Responsiveness of the Scale for the Assessment and Rating of Ataxia and Natural History in 884 Recessive and Early Onset Ataxia Patients

Andreas Traschütz, MD, PhD [1,2] Astrid D. Adarmes-Gómez, MD, [3,4] Mathieu Anheim, MD, PhD, [5,6,7] Jonathan Baets, MD, PhD, [8,9,10] Bernard Brais, MD, PhD, [11] Cynthia Gagnon, PhD [12] Janina Gburek-Augustat, MD, [13] Sarah Doss, MD, [14,15] Haşmet A. Hanağası, MD, [16] Christoph Kamm, MD, [17] Peter Klivenyi, MD, [18] Thomas Klockgether, MD, [19,20] Thomas Klopstock, MD, [21,22,23] Martina Minnerop, MD, [24,25,26] Alexander Münchau, MD [27] Mathilde Renaud, MD, [28,29] Filippo M. Santorelli, MD [30] Ludger Schöls, MD, [1,2] Andreas Thieme, MD, [31] Stefan Vielhaber, MD, [32,33,34] Bart P. van de Warrenburg, MD, PhD, [35] Ginevra Zanni, MD, PhD, [36] Ralf-Dieter Hilgers, PhD, [37] PREPARE Consortium, and Matthis Synofzik, MD [1,2]

Objective: The Scale for the Assessment and Rating of Ataxia (SARA) is the most widely applied clinical outcome assessment (COA) for genetic ataxias, but presents metrological and regulatory challenges. To facilitate trial planning,

View this article online at wileyonlinelibrary.com. DOI: 10.1002/ana.26712

Received Oct 4, 2022, and in revised form Apr 4, 2023. Accepted for publication May 22, 2023.

Address correspondence to Dr Synofzik, Research Division Translational Genomics of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research and Center of Neurology, University of Tübingen, Hoppe-Seyler-Str 3, 72076 Tübingen, Germany. E-mail: matthis.synofzik@uni-tuebingen.de

Members of the PREPARE Consortium are available as an online Supplementary file.

From the [1]Research Division "Translational Genomics of Neurodegenerative Diseases," Hertie Institute for Clinical Brain Research and Center of Neurology, University of Tübingen, Tübingen, Germany; [2]German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany; [3]Movement Disorders Unit, Department of Neurology and Clinical Neurophysiology, Institute of Biomedicine of Seville, Virgen del Rocío University Hospital/CSIC/University of Seville, Seville, Spain; [4]Center for Biomedical Research Network on Neurodegenerative Diseases, Madrid, Spain; [5]Department of Neurology, Hautepierre Hospital, University Hospitals of Strasbourg, Strasbourg, France; [6]Federation of Translational Medicine of Strasbourg, University of Strasbourg, Strasbourg, France; [7]Institute of Genetics and Molecular and Cellular Biology, INSERM-U964/CNRS-UMR7104/University of Strasbourg, Illkirch, France; [8]Translational Neurosciences, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium; [9]Laboratory of Neuromuscular Pathology, Institute Born-Bunge, University of Antwerp, Antwerp, Belgium; [10]Neuromuscular Reference Center, Department of Neurology, Antwerp University Hospital, Antwerp, Belgium; [11]Department of Neurology, McGill University, Montreal Neurological Institute, Montreal, Quebec, Canada; [12]CHUS Research Center and Health and Social Services Center of Saguenay–Lac-Saint-Jean, Faculty of Medicine, University of Sherbrooke, Quebec, Quebec, Canada; [13]Division of Neuropediatrics, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany; [14]Department of Neurology, Charité–Universitätsmedizin Berlin, corporate member of Free University of Berlin, Humboldt University of Berlin, Berlin, Germany; [15]Department of Neurological Sciences, University of Nebraska Medical Center, Omaha, NE, USA; [16]Behavioral Neurology and Movement Disorders Unit, Department of Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey; [17]Department of Neurology, University of Rostock, Rostock, Germany; [18]Interdisciplinary Excellence Center, Department of Neurology, Faculty of Medicine, Albert Szent-Györgyi Clinical Center, University of Szeged, Szeged, Hungary; [19]Department of Neurology, University Hospital Bonn, Bonn, Germany; [20]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany; [21]Department of Neurology, Friedrich Baur Institute, Ludwig Maximilian University of Munich, Munich, Germany; [22]German Center for Neurodegenerative Diseases (DZNE), Munich, Germany; [23]Munich Cluster for Systems Neurology, Munich, Germany; [24]Institute of Neuroscience and Medicine, Research Center Jülich, Jülich, Germany; [25]Department of Neurology, Center for Movement Disorders and Neuromodulation, Medical Faculty,

we characterize its responsiveness (including subitem-level relations to ataxia severity and patient-focused outcomes) across a large number of ataxias, and provide first natural history data for several of them.
Methods: Subitem-level correlation and distribution-based analysis of 1,637 SARA assessments in 884 patients with autosomal recessive/early onset ataxia (370 with 2–8 longitudinal assessments) were complemented by linear mixed effects modeling to estimate progression and sample sizes.
Results: Although SARA subitem responsiveness varied between ataxia severities, *gait/stance* showed a robust granular linear scaling across the broadest range (SARA < 25). Responsiveness was diminished by incomplete subscale use at intermediate or upper levels, nontransitions ("static periods"), and fluctuating decreases/increases. All subitems except *nose-finger* showed moderate-to-strong correlations to activities of daily living, indicating that metric properties—not content validity—limit SARA responsiveness. SARA captured mild-to-moderate progression in many genotypes (eg, *SYNE1*-ataxia: 0.55 points/yr, ataxia with oculomotor apraxia type 2: 1.14 points/yr, *POLG*-ataxia: 1.56 points/yr), but no change in others (autosomal recessive spastic ataxia of Charlevoix-Saguenay, *COQ8A*-ataxia). Whereas sensitivity to change was optimal in mild ataxia (SARA < 10), it substantially deteriorated in advanced ataxia (SARA > 25; 2.7-fold sample size). Use of a novel rank-optimized SARA without subitems *finger-chase* and *nose-finger* reduces sample sizes by 20 to 25%.
Interpretation: This study comprehensively characterizes COA properties and annualized changes of the SARA across and within a large number of ataxias. It suggests specific approaches for optimizing its responsiveness that might facilitate regulatory qualification and trial design.

ANN NEUROL 2023;94:470–485

Introduction

With mechanistic treatment trials on the horizon for many genetic ataxias, sensitive capture of clinical treatment response has become key for academia, industry, and regulatory agencies. The Scale for the Assessment and Rating of Ataxia (SARA) serves as the most widely applied primary clinical outcome assessment (COA; specifically: clinician-reported outcome) for almost all genetic ataxias.[1,2] The SARA has been applied in large observational cohort studies to estimate disease progression and trial sizes in spinocerebellar ataxia (SCA) types 1, 2, 3, and 6, or Friedreich ataxia (FA),[3,4] and as primary endpoint in several randomized controlled trials.[5–11]

However, although validity and reliability of the SARA are excellent,[1,12–14] regulatory agencies and recent studies in SCA3 have raised concerns given its insufficiently understood sensitivity to change and functional relevance, especially at the level of single subitems.[15–17] Repeated home-based video assessments have recently highlighted the strong intraindividual variability of the SARA,[18] which might be particularly driven by certain subitems. Although designed as a scale for a single construct of ataxia, the SARA score level is influenced by multisystemic features variably present across most ataxia genotypes (eg, neuropathy, spasticity, or nonataxia movement disorders), but their functional effect on the SARA properties is as yet unclear. Together, these limitations of the SARA present a major challenge in current trial planning. Although modifications to optimize the SARA are thus now being discussed (eg, the modified SARA score *f-SARA*[19]), the underlying data evidence and validation of such modifications have remained scarce.[16] Extensive real-world data from ataxia registries, with prospective SARA assessments across many years and ataxia genotypes, may in turn provide the opportunity for statistical modeling and data-driven optimization of its responsiveness as a generic COA of ataxia.

Harnessing a large multicenter prospective cohort study of autosomal recessive and early onset ataxias, we here provide an in-depth analysis of SARA's ability to capture change, both across (ie, as generic COA) and within (ie, as genotype-specific COA) a large number of ataxias. Specifically, we (1) analyze its responsiveness at the total score as well as the subitem level; (2) identify the influence of nonataxia features and ataxia severity; (3) characterize the relation to patient-focused outcomes; (4) provide first prospective natural history data for a number of genetic ataxias to facilitate the design of treatment trials; and (5) simulate data-driven rank-optimized SARA composites, illustrating their improved sensitivity by trial size estimations.

Patients and Methods

Study Cohort

Prospective cross-sectional and longitudinal data from all 948 consecutive patients enrolled between 2013 and

Heinrich Heine University, Düsseldorf, Germany; [26]Institute of Clinical Neuroscience and Medical Psychology, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany; [27]Institute of Systems Motor Science, Institute of Neurogenetics, University of Lübeck, Lübeck, Germany; [28]Clinical Genetics Service, CHRU of Nancy, Nancy, France; [29]INSERM-U1256 NGERE, University of Lorraine, Nancy, France; [30]IRCCS Foundation Stella Maris, Pisa, Italy; [31]Department of Neurology and Center for Translational Neuro and Behavioral Sciences, Essen University Hospital, University of Duisburg-Essen, Essen, Germany; [32]Department of Neurology, Otto von Guericke University, Magdeburg, Germany; [33]German Center for Neurodegenerative Diseases (DZNE), Magdeburg, Germany; [34]Center for Behavioral Brain Sciences, Otto von Guericke University Magdeburg, Magdeburg, Germany; [35]Department of Neurology, Donders Institute for Brain, Cognition, and Behavior, Radboud University Medical Center, Nijmegen, the Netherlands; [36]Unit of Neuromuscular and Neurodegenerative Disorders, Department of Neurosciences, Bambino Gesù Childrens' Hospital, IRCCS, Rome, Italy; and [37]Department of Medical Statistics, RWTH Aachen University, Aachen, Germany

Additional supporting information can be found in the online version of this article.

August 4, 2021 were retrieved from the global multicenter ARCA Registry.[20] Datasets included (1) genotypic and demographic data, (2) assessments of ataxia severity (SARA[1]) and nonataxia features (including the Inventory of Non-Ataxia Signs [INAS][21]), and (3) the functional staging (FARS-FS) and activities of daily living (ADL) scales of the Friedreich Ataxia Rating Scale (FARS)[22] as patient-focused outcomes. Patients had been eligible for inclusion into the ARCA Registry if they had (1) a genetically confirmed autosomal recessive cerebellar ataxia (ARCA); and/or (2) an early onset ataxia (EOA) with onset before age 40 years without evidence of an autosomal dominant family history, repeat-expansion SCA, or acquired cause (eg, subacute onset, rapid progression, alcohol intake, abnormal B12 levels, cerebrospinal fluid pleocytosis, or structural lesions on imaging), thus representing a stratum of ataxia patients known to be enriched for ARCAs.[23,24] Discarding eligibility failures and incomplete datasets, 931 ARCA/EOA patients provided SARA and/or phenotypic data for the current study. The total cohort included overall 1,637 SARA assessments in 884 patients (370 with 2–8 longitudinal assessments; 86 genotypes), at least one INAS assessment in 908 patients, 187 assessments with FARS-FS, and 62 assessments with FARS-ADL. As a general structuring principle, analyses were performed on this total cohort by default, and in genotype-specific subcohorts depending on available datasets. If genotype-specific data were too limited to allow for robust analysis, analyses were performed on a subset of 9 ARCAs recurring in at least 15 patients grouped and referred to as "common ARCAs" (n = 393 patients), with adjustment for genotype as covariate in the respective multivariate statistical analysis. This study was approved by the institutional review board of the medical faculty of the University of Tübingen (598/2011BO1), and patients at each contributing site had provided informed consent for pseudonymized data entry in the ARCA Registry.

Statistical Analysis

Prevalence and Influence of Nonataxia Features. The presence of multisystemic nonataxia features was determined from each patient's last available INAS and the registry's systematic history for a priori recurrent features of ARCAs (eg, epilepsy or mitochondrial features such as hearing loss or diabetes; case report form "Clinical Features"[20]). Their respective impact on the SARA as a measure of ataxia severity was analyzed by generalized linear modeling (GLM), using the SARA score of each patient from the same visit as dependent variable, and selecting age at onset, ataxia duration, ataxia duration squared, and genotype as independent variables.[25] For each individual nonataxia feature that occurred in at least 40 patients (ie, >∼5% prevalence) Fig 1C, a separate GLM was calculated with the respective nonataxia feature as independent factor. To account for multiple comparisons of nonataxia features, p values were adjusted by the Benjamini–Hochberg method with a false discovery rate of 0.05. Only nonataxia features with independent replication in separate GLMs including the most common (genetically solved and stratified) ARCAs and all unsolved EOAs, respectively, were considered significant generic determinants of ataxia severity.

Correlation-Based Analysis of SARA Subitems. The responsiveness (here defined as intraitem evolution along disease severity) of SARA subitems across the full range of ataxia severity was examined by correlating each subitem score to the corresponding total SARA score. Response curves and 95% prediction intervals were estimated by fitting third order polynomial curves, given that these aligned best with the variable patterns obtained by nonparametric smoothing of data.[26] To analyze the homogeneity of subitem response curves across ataxias, this analysis was performed across all 1,637 assessments in the cohort, and independently for each common ARCA as compared to all other ARCAs/EOAs. Patient-focused functional relevance was determined by correlating the SARA and its subitems to FARS-ADL, a 9-item composite scale of basic activities of daily living (speech, cutting food/handling utensils, dressing, personal hygiene, falling, walking, sitting) and impairments (swallowing, bladder function; Fig 1A).

Progression Analysis. Annualized progression rates for the SARA and its subitems were estimated using linear mixed effects modeling (LMEM; restricted maximum likelihood, covariance structure: variance components, Kenward–Roger degrees of freedom) with random effects on intercept (score at baseline) and slope (points per year of follow-up).[4] Annualized progression rates were compared between ataxia severities, SARA subitems, and different composites by adding interaction terms between follow-up time (months since baseline visit/12) and the respective groups. Ataxia severity was binned according to the SARA at baseline (<10, 10–25, and >25 for mild, moderate, and advanced ataxia) and separated into ambulatory and nonambulatory patients (operationalized by SARA subitem *gait* > 6, indicating inability to walk >10m even with strong support or full inability to walk). Although always arbitrary, the pragmatic binning according to baseline SARA was data-driven, with the aim of capturing relevant floor, ceiling, and plateau segments of SARA subitems in cross-sectional data. All bins were evaluated by comparing available FARS-FS levels as external anchor. Genotype (of the common ARCA genotypes) was added as fixed effect to the LMEM in control analyses, allowing capture of both genotype-specific effects and also across-genotype

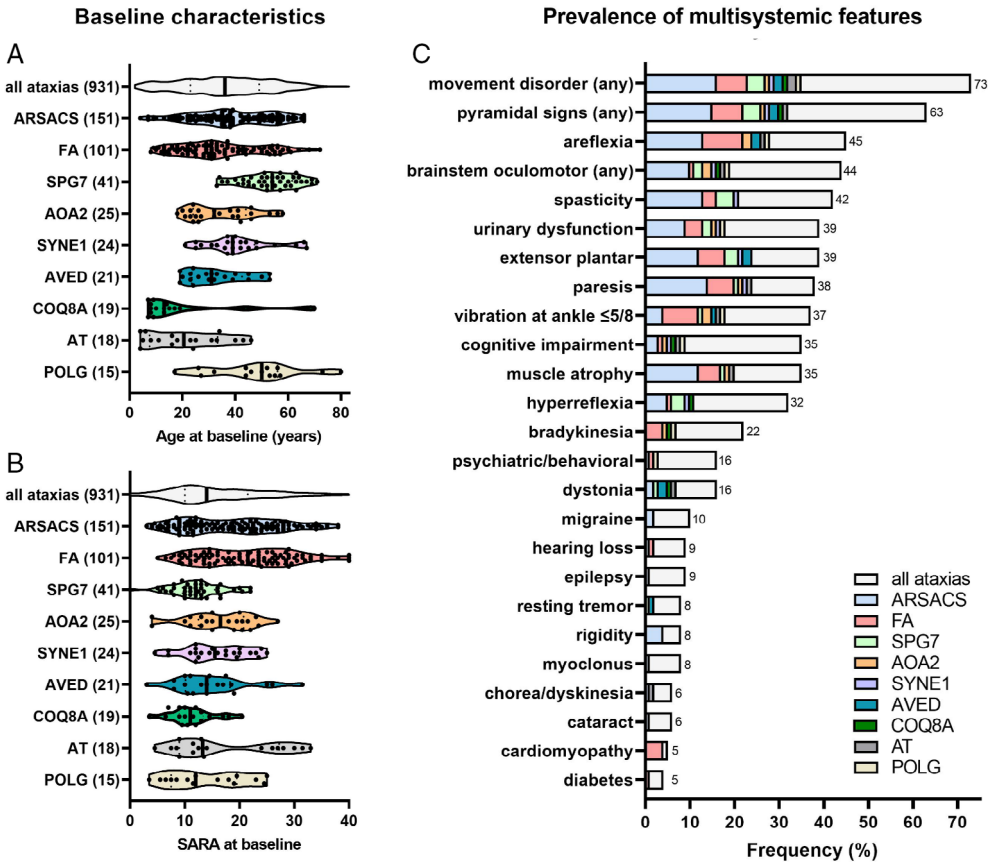

FIGURE 1: Characterization of multisystemic ataxia cohort. (A, B) Characterization of patient age (A) and ataxia severity (B) at baseline assessment for the autosomal recessive and unsolved early onset cerebellar ataxias that are part of this study. Patients in this study represent a wide distribution of these features, particularly also in the "common autosomal recessive cerebellar ataxias" (color-coded). Solid and dashed lines in violin plots indicate median and quartiles. Number of patients is shown in parentheses. (C) Frequency of extracerebellar features observed in at least 5% of patients. Multisystemic phenotypes were representative for the ataxia cohort, and—with the exception of cardiomyopathy—not predominantly caused by one of the most common genotypes. AOA2 = ataxia with oculomotor apraxia type 2; ARSACS = autosomal recessive spastic ataxia of Charlevoix-Saguenay; AT = ataxia–telangiectasia; AVED = ataxia with vitamin E deficiency; FA = Friedreich ataxia; SARA = Scale for the Assessment and Rating of Ataxia; SPG7 = spastic paraplegia type 7. COQ8A/POLG/SYNE1 indicate ataxia related to the respective genes.

effects on disease progression. We considered potential effect of ceiled and floored observation on the model fit and compared our results for LMEM with Tobit models. As a methodological tool to compare sensitivity to change between different SARA composites and ataxia severities, we calculated hypothetical sample sizes (1:1 allocation ratio) based on the progression estimates from the LMEM, simulating a 50% reduction in SARA progression in an arbitrary parallel group, 2-year, 5-visit interventional trial with $\alpha = 0.05$ and 90% power.

Annualized progression rates (mean, standard error [SE], and 95% confidence interval [CI]) and sample sizes were calculated with SAS, version 9.4 (procedure MIXED, NLMIXED). SPSS 25 (IBM, Armonk, NY) was used to perform GLMs (function GENLIN). Correlations, curve fitting, descriptive, and confirmatory statistics were calculated with Prism 9 (GraphPad Software, La Jolla, CA), using nonparametric measures (Spearman rho, median, interquartile range (IQR)). MATLAB 2019b (The MathWorks, Natick, MA) was used to determine bootstrapped 95% CIs of Spearman correlations.

Results

Baseline Characteristics

We included 931 patients (51% female, median age at baseline = 36 years, range = 2–83; see Fig 1A, Fig 2) with genetically confirmed ARCA or genetically unsolved EOA highly enriched for ARCA of global origin, predominantly from Europe (n = 645), followed by Canada (n = 109), Asia (n = 96), and Africa (n = 47). The genetic cause of ataxia was identified in 623 of 931 patients (67%, 86 genotypes; see Fig 1A,B for frequency of "common ARCAs"). Disease onset in the genetically confirmed ARCA cases was before age 40 years in 94%, thus supporting the pragmatic cutoff of age at onset

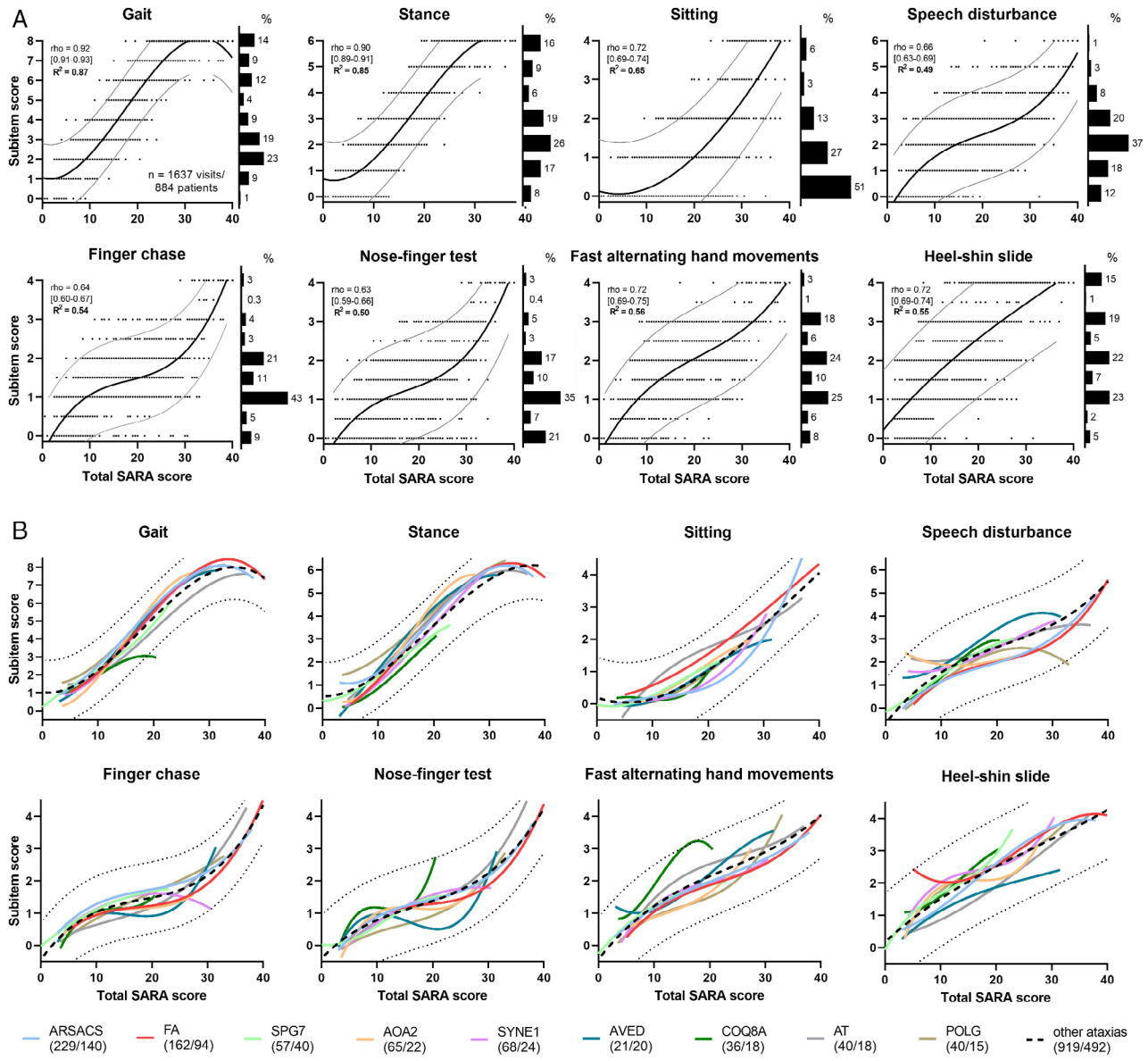

FIGURE 2: Responsiveness of Scale for the Assessment and Rating of Ataxia (SARA) subitems across ataxia severity and genotypes. Cross-sectional analysis of 1,637 SARA assessments in 884 patients. (A) Spearman correlation [95% confidence interval] between SARA subitem and total SARA (equal to overall ataxia severity). Response curves and 95% prediction intervals were estimated by third order polynomial curve fits. SARA subitems show ceiling effects (*gait*, *stance*), floor effects (*sitting*), and nonresponsive "static" periods at moderate ataxia severity (particularly *speech*, *finger-chase*, and *nose-finger*). Distribution of subitem score levels, depicted in the vertical bar charts next to each subitem, shows considerable ceiling effects (*gait*, *stance*) and floor effects (*sitting*), as well as under-representation of intermediate scores (*gait*, *stance*, and *sitting*, but also uneven scores for appendicular items) and high scores (*speech*, *finger-chase*, *nose-finger*). (B) Comparison of polynomial response curves between the "common autosomal recessive cerebellar ataxias" and all other ataxias in the cohort. Numbers in parentheses indicate total number of SARA assessments per number of patients. The relative contribution of each subitem to the total SARA is overall similar across ataxias, with genotype-specific differences particularly for *COQ8*-ataxia. AOA2 = ataxia with oculomotor apraxia type 2; ARSACS = autosomal recessive spastic ataxia of Charlevoix-Saguenay; AT = ataxia–telangiectasia; AVED = ataxia with vitamin E deficiency; FA = Friedreich ataxia; SPG7 = spastic paraplegia type 7. COQ8A/POLG/SYNE1 indicate ataxia related to the respective genes.

< 40 years chosen for the genetically still unsolved cohort. The cohort spanned the full range of ataxia severities (SARA from 3 to 40; see Fig 1B) across the common ARCAs, thus allowing for detailed metrological SARA analyses, including disease progression analysis.

Nonataxia Features Determine Ataxia Severity

ARCAs and EOAs were inherently multisystemic, with an INAS count of zero (indicating absence of noncerebellar involvement) in only 4% of patients (median INAS count = 5, IQR = 3–6). Nonataxia movement disorders

(73%; bradykinesia, rigidity, resting tremor, dystonia, myoclonus, and/or chorea/dyskinesia), pyramidal signs (63%; spasticity, extensor plantar response, and/or hyperreflexia), and signs of neuropathy (areflexia: 45%) were particularly prevalent nonataxia features (see Fig 1C for complete list). Cognitive impairment was present in 35%, with predominant childhood onset (median = 6 years, IQR = 2–24), and reportedly nonprogressive in 59%. Ataxia severity as assessed by the SARA was independently associated with several nonataxia features in both common ARCAs (n = 393, adjusted for genotype) and unsolved EOAs (n = 335). In patients with otherwise similar genotype, age at onset, and disease duration, SARA scores were higher with the presence of paresis (common ARCAs: 5.0 points/unsolved EOAs: 4.8 points), cognitive impairment (4.4/3.0 points), impaired vibration sense (3.6/3.9 points), urinary dysfunction (3.1/5.4 points), brainstem oculomotor signs (2.7/3.8 points), muscle atrophy (2.7/4.4 points), and spasticity (2.2/3.5 points; Table).

Responsiveness of SARA Subitems Depends on Ataxia Severity

Individual subitems of the SARA showed variable responsiveness across the ataxia severity range (see Fig 2A). *Gait* and *stance* covered a large range of severities (87% SARA scores < 25) with a linear scaling, including relatively granular progression steps of 8 and 6 metric levels, respectively, and relatively small, uniform prediction intervals even across the heterogeneity of ARCAs and EOAs in the cohort. In contrast, subitem *sitting* appeared only responsive for SARA scores > 10, but continued to increase with a linear slope until most advanced ataxia (SARA > 30), where *gait* and *balance* no longer contributed to responsiveness. For subitems *speech*, *finger-chase*, and *nose-finger*, responsiveness relative to the total SARA flattened in the range in which most patients were assessed (57% SARA scores = 10–30), indicating "static periods" in an important SARA range, with steep increases only beyond that range. Accordingly, these subitems only showed moderate correlations with the total SARA score (rho = 0.6–0.7), as compared to strong correlations of all other subitems (rho = >0.7–0.9). In contrast, *fast alternating hand movements* captured ataxia severity also within the range of SARA = 10–30, rendering this the most promising of all SARA upper limb subitems to capture appendicular upper limb ataxia. For appendicular lower limb ataxia, *heel-shin slide* was responsive across the full range of SARA scores, although with higher variability as compared to truncal ataxia items (*gait*, *stance*, *sitting*). The variable responsiveness between SARA subitems was not explained by their different scoring ranges alone, as exemplified by the absence versus presence of static periods in the comparison of subitems with the same 6-level (*stance* vs *speech*) or 4-level (*alternating hand movements* vs *finger-nose* or *nose-finger*) scoring range. It could also not be explained by a higher granularity on the item level: while the appendicular items principally allow for a larger granularity (equal to 9-level scoring range), this was effectively not used given the under-representation of half-point levels (see Fig 2A).

Despite the heterogeneity of ARCAs (including partly very different neural system damage, eg, paresis, spasticity, or sensory neuropathy), independent

TABLE. Nonataxia Features Associated with Ataxia Severity

	Common ARCAs, n = 393			Unsolved EOAs, n = 335		
Feature	**Beta**	**SE**	p_{adj}	**Beta**	**SE**	p_{adj}
Paresis	4.9912	0.8627	<0.0001	4.8052	0.9916	<0.0001
Cognitive impairment	4.4418	0.9927	<0.0001	2.9897	0.9058	0.0023
Vibration at ankle ≤ 5/8	3.6171	0.8125	<0.0001	3.9282	1.0253	0.0006
Urinary dysfunction	3.0997	0.7897	0.0006	5.4036	0.8950	<0.0001
Brainstem oculomotor sign (any)	2.6768	0.8190	0.0039	3.8300	0.8951	<0.0001
Muscle atrophy	2.6501	0.8135	0.0046	4.3656	1.0907	0.0005
Spasticity	2.1950	1.0034	0.0598	3.4616	0.9204	0.0008

Note: Generalized linear model with Scale for the Assessment and Rating of Ataxia score as dependent variable, and genotype, age at onset, ataxia duration, and ataxia duration squared as independent variables. Nonataxia features with a prevalence > 5% were individually entered into separate models as independent factor. Listed features were identified in both common genetically defined ARCAs as well as in a separate cohort of unsolved EOAs.
Abbreviation: ARCA = autosomal recessive cerebellar ataxia; Beta = regression coefficient; EOA = early onset ataxia; $p_{adj} = p$ values adjusted by Benjamini–Hochberg method; SE = standard error.

polynomial fits for each of the common ARCA genotypes as compared to all residual ataxias (noncommon ARCAs and unsolved EOAs) revealed a relatively uniform progression pattern for each subitem relative to the total SARA, particularly seen for *gait* and *stance* (see Fig 2B). A notable exception was *COQ8A*-ataxia; here, the contribution of individual subitems to the total SARA was lower for item *gait*, but higher for items *nose-finger* and *alternating hand movements* (both compared to the 95% prediction interval of other ataxias.) This might be due to the highly prevalent upper limb movement disorders in *COQ8A*-ataxia such as tremor, myoclonus, and dystonia.

Metric Limitations of SARA Responsiveness

The analysis of the distribution of SARA subitem scores across all 1,637 assessments (see Fig 2A), and of the longitudinal score level changes of each subitem across all 202 patients with a follow-up assessment after 1 year (mean = 12.8 months, range = 9–19), as a proxy for a standard trial duration (Sankey diagram, Fig 3), showed several metric limitations for the SARA responsiveness. Incomplete coverage of the full range of the scale was observed for all subitems. Intermediate score levels were under-represented in subitems *gait* (5 = "severe staggering, permanent support of one stick or light support by one arm required"), *stance* (4 = "able to stand for >10 seconds in natural position only with intermittent support"), and *sitting* (3 = "able to sit for > 10 seconds only with intermittent support"). Also half-point scores (reflecting potential asymmetric severity between left and right) were under-represented in all appendicular subitems. Upper score levels were under-represented in subitems *speech*, *finger-chase*, *nose-finger*, and *alternating hand movements*, where fewer than 1 to 6% of patients were rated with the highest score. Inconsistent transitions between score levels comprised the following: a substantial share of subitems with (neurologically unexpected) score decreases ("improvements") from baseline to follow-up (particularly for *speech*); fluctuating large changes of >1–2-point decreases ("improvements") or increases ("deteriorations") for subitems *finger-chase* and *nose-finger*; and nontransition between score levels, for example, from *speech* score 3 to 4 ("occasional" to "many words difficult to understand").

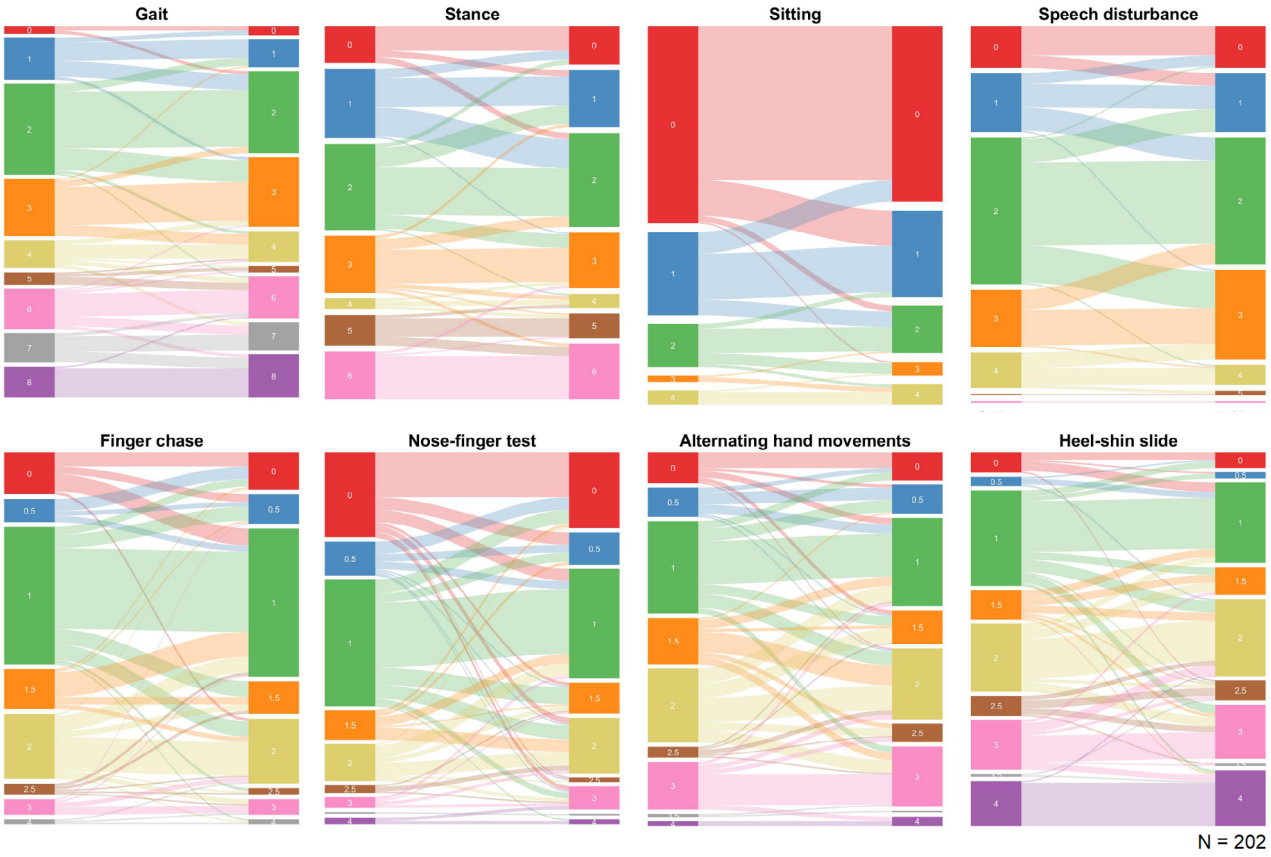

FIGURE 3: Distribution of the Scale for the Assessment and Rating of Ataxia (SARA) subitems cross-sectionally and in longitudinal progression. Sankey diagram of 202 patients with approximately annual follow-up is shown. In addition to the incomplete use of score ranges, SARA subitems with nonresponsive periods were characterized longitudinally by substantial score decrease ("improvement") for *speech* (scores 2 to 1, 3 to 2, and 4 to 3), and by volatile score decrease ("improvement") and increase ("deterioration") in appendicular items (eg, scores 0 to 2 in *finger-chase* and *nose-finger*, or scores 2 to 1 in all upper limb items). [Color figure can be viewed at www.annalsofneurology.org]

SARA Correlations with Patient-Focused Outcome

We next analyzed whether the limitations observed in the metrics of the SARA might be paralleled by low patient-focused functional relevance. To this end, we correlated the SARA and its subitems to FARS-ADL as a patient-focused anchor. The total SARA showed a strong correlation with FARS-ADL (rho = 0.85), indicating the capacity of this COA to capture everyday-relevant functional impairment across domains (Fig 4). Strong correlations were also observed for its subitems *sitting* and *fast alternating hand movements* (rho > 0.7, comparable to subitems *gait* and *stance)*, and moderate correlations were still found for subitems *speech*, *finger-chase*, and *heel-shin slide* (rho > 0.5). In contrast, the weak correlation of *nose-finger* with FARS-ADL in addition to its metric limitations suggests that this SARA subitem might be of overall limited added value for the SARA as a COA of ataxia.

Modeling of Longitudinal Progression

The SARA was generally able to capture progression of ataxia severity in the cohort, with overall mild-to-moderate annualized progression across all ARCAs/EOAs (LMEM = +0.60 [SE = 0.06] points/yr; Fig 5), but with substantial variability between genotypes, ataxia severities, and SARA subitems.

Disease Progression according to Genotypes. Annual SARA progression was highest in *POLG*-ataxia (1.56 [SE = 0.32]), FA (1.49 [SE = 0.19]), and ataxia with oculomotor apraxia type 2 (AOA2; 1.14 [SE = 0.24]). In contrast, SARA progression was only moderate in spastic

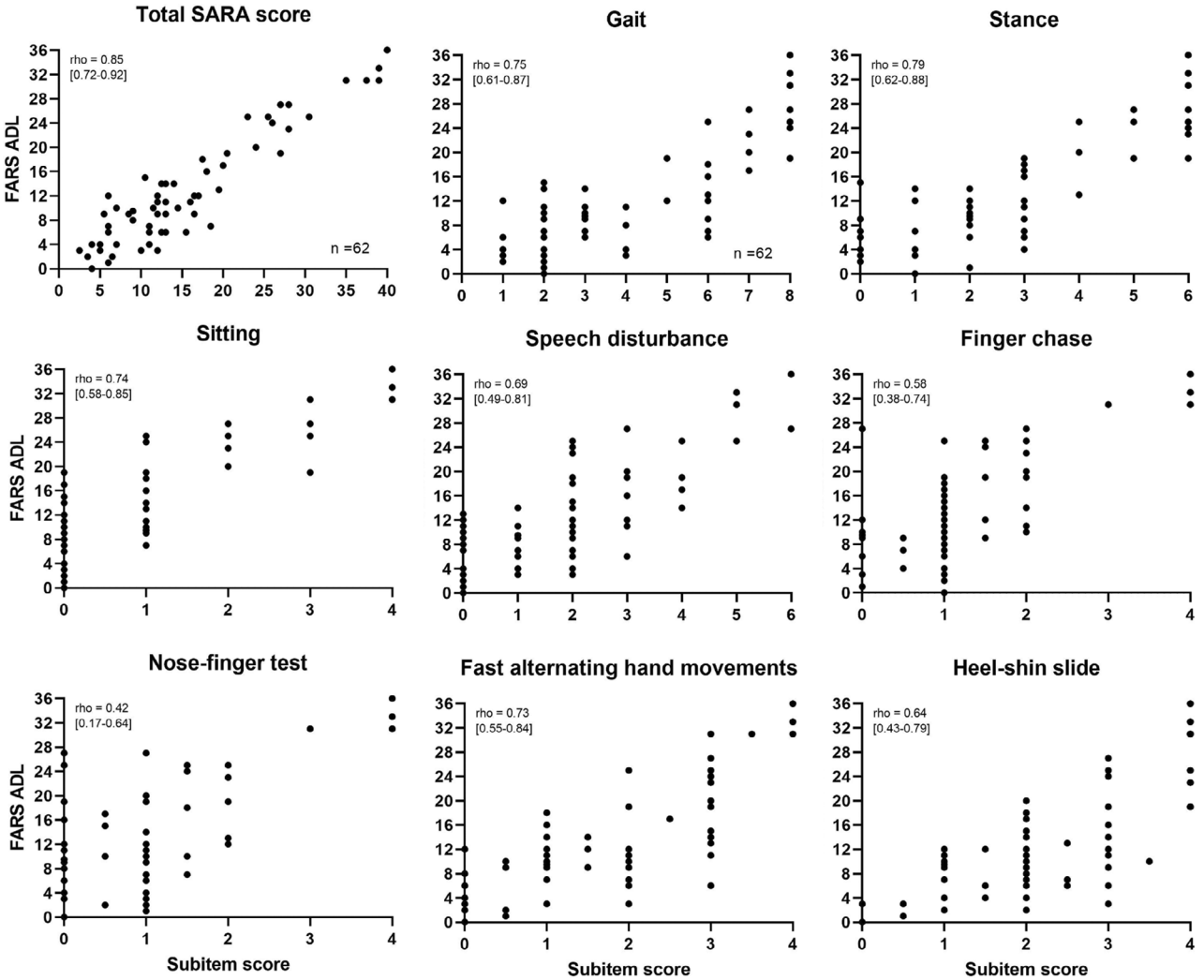

FIGURE 4: Correlation of the total Scale for the Assessment and Rating of Ataxia (SARA) and its subitems with Friedreich Ataxia Rating Scale (FARS) activities of daily living (ADL). Spearman correlations [95% confidence intervals] are shown. The total SARA and all SARA subitems—except *nose-finger*—show at least moderate correlations with FARS-ADL, which can be considered a clinically meaningful endpoint.

paraplegia type 7 (SPG7; 0.57 [SE = 0.41]), *SYNE1*-ataxia (0.55 [SE = 0.26]), and ataxia–telangiectasia (AT; 0.62 [SE = 0.38]). No change within 1 year was captured by the SARA in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS; 0.12 [SE = 0.15]) and *COQ8A*-ataxia (−0.06 [SE = 0.41]; see Fig 5).

Disease Progression according to Ataxia Severity. Ataxia severity was binned by baseline SARA (mild: <10, moderate: 10–25, advanced: >25) and by ambulatory status as critical ataxia milestone (baseline SARA *gait* > 6 as proxy for loss of ambulation). This pragmatic binning was supported by the FARS-FS, which was available for a subset of assessments, and which differed between mild (n = 58, median = 2 [IQR = 1.5–2.5]), moderate (n = 101, median = 3 [IQR = 2–4]), and severe ataxia (n = 27, median = 5 [IQR = 4.5–6]; Kruskal–Wallis test with Dunn multiple comparisons: all $p < 0.001$), as well as between ambulatory

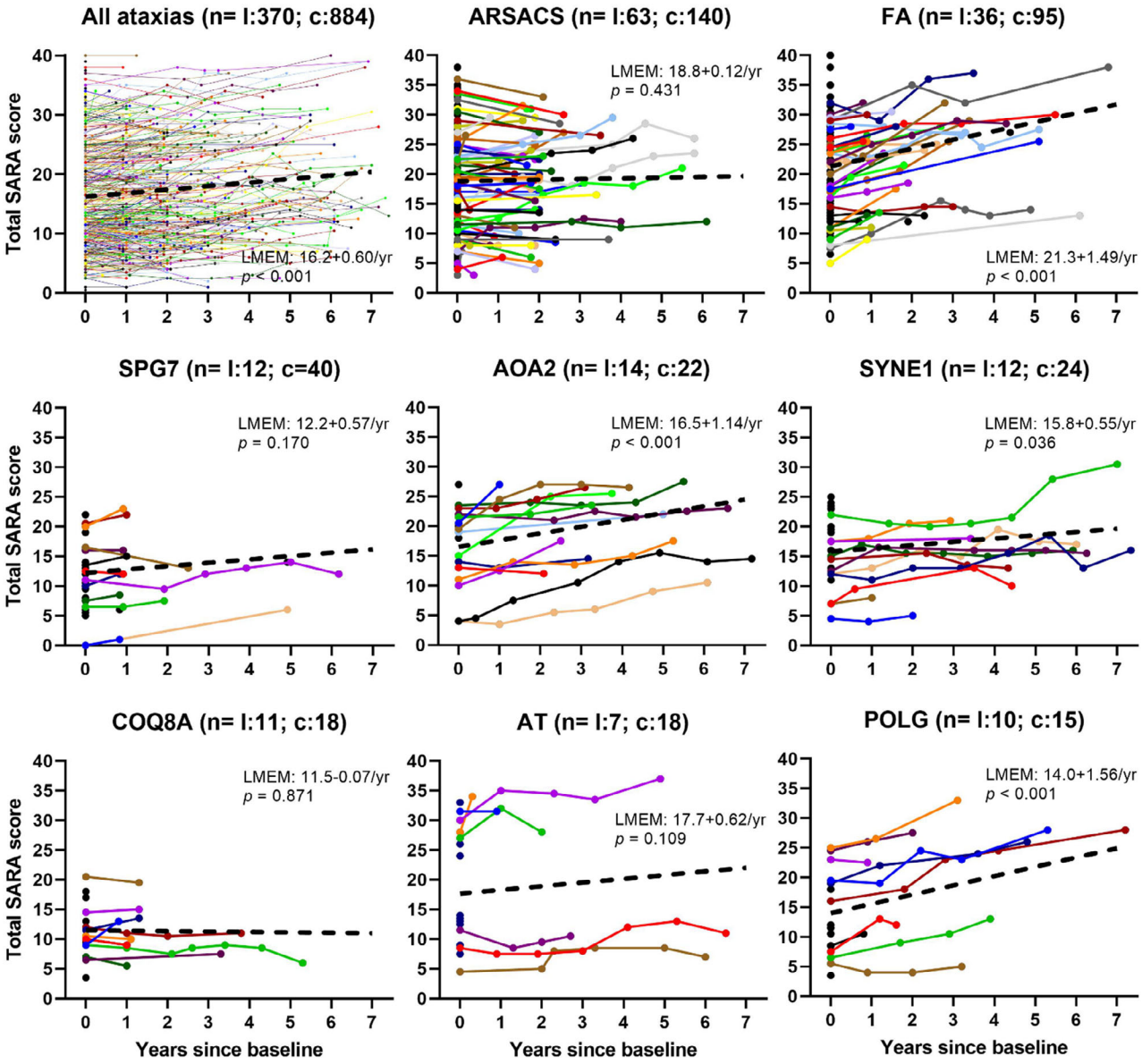

FIGURE 5: Annualized progression of the Scale for the Assessment and Rating of Ataxia (SARA) in the total cohort and the most common autosomal recessive cerebellar ataxias. Dashed lines and corresponding equations present linear mixed effects models (LMEMs) with random intercept and slope; *p* values < 0.05 indicate progression significantly different from zero. I: number of subjects with longitudinal data (colored lines), c: total number of subjects including cross-sectional baseline data without follow-up (black dots). The SARA captures change in Friedreich ataxia (FA), ataxia with oculomotor apraxia type 2 (AOA2), *SYNE1*, and *POLG*, but does not show change in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) or *COQ8A*-ataxia within 1 year. Intercepts of baseline SARA scores can be interpreted as expectable distributions of patients' ataxia severity for trial planning. Ataxia with vitamin E deficiency is not shown because only one patient had longitudinal data. AT = ataxia–telangiectasia; SPG7 = spastic paraplegia type 7. COQ8A/POLG/SYNE1 indicate ataxia related to the respective genes.

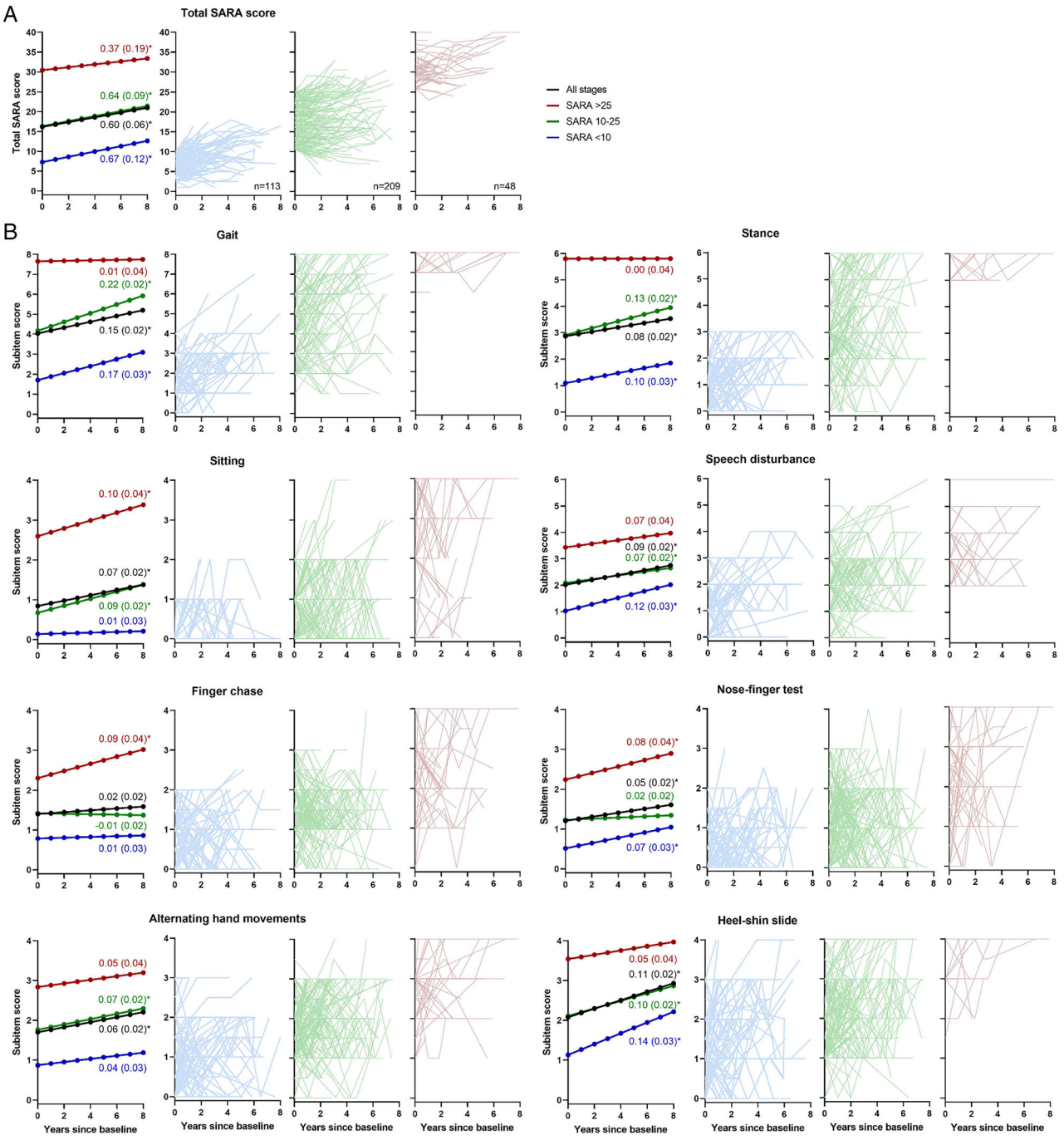

FIGURE 6: Sensitivity to change of the total Scale for the Assessment and Rating of Ataxia (SARA) and its subitems by ataxia severity. Raw data and mean (standard error) estimated annualized progression are based on linear mixed effects modeling of all longitudinal data in the full cohort. Results are shown aggregated across all disease stages as well as separate for mild (SARA < 10), moderate (SARA = 10–25), and advanced (SARA > 25) ataxia. Asterisks mark models with sensitivity to change. (A) Sensitivity of the SARA decreases in advanced ataxia due to smaller progression and higher variability. (B) The SARA subitems *gait*, *stance*, *speech*, and *heel-shin* show sensitivity to change in mild and moderate ataxia (ie, SARA ≤ 25), thus likely driving the sensitivity to change of the total SARA in these disease stages. In advanced ataxia (ie, SARA > 25), *sitting*, *finger-chase*, and *nose-finger* show sensitivity to change, thus here likely driving the changes in the total SARA. Note the marked intraindividual variability in the longitudinal trajectory of patients, where a biological progression, that is, progressive worsening, is expected for all underlying ataxias, irrespective of genotype. [Color figure can be viewed at www.annalsofneurology.org]

(n = 149, median = 2.5 [IQR = 2–3.5]) and nonambulatory bins (n = 38, median = 5 [IQR = 5–5.1], Mann–Whitney test: p < 0.001). Sensitivity to change of the SARA was similar in mild (0.67 [0.12]) and moderate ataxia (0.64 [0.09]), whereas progression decreased and variability increased in advanced ataxia (0.37 [0.19]; Fig 6A). Accordingly, progression

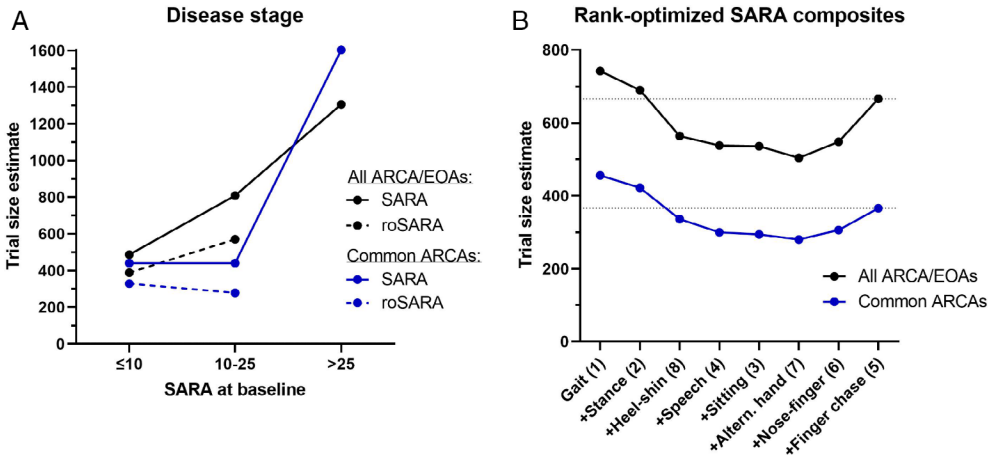

FIGURE 7: Sample size estimates for the Scale for the Assessment and Rating of Ataxia (SARA) and a rank-optimized SARA (roSARA) relative to ataxia severity (A) and for rank-optimized SARA composites (B). Simulations illustrate the relative impact of altered sensitivity on hypothetical sample sizes in trial scenarios, estimated for the detection of 50% reduced progression in a 2-year 5-visit trial with a power of 90%. (A) Sample size estimates for the SARA depend on ataxia severity, with sharp increases (about 3-fold) in advanced ataxia. The roSARA (dashed line) shows an improved sensitivity compared to the classical SARA (uninterrupted line), both for the full cohort (equal to all autosomal recessive cerebellar ataxias [ARCAs]/early onset ataxias [EOAs]; black) and for the subgroup of the most common ARCAs stratified by genotype (blue). Sample size for roSARA was calculated only for the trial-relevant mild (SARA < 10) and moderate (SARA = 10–25) disease stage; data were insufficient for modeling the advanced stage. (B) Sensitivity of a series of rank-optimized SARA composites with successive step-by-step inclusion from the most (left side: *gait*) to least sensitive subitem (right side: *finger-chase*), calculated for baseline SARA ≤ 25 points. Trial sizes would decrease by 20 to 30% if subitems *nose-finger* and *finger-chase* were omitted. Note the comparable sample sizes between the total SARA (dashed lines) and a composite of only 2 or 3 most optimal subitems. [Color figure can be viewed at www.annalsofneurology.org]

decreased, and variability increased when ataxia severity was stratified by loss of ambulation (baseline SARA *gait* > 6: 0.34 [0.18] vs. SARA *gait* ≤ 6: 0.63 [0.07]).

Disease Progression of SARA Subitems. Overall, all SARA subitems except *finger-chase* were sensitive to change, with the largest annual progression for *gait*, followed by *heel-shin*, *speech*, and *stance* (see Fig 6B). All of these 4 most responsive items consistently lost sensitivity to change in advanced ataxia (SARA > 25), whereas subitem *sitting* was not sensitive to change in mild ataxia (SARA < 10; see Fig 6B). Among upper limb subitems, *alternating hand movements* was only sensitive to change in moderate ataxia, whereas *finger-chase* and *nose-finger* were sensitive in advanced ataxia despite their metric limitations. When stratified by ambulatory status, all SARA subitems except *finger-chase* were sensitive to change in ambulatory patients (SARA *gait* ≤ 6), whereas sensitivity to change was limited to subitems *sitting* (0.10 [0.04]), *finger-chase* (0.09 [0.04]), and *nose-finger* (0.11 [0.04]) in nonambulatory patients (SARA *gait* > 6).

Rank Optimization of SARA for Clinical Trials

Based on the annualized progression estimated by LMEM, we performed hypothetical sample size calculations as a methodological tool to test and illustrate the sensitivity of different SARA composites in trial scenarios (rather than for use in actual trials, which will likely be conducted in a disease−/genotype-specific fashion, except for symptomatic ataxia drugs). As use cases, we analyzed the sensitivity of the SARA across the full cohort and the subset of common ARCAs, which allowed characterization and optimization of its sensitivity as an overall, generic COA of ataxia (rather than for a particular genotype). Compared to a trial in ARCA/EOA patients with mild ataxia (SARA < 10, N = 486), sample size increased 1.7-fold (N = 808) when applying the SARA as COA in moderate ataxia (SARA = 10–25), and 2.7-fold (N = 1,306) in advanced ataxia (SARA > 25; Fig 7A). In the common ARCAs, estimated sample sizes were similar in early and moderate ataxia (both N = 440), but increased 3.6-fold in advanced ataxia (N = 1,602).

We next explored whether sensitivity of the SARA could be optimized by including (and omitting) subitems according to their ranked annual progression rate, here focusing on the most trial-relevant severity range of ataxia, namely mild-to-moderate ataxia. Successive inclusion of the 6 most responsive subitems (*gait*, *stance*, *heel-shin*, *speech*, *sitting*, *alternating hand movements*) into such a generic rank-optimized SARA (roSARA), calculated across all ARCA/EOA patients with a baseline SARA ≤ 25, led to a continuous decrease in sample size from N = 742 to N = 504, that is, yielding a reduction of trial size by 18% for the roSARA as compared to total SARA (see Fig 7B).

Adding *nose-finger* and *finger-chase* to the roSARA, that is, simulating all items of the SARA (equal to total SARA), worsened sensitivity (N = 666), with a suboptimal level comparable to the sensitivity of only the 2 or 3 most responsive items. Similar results were obtained for the roSARA in the common ARCAs with a baseline SARA ≤ 25, with a reduction of trial size by 24%.

Discussion

To facilitate trial planning and regulatory qualification, this study comprehensively characterizes responsiveness of the SARA as generic COA for ataxia, harnessing a large prospective genetic ataxia cohort as an exemplary showcase cohort for multisystemic ataxias, and provides first natural history data for a wide range of ataxias.

Responsiveness of the SARA

As a major finding, our subitem-level metric analysis delineated 3 types of limitations that hamper responsiveness of the SARA. First, our correlation-based analysis showed limited responsiveness in several subitems and disease stages, with varying patterns: ceiling effects for *gait* and *stance*, floor effects for *sitting*, and nonresponsive ("static") periods at moderate stages particularly for *speech*, *finger-chase*, and *nose-finger*. These findings extend partly similar recent observations in a monocentric FA[26] and multicentric SCA3 cohort,[15] but now demonstrate these SARA properties for a more comprehensive set of items in a multicenter cohort and in particular across a large number of genetic ataxias, that is, for the SARA as a generic COA across ataxias. Second, our distribution-based analysis identified incomplete coverage ("use") of the full-scale range in almost all subitems. Under-representation of SARA *gait* score = 5 may reflect the patients' need for better stabilization by bilateral walking aids or a stroller after they lose free ambulation.[27] For *stance* and *sitting*, we hypothesize that "intermittent support" (score = 4 and 3, respectively) is barely needed in a static open-eye balance task of only 10-second duration. The under-representation of upper scores in *speech* and upper limb subitems indicate that they are biologically rare (in particular in the milder ataxia types) and/or that such advanced patients are no longer seen at ataxia referral centers.

As a third type of limitation, our analysis of longitudinal assessments—with 1-year intervals, reflecting the most trial-relevant time interval—revealed nonprogression of SARA subitems as major cause of their nonresponsive ("static") periods. *Speech* was nonresponsive due to nonprogression from "occasional" (score = 3) to "many words difficult to understand" (score = 4), or even substantial improvement at follow-up (37% and 41% of patients from score = 3 and 4, respectively). This may reflect interfering speech therapy,[28] rater bias (eg, listener experience), and/or effective compensatory mechanisms, especially because assessment of speech in the SARA is based on intelligibility, and on "normal conversation," which may become verbally scarce and simplified in advanced ataxia. The *speech* subitem of the SARA may benefit from standardized speech tasks with a minimum quantity and complexity of speech production (eg, reading task or syllable/sentence repetition, as used for outcome assessment in ataxia speech trials[28,29]). For upper limb subitems of the SARA, we demonstrate fluctuating score decreases ("improvements") and increases as major cause of their nonprogression. We hypothesize that this variability of upper limb subitems is caused by the variable speed–accuracy trade-off that a patient is still free to select despite task instructions ("as fast *and* as precise as possible," "at *moderate* speed") and by difficulties of the rater to visually estimate spatial deviations (dysmetria, tremor amplitude) in an objective and reproducible manner.[30] Subitem *nose-finger* may additionally suffer from an incomplete definition of "kinetic tremor"; it might be interpreted by clinical raters as "*rhythmic*, oscillatory movement" according to movement disorders criteria[31] (consistent with intention tremor), or as any irregularity due to decomposition and dysmetria of arm movement, as rated in the International cooperative ataxia rating scale.[32]

Patient-Focused Functional Relevance

Regulatory agencies have expressed concerns not only on the intraindividual and subitem variability of the SARA, but also on the functional relevance of its individual subitems.[19] Our findings from correlation analysis of SARA with FARS-ADL help to inform this discussion. First, the moderate-to-strong correlations of all SARA subitems—except *nose-finger*—with FARS-ADL indicate that all of these subitems might capture everyday functions in ataxia, given that FARS-ADL captures complex functional impairment across multiple domains.[22] In particular, the high degree of correlation of SARA *gait* and *stance* with FARS-ADL indicates that they reflect meaningful items in everyday living, which may explain their use as primary outcome or predictors in clinical trials as well as their minimal placebo effects, as suggested earlier for FA,[33–35] and now shown here across multiple ataxia genotypes. Second, the discrepancy between poor sensitivity to change and good correlations with FARS-ADL of several SARA subitems indicates that metric problems, not content validity of the respective item, impair responsiveness. This discrepancy was particularly striking for subitem *alternating hand movements*; its strong correlation with FARS-ADL (rho = 0.73; comparable to *gait*) suggests that this motor

task captures upper limb ataxia functionally meaningful to patients. This item, like *speech* or *heel-shin*, might thus not be discarded from the SARA, but rather needs improvement in task design and/or scoring to increase its metric properties. Although FARS-ADL reflects functional impairment also in patient-reported outcome measures of ataxia,[36] it remains unclear to what extent correlations between SARA and FARS-ADL reflect functional meaningfulness or rather overall disease severity of a shared construct, namely, ataxia. Here, item-specific correlations between SARA and FARS-ADL in larger cohorts specifically powered for this question will help to better characterize the functional relevance of SARA items and domains.

Longitudinal Natural History Data

Our study provides natural history data for several ARCAs based on the SARA, informing on annualized progression for trial size estimations (mean and variability of slope) and ataxia severities that need to be expected upon trial planning (distribution of intercepts). Specifically, we provide for the first time natural history data for AOA2, *SYNE1*-ataxia, and AT. Albeit preliminary, given the limited number of observations and heterogeneous time courses, such data are urgently required for trial planning, as treatments are on the horizon for several of these ARCAs, either for the whole ataxia disease type[37] or for individual patients thereof susceptible to individualized genetic treatments.[38] Our longitudinal data partly confirm, partly revise earlier results on ARCA progression estimates from a smaller, European, and cross-sectional only cohort.[39] The progression rates of 1.56 SARA points/yr in *POLG*-ataxia and 0.57 SARA points/yr in SPG7 corroborate and extend earlier findings in these ataxias.[40,41] In ARSACS, where responsiveness of the SARA in annual intervals is controversial,[14,42] our data suggest that even without ceiling effects, the SARA may not be sensitive to change over 1 year. Larger natural history studies accounting, for example, for different mutations (c.8844delT *SACS* founder mutation vs other mutations) are needed to address this question.

Modeling Longitudinal Progression

As one of the most important findings, our progression and sample size estimations demonstrate that using the SARA as COA in future trials not only is affected by nonataxia features, but will particularly benefit from specifically considering ataxia severity and subitem composition. Regarding ataxia severity, the total SARA performed best in mild ataxia (baseline SARA < 10). This is important, as it is this disease stage where disease-modifying therapies might likely be most effective, and which thus represents the disease stratum mainly enrolled in current treatment trials.[43] In moderate ataxia (SARA = 10–25), sensitivity of the SARA was equal to mild ataxia only after stratification for ataxia genotype. This reflects heterogeneity across genetic ataxias at this stage, and indicates that genotype-specific natural history studies may be particularly recommended before applying the SARA as a trial COA to this stage. In advanced ataxia (SARA > 25), a relative 2- to 3-fold increase in trial size—irrespective of genetic stratification—renders the use of the SARA as primary COA in rare genetic ataxias practically impossible. For these patients—almost 15% of our cohort—better COAs, or acceptance of other, nonclinical outcome measures, are thus needed, for example, fluid markers such as neurofilament light chain (with reductions of sample size in SCA3 by possibly up to a factor of 40)[3,44] or digital-motor measures (with reduction by possibly up to a factor of 6).[45,46]

Rank-Optimized SARA

An improved version of the SARA might be established by a generic ranking and selection of only the most responsive SARA items, across ataxia types (roSARA). Our study demonstrates that *finger-chase* and *nose-finger* are not beneficial or "neutral" subitems, but detrimental to the performance of the SARA as COA. We show that a data-driven roSARA that omits these two upper limb subitems allows reduction of trial size by up to 25%, and that a composite of only the 3 or 4 most responsive subitems may appear superior to the total SARA, but still have suboptimal sensitivity. This finding extends recent findings limited to SCA3[15] to a large across-genotype cohort, thus demonstrating their general applicability to genetic ataxias. In combination with the suggested metric optimizations, such an roSARA might help to facilitate its regulatory qualification for upcoming trials.[47]

Limitations of the Study

In line with the concept and development of the SARA,[1] our across-genotype analyses assume a unified core construct of ataxia underlying the large overall data aggregate. Although we controlled for genotype in the common ARCA cohort, key findings that could not be analyzed on the genotype-specific level (eg, the added value of the roSARA) would warrant further validation in larger cohort studies of the respective genetic ataxia type. However, our study is fully in line with smaller genotype-specific studies, for example, in FA[26] or SCA3,[15] which focused on some of the aspects now identified here as part of a larger set of findings with general applicability across genetic ataxias. Moreover, for many of these ultrarare ataxias, genotype-specific validation will not be realistic in the near future.

Taking ARCAs as a whole group allows delineation of a distribution range of possible disease trajectories across all ARCAs, including the rarest ones. This can be used for estimating and narrowing the probability distributions of disease trajectories even for ataxias where no within-genotype natural history studies can be done, using it, for example, as prior for Bayesian trajectory estimations.[48]

This study is also limited in that it resorts to multicenter registry rather than clinical trial data. Although this adds data variability, such use of real-world data allows application of the findings reported here to real-world applications of SARA as COA.[49] Finally, our sample size calculations aimed to compare the relative sensitivity of the SARA between ataxia severities and SARA composites as a generic, that is, across-genotype COA for ataxia. Although this allows illustration of the COA properties of the SARA for ataxia as a core construct across the multitude of genetic ataxias, the heterogeneity of underlying ataxia types is expected to increase sample sizes, and extrapolation to genotype-specific treatment trials has to be interpreted with caution. However, in the absence of treatment trial data for literally all ataxias analyzed here (except FA), our sample size calculations provide at least a starting point allowing estimation of relative sample size magnitudes for upcoming trial planning. As more data become available in rigorously performed trial-like natural history studies (eg, ARSACS, SPG7, or RFC1)[50,51] and future treatment trials, this will allow validation and further refinement of the findings of the present study in disease-specific trial contexts.

Acknowledgments

This work was supported via the European Union's Horizon 2020 research and innovation program as part of the innovation project EVIDENCE-RND under the EJP RD COFUND-EJP (No. 825575, to R.-D.H. and M.S.), as part of Solve-RD (No. 779257, to J.B., M.S., and B.P.v.d.W.), by the DFG under the framework of EJP-RD network PROSPAX (No. 441409627; M.S., B.P.v.d.W.), and by the Clinician Scientist program PRECISE.net, funded by the Else Kröner-Fresenius-Stiftung (to A.Tr. and M.S.). The study was further funded by the Federal Ministry of Education and Research, Germany, and through the TreatHSP network (01GM1905 to L.S.). B.P.v.d.W. receives additional research support from ZonMW, Hersenstichting, Gossweiler Foundation, and Radboud University Medical Center. L.S., T.Klop., T. Kloc., G.Z., B.P.v.d.W., and M.S. are members of the European Reference Network for Rare Neurological Diseases–Project ID No. 739510. A.Tr. receives funding from the University of Tübingen, medical faculty, for the Clinician Scientist Program Grant #439-0-0. P.K. receives funding from University of Szeged (Hetényi Géza: 5S330 A202) and Ministry of Innovation and Technology of Hungary, National Research, Development and Innovation Fund (TKP2021-EGA). J.B. is supported by a Senior Clinical Researcher mandate of the Research Fund–Flanders (FWO) under grant agreement number 1805021N, and is a member of the μNEURO Research Center of Excellence of the University of Antwerp. F.M.S. is supported in part by the Italian Ministry of Health (EJP-RD network PROSPAX; Ricerca Finalizzata RF-2016-02361610; RF-2019-12370417; Ricerca Corrente RC 5x1000). Several authors of this publication are members of the European Reference Network for Rare Neuromuscular Diseases and of the European Reference Network for Rare Neurological Diseases.

We are grateful to S. Reich for her coordinating support as part of the EOA/PREPARE consortium and to T. Heger for monitoring the datasets of the ARCA registry. Open Access funding enabled and organized by Projekt DEAL.

Author Contributions

A.Tr., R.-D.H., and M.S. contributed to the conception and design of the study. All authors contributed to the acquisition and analysis of data. A.Tr. and M.S. contributed to drafting the text and preparing the figures.

Potential Conflicts of Interest

Nothing to report.

References

1. Schmitz-Hübsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 2006;66:1717–1720.

2. Perez-Lloret S, van de Warrenburg B, Rossi M, et al. Assessment of ataxia rating scales and cerebellar functional tests: critique and recommendations. Mov Disord 2021;36:283–297.

3. Jacobi H, du Montcel ST, Bauer P, et al. Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study. Lancet Neurol 2015;14:1101–1108.

4. Reetz K, Dogan I, Hilgers R-D, et al. Progression characteristics of the European Friedreich's ataxia consortium for translational studies (EFACTS): a 4-year cohort study. Lancet Neurol 2021;20:362–372.

5. Coarelli G, Heinzmann A, Ewenczyk C, et al. Safety and efficacy of riluzole in spinocerebellar ataxia type 2 in France (ATRIL): a multicentre, randomised, double-blind, placebo-controlled trial. Lancet Neurol 2022;21:225–233.

6. Lei L-F, Yang G-P, Wang J-L, et al. Safety and efficacy of valproic acid treatment in SCA3/MJD patients. Parkinsonism Relat Disord 2016;26:55–61.

7. Giordano I, Bogdanow M, Jacobi H, et al. Experience in a short-term trial with 4-aminopyridine in cerebellar ataxia. J Neurol 2013;260: 2175–2176.
8. Miyai I, Ito M, Hattori N, et al. Cerebellar ataxia rehabilitation trial in degenerative cerebellar diseases. Neurorehabil Neural Repair 2012; 26:515–522.
9. França C, de Andrade DC, Silva V, et al. Effects of cerebellar transcranial magnetic stimulation on ataxias: a randomized trial. Parkinsonism Relat Disord 2020;80:1–6.
10. Feil K, Adrion C, Boesch S, et al. Safety and efficacy of acetyl-DL-leucine in certain types of cerebellar ataxia: the ALCAT randomized clinical crossover trial. JAMA Netw Open 2021;4:e2135841-e.
11. Benussi A, Dell'Era V, Cantoni V, et al. Cerebello-spinal tDCS in ataxia: a randomized, double-blind, sham-controlled, crossover trial. Neurology 2018;91:e1090–e1101.
12. Weyer A, Abele M, Schmitz-Hübsch T, et al. Reliability and validity of the scale for the assessment and rating of ataxia: a study in 64 ataxia patients. Move Disord 2007;22:1633–1637.
13. Bürk K, Mälzig U, Wolf S, et al. Comparison of three clinical rating scales in Friedreich ataxia (FRDA). Mov Disord 2009;24:1779–1784.
14. Bourcier D, Bélanger M, Côté I, et al. Documenting the psychometric properties of the scale for the assessment and rating of ataxia to advance trial readiness of autosomal recessive spastic ataxia of Charlevoix-Saguenay. J Neurol Sci 2020;417:117050.
15. Maas R, Teerenstra S, Lima M, et al. Differential temporal dynamics of axial and appendicular ataxia in SCA3. Mov Disord 2022;37:1850–1860.
16. Klockgether T, Ashizawa T, Brais B, et al. Paving the way toward meaningful trials in ataxias: an ataxia global initiative perspective. Mov Disord 2022;37:1125–1130.
17. Biohaven Provides Update on Phase 3 Clinical Trial Evaluating Troriluzole for Spinocerebellar Ataxia (SCA). [August 22, 2022]; Available from: https://www.biohavenpharma.com/investors/news-events/press-releases/05-23-2022.
18. Grobe-Einsler M, Taheri Amin A, Faber J, et al. Development of SARAhome, a new video-based tool for the assessment of ataxia at home. Mov Disord 2020.
19. Troriluzole in Adult Subjects With Spinocerebellar Ataxia. [July 25, 2022]; Available from: https://clinicaltrials.gov/ct2/show/NCT03701399.
20. Traschütz A, Reich S, Adarmes AD, et al. The ARCA registry: a collaborative global platform for advancing trial readiness in autosomal recessive cerebellar ataxias. Front Neurol 2021;12:12.
21. Jacobi H, Rakowicz M, Rola R, et al. Inventory of non-ataxia signs (INAS): validation of a new clinical assessment instrument. Cerebellum 2013;12:418–428.
22. Subramony S, May W, Lynch D, et al. Measuring Friedreich ataxia: interrater reliability of a neurologic rating scale. Neurology 2005;64: 1261–1262.
23. Gama MTD, Braga-Neto P, Rangel DM, et al. Autosomal recessive cerebellar ataxias in South America: a multicenter study of 1338 patients. Mov Disord 2022;37:1773–1774.
24. Coutelier M, Hammer MB, Stevanin G, et al. Efficacy of exome-targeted capture sequencing to detect mutations in known cerebellar ataxia genes. JAMA Neurol 2018;75:591–599.
25. Schüle R, Wiethoff S, Martus P, et al. Hereditary spastic paraplegia: clinicogenetic lessons from 608 patients. Ann Neurol 2016;79: 646–658.
26. Pandolfo M. Neurologic outcomes in Friedreich ataxia: study of a single-site cohort. Neurol Genet 2020;6:e415.
27. Lessard I, St-Gelais R, Hébert LJ, et al. Functional mobility in walking adult population with ataxia of Charlevoix-Saguenay. Orphanet J Rare Dis 2021;16:1–10.
28. Vogel AP, Graf LH, Magee M, et al. Home-based biofeedback speech treatment improves dysarthria in repeat-expansion SCAs. Ann Clin Transl Neurol 2022;9:1310–1315.
29. Vogel AP, Stoll LH, Oettinger A, et al. Speech treatment improves dysarthria in multisystemic ataxia: a rater-blinded, controlled pilot-study in ARSACS. J Neurol 2019;266:1260–1266.
30. Ladavas E, Serino A. Action-dependent plasticity in peripersonal space representations. Cogn Neuropsychol 2008;25:1099–1113.
31. Bhatia KP, Bain P, Bajaj N, et al. Consensus statement on the classification of tremors. From the task force on tremor of the International Parkinson and Movement Disorder Society. Mov Disord 2018;33: 75–87.
32. Trouillas P, Takayanagi T, Hallett M, et al. International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. J Neurol Sci 1997;145:205–211.
33. Rummey C, Corben LA, Delatycki M, et al. Natural history of Friedreich's ataxia: heterogeneity of neurological progression and consequences for clinical trial design. Neurology 2022;99:e1499–e1510.
34. Rummey C, Farmer JM, Lynch DR. Predictors of loss of ambulation in Friedreich's ataxia. EClinicalMedicine 2020;18:100213.
35. Meier T, Perlman SL, Rummey C, et al. Assessment of neurological efficacy of idebenone in pediatric patients with Friedreich's ataxia: data from a 6-month controlled study followed by a 12-month open-label extension study. J Neurol 2012;259:284–291.
36. Schmahmann JD, Pierce S, MacMore J, L'Italien GJ. Development and validation of a patient-reported outcome measure of ataxia. Mov Disord 2021;36:2367–2377.
37. Synofzik M, Puccio H, Mochel F, Schöls L. Autosomal recessive cerebellar ataxias: paving the way toward targeted molecular therapies. Neuron 2019;101:560–583.
38. Synofzik M, van Roon-Mom WM, Marckmann G, et al. Preparing n-of-1 antisense oligonucleotide treatments for rare neurological diseases in Europe: genetic, regulatory, and ethical perspectives. Nucleic Acid Ther 2022;32:83–94.
39. Traschütz A, Adarmes AD, Anheim M, et al. Autosomal Recessive Cerebellar Ataxias in Europe: Frequency, onset and severity in 677 patients. Mov Disord. 2023.
40. Coarelli G, Schule R, van de Warrenburg BP, et al. Loss of paraplegin drives spasticity rather than ataxia in a cohort of 241 patients with SPG7. Neurology 2019;92:e2679–e2690.
41. Bender F, Timmann D, van de Warrenburg BP, et al. Natural history of polymerase gamma–related ataxia. Mov Disord 2021;36:2642–2652.
42. Gagnon C, Lessard I, Lavoie C, et al. An exploratory natural history of ataxia of Charlevoix-Saguenay: a 2-year follow-up. Neurology 2018;91:e1307–e1311.
43. A Pharmacokinetics and Safety Study of BIIB132 in Adults With Spinocerebellar Ataxia 3. [August 2, 2022]; Available from: https://clinicaltrials.gov/ct2/show/NCT05160558.
44. Wilke C, Haas E, Reetz K, et al. Neurofilaments in spinocerebellar ataxia type 3: blood biomarkers at the preataxic and ataxic stage in humans and mice. EMBO Mol Med 2020;12:e11803.
45. Corben LA, Nguyen KD, Pathirana PN, et al. Developing an instrumented measure of upper limb function in Friedreich ataxia. Cerebellum 2021;5:1–9.
46. Ilg W, Müller B, Faber J, et al. Digital gait biomarkers allow to capture 1-year longitudinal change in spinocerebellar ataxia type 3. Mov Disord 2022;37:2295–2301.
47. Klockgether T, Ashizawa T, Brais B, et al. Paving the way toward meaningful trials in ataxias: an ataxia global initiative perspective. Mov Disord. 2022;37:1125–1130.

48. Fouarge E, Monseur A, Boulanger B, et al. Hierarchical Bayesian modelling of disease progression to inform clinical trial design in centronuclear myopathy. Orphanet J Rare Dis 2021;16:1–11.

49. European Medical Agency: Guideline on registry-based studies. [January 6, 2021]; Available from: https://www.ema.europa.eu/en/guideline-registry-based-studies.

50. Phenotypes, Biomarkers and Pathophysiology in Spastic Ataxias (SPAX-PBP). [August 2, 2022]; Available from: https://clinicaltrials.gov/ct2/show/NCT04297891.

51. RFC1 Natural History Study (RFC1-NHS). [August 2, 2022]; Available from: https://clinicaltrials.gov/ct2/show/NCT05177809.