
rating of ICC values are as follows: ICC <0.40, poor; 0.40 to 0.59, fair; 0.60 to 0.74, good; 0.75 to 1.00, 
excellent. For uniformity reasons with previously published data, 5,12 we also interpreted outcomes 
by Landis & Koch criteria; 21 which are originally described for categorical data. According to Landis & 
Koch, we characterized ICC outcomes by: ICC <0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 
0.61 to 0.80, substantial; >0.81, almost perfect. We determined the correlation between the ARS 
outcomes by Pearson coefficient (in case of non-normality we used Spearman’s rho coefficient). We 
determined the correlation between the ataxia severity grading system proposed by Klockgether 
et al. 18 and the phenotypic severity of the movement disorder, and we also correlated outcomes 
with total ARS scores by Spearman’s rho coefficient. In perspective of previously reported ARS age-
dependency in typically developing children, we compared the paediatric EOA scores with these 
historic age-related mean control values, by Mann–Whitney U test. 12 To determine the discriminant 
validity of ARS for ataxia severity, we determined the association between the primary movement 
disorder features (i.e. ataxia, dystonia, myoclonus, chorea, spasticity, tremor, and ‘sloppiness’) 
and the total ARS scores by the Kruskall–Wallis test. We performed a multiple regression analysis 
to determine the effect of age, sex, disease duration, primary movement disorder feature, and 
the severity of the primary movement disorder feature on the total ARS scores. Because ARS 
are specifically designed to reflect ataxia severity, we deliberately included semi-quantitative 
information about the perceived phenotypic severity of the most dominant movement disorder 
(including other movement disorders then ataxia) in our model. We applied a stepwise regression 
analysis with forward selection starting with age, 22 and we explored which variables would have 
added predictive value over and above variables already in the model. 22 All statistical tests were 
two-tailed. Statistical significance was set at \( p < 0.05 \).

RESULTS

Patient characteristics
There were no missing data. In two of the 40 included patients, none of the assessors recognized 
ataxia as part of the movement disorder. These two patients (diagnosed with DYT-6 and SPG-11) were 
therefore excluded from further analysis. Thus, the EOA data were obtained from the remaining 38 
patients. Subdivision into EOA sub-groups with core ataxia and comorbid ataxia revealed 26 (68.4%) 
patients in the former group and 12 (32.6%) patients in the latter group. For patient characteristics, 
see Table I. Probability plots revealed normally distributed disease duration. Age and the total ARS 
scores (ICARS, SARA, and BARS) were not normally distributed. Comparing age (16y 11mo vs 12y)
Reliability and Validity of ataxia rating scales in EOA

and disease duration (12y 1mo vs 7y 6mo) between EOA with core ataxia and EOA with comorbid ataxia revealed no significant difference ($p=0.112$ and $p=0.096$ respectively). Comparing the ataxia severity grading system proposed by Klockgether et al.$^{18}$ between the EOA subgroups revealed no significant difference ($p=0.436$). However, comparing the phenotypic movement disorder severity grading system between the EOA sub-groups revealed higher movement disorder severity in the EOA subgroup with core ataxia than in the EOA subgroup with comorbid ataxia ($p=0.040$). Total ICARS, SARA, and BARS scores were significantly higher in EOA with core ataxia than in EOA with comorbid ataxia ($p=0.001$ for ICARS, SARA, and BARS, see Table I). In 21 out of 26 (80.7%) patients with EOA with core ataxia, all three assessors recognized ataxia as the primary movement disorder feature. The remaining 5 out of 26 (19.3%) patients with EOA were assigned to the core ataxia subgroup by the underlying genetic or metabolic diagnosis (AVED n=2; GOSR2 mutation n=3). In all of these five patients, two of the three assessors recognized ataxia as the primary movement disorder feature, and one assessor recognized ataxia as a secondary movement disorder feature. Total ARS scores were similar between the two EOA core ataxia subgroups (i.e. either identification by all three assessors, or identification by two assessors and the underlying diagnosis [$p=0.753$, $p=0.659$, and $p=0.613$ for ICARS, SARA, and BARS respectively]).

ARS subscales in patients with EOA

ARS subscale scores were not significantly different between EOA subgroups with core ataxia and comorbid ataxia (Table SI, online supplementary information).

Reliability of ARS in patients with EOA.

The quantitative ARS scores were characterized by an interobserver agreement (ICC) of 0.969, 0.977, and 0.913 (for ICARS, SARA, and BARS respectively; all $p<0.001$, i.e. excellent and almost perfect, according to Cicchetti and Landis & Koch).$^{20,21}$ The ICC of the subscales varied between 0.705 and 0.982 (for ICARS, SARA, and BARS; all $p<0.001$ [good to excellent and substantially to nearly perfect, according to Cicchetti and Landis & Koch])$^{20,21}$ The ARS ICC varied between 0.966 and 0.994 (all $p<0.001$; i.e. excellent and nearly perfect according to Cicchetti and Landis & Koch).$^{20,21}$ See Table SII (online supporting information).

Discriminant validity of ARS in patients with EOA.

All three ARS were strongly correlated ($r_s = 0.988$, 0.958, and 0.941 for ICARS and SARA, ICARS and BARS, SARA and BARS respectively [all $p<0.001$]). The ataxia severity grading system proposed by Klockgether et al.$^{18}$ was moderately correlated ($r_s = 0.450–0.476$, $p<0.001$) and the phenotypic movement disorder severity grading system was strongly correlated ($r_s = 0.775–0.801$, $p<0.001$) with total ARS scores. The ataxia severity grading system proposed by Klockgether et al.$^{18}$ and the phenotypic movement disorder severity grading system were also significantly correlated with each other ($r_s = 0.513; p=0.001$). Comparing quantitative ARS scores between included children with EOA (<18y of age; n=25) and historic age-related mean control values,$^{12}$ revealed significantly higher ARS
Figure 1: Ataxia rating scale (ARS) scores according to age, in children with early onset ataxia (EOA) and typically developing participants.

The x-axis indicates the age of the children in years. The y-axis indicates ARS scores: (A) ICARS, (B) SARA, (C) and BARS scores. The blue dots represent individual outcomes in children with EOA (n=25; ≤18y of age), connected by the blue linear regression line. The red dots represent individual outcomes in typically developing participants (derived from Brandsma et al.),12 connected by a red one phase decay trend line. Outcomes reveal significantly higher total ARS scores in patients with EOA compared with typically developing participants (for ICARS, SARA, and BARS, p<0.001 [Mann–Whitney U test]). ICARS, International Cooperative Ataxia Rating Scale; SARA, Scale for Assessment and Rating of Ataxia; BARS, Brief Ataxia Rating Scale.

Figure 2: Ataxia rating scale (ARS) scores according to the primary movement disorder feature.

The x-axis indicates the phenotypically assessed primary movement disorder (ataxia [n=27];* dystonia [n=2]; chorea [n=4]; spasticity [n=1]; sloppiness [n=3]; and myoclonus [n=1]) in early onset ataxia (EOA) with core ataxia and EOA with comorbid ataxia subgroups. The y-axis indicates ICARS (A), SARA (B), and BARS (C) scores respectively. ARS scores do not significantly differ between primary ataxia and other primary movement disorder (p=0.062; p=0.068, and p=0.072 for ICARS, SARA, and BARS respectively). *26 out of 27 patients with ataxia as median primary movement disorder also fulfilled the criteria for EOA with core ataxia. In one patient unfulfilling the criteria for EOA with core ataxia, ataxia was recognized by two of three assessors as the primary movement disorder. ICARS, International Cooperative Ataxia Rating Scale; SARA, Scale for Assessment and Rating of Ataxia; BARS, Brief Ataxia Rating Scale.
scores in the children with EOA ($p<0.001$ for ICARS, SARA, and BARS, Fig. 1). Phenotypic assessment of the primary movement disorder feature revealed 27 patients with ataxia (27 out of 38[71.1%]); one with myoclonus (1 out of 38 [2.6%]); two with dystonia (2 out of 38 [5.3%]); four with chorea (4 out of 38 [10.5%]); one with spasticity (1 out of 38 [2.6%]); and three with ‘sloppiness’ (3 out of 38 [7.9%]). Comparing quantitative ARS scores between the phenotypically determined primary movement disorder features revealed no statistically significant differences ($p=0.062$, $p=0.068$, and $p=0.072$ for ICARS, SARA, and BARS respectively [Kruskall–Wallis test], see Fig. 2). Multiple regression analysis showed that total ARS scores are significantly predicted by the severity of the primary movement disorder in ICARS ($\beta=0.86; \ p=0.026$), SARA ($\beta=0.83; \ p=0.026$), and BARS ($\beta=0.88; \ p=0.024$), independent of whether the primary movement disorder features concern ataxia or not. The severity of the prevailing movement disorder explained a significant proportion in the variance of the ARS scores for ICARS ($R^2=0.764; p<0.001$), SARA ($R^2=0.775; p<0.001$), and BARS ($R^2=0.754; p<0.001$). The type of primary movement disorder did not render a significant F-change, nor did age, sex, or disease duration, implicating that these variables could be omitted from our regression model (for further analysis, see Table SIII, online supporting information).

**DISCUSSION**

In patients with EOA, ICARS, SARA, and BARS reveal high interobserver and intraobserver agreement, reflecting the reliability of the scores. However, the discriminant validity of ARS failed to discern between the influence of ataxia and the influence of other movement disorders. In EOA with the phenotype core ataxia, ARS can thus be regarded as reliable and reproducible biomarkers for ataxia severity. However, in children with EOA with the phenotype comorbid ataxia, ARS scores can be confounded by the influence of other concurrent movement disorders. This implies that ARS scores do not necessarily reflect the severity of ‘ataxia’ alone.

In patients with EOA, total ARS scores revealed similarly high ICC outcomes for interobserver and intraobserver agreement, as previously reported in adults with ataxia (0.91–0.99 vs 0.91–0.98 respectively).6–11 This implies that the total scores of all three ARS are highly reproducible and that one may choose a scale for its own intrinsic properties, instead of for reasons of interobserver agreement, alone. However, subscale analysis reveals relatively low interobserver agreement for the oculomotor subscale. As oculomotor parameters are not included in the SARA, SARA might be preferred above other ARS;7 but, this would only be valid under the premise that information on oculomotor function could be intentionally left out. Although previously published ICC results in typically developing children (0.62–0.96)12 appear lower than the present data in children with EOA, this does not necessarily imply that the reproducibility in typically developing children is lower. This is explained by the method of ICC calculation, in which a small variation in (typically developing age-related) scores will mathematically induce a low numerical ICC outcome, whereas the absolute observer differences can be the same. This implicates that the numerical ICC value is not necessarily indicative of the score agreeability alone.
Chapter 7

Interestingly, we observed that cross-sectional EOA ARS scores were not significantly predicted by age. This is understandable as the severity of the primary movement disorder exerted a much stronger effect on the EOA ARS scores than did age (i.e. >87% more). Despite this, consideration of ARS age-dependency is advisory, especially when longitudinal ARS scores with minimal changes (cut-off margins) are being considered as relevant for therapeutic gain.23,24

Regarding discriminant validity, multiple regression analysis revealed that the severity of the primary movement disorder influenced ARS scores, independent of the phenotype of the primary movement disorder. In the EOA subgroup with core ataxia, ARS outcomes were thus reflective of ataxia severity. However, in the EOA subgroup with comorbid ataxia, ARS scores were confounded by the influence of other concurrent movement disorders. In addition to previously described confounding factors (such as paediatric age and muscle strength in patients with Friedreich’s ataxia),12,13 one might anticipate that additional influences, such as neuropathy, could also confound ARS scores. As different patient groups are needed to substantiate this hypothesis, we hope that future studies will elucidate this point.

Overall, the insight provided into the ARS construct has direct implications for the assessment of therapeutic interventions in children with EOA.23,24 When small changes in ARS scores are being considered as indicators for ‘therapeutic’ ataxia improvement, one should strive to include homogeneous patients (regarding both age and phenotype).16,17

There are some weaknesses in this study. First, patients were quantitatively scored and phenotyped by the same assessors. However, because there was a time interval (of 6mo) between both assessments, and because assessors were not allowed to review their previous scores, a bias appears unlikely. Second, in the absence of quantitative ARS data in children with EOA, our sample size calculation was based on ARS ICC data in adults with ataxia.7 We are aware of the potential limitation on quantitative ARS data by the relatively small sample size, especially regarding the applied multiple regression analysis. However, because the underlying disorders of the included patients with EOA are rare, we suggest that these data can be interpreted as indicative. We hope that future international studies will have larger sample sizes to allow elucidation of these findings to further extent. Third, we are aware that the Friedreich Ataxia Rating Scale was not included in the present analysis. However, because SARA was recently characterized as a reliable scale in patients with Friedreich’s ataxia,25 and because SARA is highly correlated with ICARS and BARS, one may deduce that ICARS, SARA, and BARS are applicable in all patients of the EOA group, including Friedreich’s ataxia.

To conclude, ARS are reliably reproducible in patients with EOA. In patients with EOA with a core ataxic phenotype, total ICARS, SARA, and BARS scores can be regarded as sufficiently reliable for assessment of the ataxia severity. However, in patients with EOA with a comorbid ataxic phenotype, ARS are not only influenced by ataxia, but also by other concurrent movement disorders. Despite high reliability of ARS scores, discriminant validity appears insufficient for phenotypic EOA subgroups with comorbid ataxia. For reliable data interpretation of ARS scores, we conclude that the scores should be interpreted in homogeneous phenotypic EOA groups.
REFERENCES

Chapter 7


SUPPLEMENTARY DATA

Table SI: Percentage (%) of ataxia rating scales (ARS) sub-scale scores in EOA subgroups with ‘core ataxia’ and ‘comorbid ataxia’

<table>
<thead>
<tr>
<th></th>
<th>EOA with core ataxia</th>
<th>EOA with co-morbid ataxia</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICARS total score</td>
<td>41.89</td>
<td>18.03</td>
</tr>
<tr>
<td>Gait</td>
<td>41%</td>
<td>39%</td>
</tr>
<tr>
<td>Kinetic</td>
<td>47%</td>
<td>51%</td>
</tr>
<tr>
<td>Speech</td>
<td>7%</td>
<td>9%</td>
</tr>
<tr>
<td>Oculomotor</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>SARA total score</td>
<td>16.59</td>
<td>6.56</td>
</tr>
<tr>
<td>Gait</td>
<td>51%</td>
<td>43%</td>
</tr>
<tr>
<td>Kinetic</td>
<td>37%</td>
<td>42%</td>
</tr>
<tr>
<td>Speech</td>
<td>12%</td>
<td>15%</td>
</tr>
<tr>
<td>BARS total score</td>
<td>12.97</td>
<td>5.65</td>
</tr>
<tr>
<td>Gait</td>
<td>29%</td>
<td>26%</td>
</tr>
<tr>
<td>Kinetic</td>
<td>50%</td>
<td>54%</td>
</tr>
<tr>
<td>Speech</td>
<td>13%</td>
<td>18%</td>
</tr>
<tr>
<td>Oculomotor</td>
<td>8%</td>
<td>2%</td>
</tr>
</tbody>
</table>

%ARS sub-scores per EOA subgroup. Percentage of ARS sub-scales are calculated by the formula: sub-scores/total scores x 100%. For each ataxia rating scale, we provide outcomes (mean total score and sub-scale-score). EOA with comorbid ataxia tended to reveal a slightly higher %kinetic function and a slightly lower %gait function than EOA with core ataxia, although the level of significance was not reached.

Table SII: ICCs for ataxia rating scale (ARS) scores

<table>
<thead>
<tr>
<th></th>
<th>Interobserver agreement</th>
<th>Intraobserver agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>‘EOA with core ataxia’</td>
<td>‘EOA with co-morbid ataxia’</td>
</tr>
<tr>
<td>ICARS total</td>
<td>.969</td>
<td>.967</td>
</tr>
<tr>
<td>Gait</td>
<td>.982</td>
<td>.986</td>
</tr>
<tr>
<td>Kinetic</td>
<td>.918</td>
<td>.913</td>
</tr>
<tr>
<td>Speech</td>
<td>.818</td>
<td>.827</td>
</tr>
<tr>
<td>Oculomotor</td>
<td>.771</td>
<td>.768</td>
</tr>
<tr>
<td>SARA total</td>
<td>.977</td>
<td>.977</td>
</tr>
<tr>
<td>Gait</td>
<td>.982</td>
<td>.979</td>
</tr>
<tr>
<td>Kinetic</td>
<td>.906</td>
<td>.902</td>
</tr>
<tr>
<td>Speech</td>
<td>.866</td>
<td>.864</td>
</tr>
<tr>
<td>BARS total</td>
<td>.913</td>
<td>.917</td>
</tr>
<tr>
<td>Gait</td>
<td>.958</td>
<td>.973</td>
</tr>
<tr>
<td>Kinetic</td>
<td>.782</td>
<td>.784</td>
</tr>
<tr>
<td>Speech</td>
<td>.807</td>
<td>.820</td>
</tr>
<tr>
<td>Oculomotor</td>
<td>.705</td>
<td>.710</td>
</tr>
</tbody>
</table>

ICC = Intra-class Correlation Coefficient; Interobserver agreement subdivided according to total, ‘EOA with core ataxia’ and ‘EOA with comorbid ataxia’. ICC’s without indication = $p<.005$; $p=.024$; $p=.120$
### Table SIII: Multiple regression analysis of total ataxia rating scale scores

<table>
<thead>
<tr>
<th></th>
<th>ICARS total score</th>
<th>SARA total score</th>
<th>BARS total score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F change</td>
<td>$B^*$</td>
<td>$\beta$</td>
</tr>
<tr>
<td>Age</td>
<td>0.19</td>
<td>0.48</td>
<td>0.04</td>
</tr>
<tr>
<td>Gender</td>
<td>0.05</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>Disease duration</td>
<td>0.02</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>Primary movement disorder</td>
<td>1.11</td>
<td>0.96</td>
<td>1.12</td>
</tr>
<tr>
<td>Primary movement disorder severity</td>
<td>21.98***</td>
<td>23.94***</td>
<td>20.32***</td>
</tr>
<tr>
<td>Non vs. Mild</td>
<td>-5.77 (20.05)</td>
<td>-.12</td>
<td>-3.5 (8.36)</td>
</tr>
<tr>
<td>Non vs. Moderate</td>
<td>14.09 (19.03)</td>
<td>.32</td>
<td>5.31 (7.93)</td>
</tr>
<tr>
<td>Non vs. Severe</td>
<td>47.90 (20.24)</td>
<td>.86*</td>
<td>19.87 (8.43)</td>
</tr>
</tbody>
</table>

Regression analysis results for the effects of age, gender, disease duration, primary movement disorder and phenotypically assessed primary movement disorder severity on ataxia rating scale total scores. $B^*$ (unstandardized coefficients with standard error in parenthesis) and $\beta$ (standardized regression coefficient). * $p<.05$; ** $p<.01$; *** $p<.001$. 