ANNALS of Clinical and Translational Neurology
Open Access
AMERICAN NEUROLOGICAL ASSOCIATION

RESEARCH ARTICLE

Digital endpoints for self-administered home-based functional assessment in pediatric Friedreich's ataxia

Arne Mueller[1], Elaine Paterson[2], Avery McIntosh[3], Jens Praestgaard[3], Mary Bylo[2], Holger Hoefling[4], McKenzie Wells[5], David R. Lynch[5], Christian Rummey[6], Michelle L. Krishnan[1], Meredith Schultz[2] & C. J. Malanga[2]

[1]Translational Medicine, Novartis Institutes for BioMedical Research, Basel, Switzerland
[2]Translational Medicine, Novartis Institutes for BioMedical Research, Cambridge, Massachusetts, USA
[3]Novartis Development, Cambridge, Massachusetts, USA
[4]NIBR Informatics, Novartis Institute of Biomedical Research, Basel, Switzerland
[5]Division of Neurology, Departments of Neurology and Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA
[6]Clinical Data Science GmbH, Basel, Switzerland

Correspondence
Arne Mueller, Translational Medicine, Novartis Institutes for BioMedical Research, Novartis Pharma AG, CH–4002 Basel, Switzerland. Tel: +41 792740081; E-mail: arne.mueller@novartis.com

Received: 14 May 2021; Revised: 16 July 2021; Accepted: 17 July 2021

Annals of Clinical and Translational Neurology **2021; 8(9): 1845–1856**

doi: 10.1002/acn3.51438

Abstract

Background: Friedreich's ataxia is an inherited, progressive, neurodegenerative disease that typically begins in childhood. Disease severity is commonly assessed with rating scales, such as the modified Friedreich's Ataxia Rating Scale, which are usually administered in the clinic by a neurology specialist. **Objective**: This study evaluated the utility of home-based, self-administered digital endpoints in children with Friedreich's ataxia and unaffected controls and their relationship to standard clinical rating scales. **Methods**: In a cross-sectional study with 25 participants (13 with Friedreich's ataxia and 12 unaffected controls, aged 6–15 years), home-based digital endpoints that reflect activities of daily living were recorded over 1 week. Domains analyzed were hand motor function with a digitized drawing, automated analysis of speech with a recorded oral diadochokinesis test, and gait and balance with wearable sensors. **Results**: Hand-drawing and speech tests were easy to conduct and generated high-quality data. The sensor-based gait and balance tests suffered from technical limitations in this study setup. Several parameters discriminated between groups or correlated strongly with modified Friedreich's Ataxia Rating Scale total score and activities of daily living total score in the Friedreich's ataxia group. Hand-drawing parameters also strongly correlated with standard 9-hole peg test scores. **Interpretation**: Deploying digital endpoints in home settings is feasible in this population, results in meaningful and robust data collection, and may allow for frequent sampling over longer periods of time to track disease progression. Care must be taken when training participants, and investigators should consider the complexity of the tasks and equipment used.

Introduction

Friedreich's ataxia (FA) is a progressive, autosomal recessive, systemic neurodegenerative disorder. FA is most often caused by a pathogenic guanine-adenine-adenine (GAA) trinucleotide repeat expansion in intron 1 of the frataxin (*FXN*) gene on chromosome 9q21, resulting in reduced transcription of the *FXN* gene and a decrease in FXN protein.[1] Age of onset, disease severity, survival, and other clinical aspects correlate with the length of the shorter GAA expansion allele in homozygotes, although the relationship between GAA-repeat length and the clinical phenotype is not sufficiently robust to be a clinically useful indicator of disease.[2,3]

Friedreich's ataxia typically presents in late childhood with progressive gait instability, dysmetria, and dysarthria, leading to loss of independent gait and severe disabilities that are usually followed by onset of non-neurologic

features such as cardiomyopathy and glucose intolerance.[4] Patients with disease onset before 15 years of age lose ambulation and become wheelchair dependent, on average, 11 years after symptoms begin. Loss of ambulation is preceded by changes in stance and balance parameters.[5]

Disease status and progression in FA can be clinically assessed with neurologic composite scales that cover different functional domains such as ambulation, balance, speech, and upper limb and hand function.[6,7] Currently, two rating scales are generally accepted for FA: the Scale for the Assessment and Rating of Ataxia (SARA)[8] and the Friedreich's Ataxia Rating Scale (FARS).[9] The FARS is a three-part test specifically developed for FA and recently modified (mFARS) with a lower maximum severity score of 93 based on psychometric properties[10] and an intraclass correlation coefficient (ICC) of 0.95, which indicates excellent test-retest reliability.[11] Both SARA and mFARS have been validated in two large natural history studies[12,13] and used in clinical trials.

However, clinic-based assessments have limitations. First, because few clinicians have FA-specific expertise, patients and families often need to overcome challenges with mobility, fatigue, and other medical complications from FA and travel long distances to specialist clinics. Second, in-clinic assessments are typically conducted at infrequent intervals (sometimes annually). This can make using these measurements as endpoints challenging in a clinical trial setting that requires rapid insights.

Alternatively, home-based measures, such as a video-based remote assessment of the SARA[14,15] or feedback derived from wearable sensors, can provide high-frequency data, quantify day-to-day variability, and improve the clinician's ability to understand and predict disease progression.[16] Greater numbers in readings of these assessments can increase confidence in pathologic trends and disease progression. Furthermore, quantitative, continuous digital endpoints can objectively measure functions, such as speech or gait parameters, without subjective reader bias.[17] There is also increasing demand from regulators to develop endpoints that are meaningful to patients.[18] Collecting digital endpoints frequently over greater periods of time may allow for objective quantification of activities of daily living (ADL) that are directly meaningful.[19] Thus, home-based measures may facilitate clinical trial conduct, particularly when patients cannot easily attend clinic visits because of physical impairments and travel restrictions.

We report results from a cross-sectional observational study of different home-based, self-administered digital endpoints and in-clinic mFARS assessments in children with FA and unaffected (control) children. The objectives of this study were to assess the usability of these digital devices, to determine how well these endpoints discriminate between groups, and to compare them with the mFARS and complementary assessments.

Materials and Methods

Study design

We conducted a single-center cross-sectional study with 13 children with FA and 12 unaffected age-matched controls (please see Supplementary Material for inclusion and exclusion criteria). The study protocol was reviewed and approved by the Institutional Review Board at the Children's Hospital of Philadelphia (Supplemental Ethics Statement).

During screening, standard FA assessments were conducted in the clinic for all participants with the mFARS[9,20] and with complementary clinical measures for ADL (FA-ADL with 9 categories), the 9-hole peg test (9HPT) for hand motor function, the timed 25-foot walk for ambulation, and the "PATA" bedside speech test (a physician-administrated oral diadochokinesis test, counting syllable utterances in 10 s).[20,21] We report reciprocal values for the 9HPT (1 boards/second) and 25-foot walk test (meters/second).[20]

In the screening clinic visit, families received equipment and training with the home-based digital devices and recorded their first session. At home, participants recorded digital endpoints for three assessments at specified days (but no specific time of the day) within 1 week and responded to a usability questionnaire for the tests and devices before returning the equipment. A visit and assessment schedule is presented in Figure S1.

Digital endpoints and data processing

Hand-drawing assessment

Hand motor function was analyzed using Archimedean spirals drawn with a digital pen (Anoto Pen, DPC Solutions, Morges, Switzerland), similar to the manual hand-drawing assessment in SARA. Spirals were drawn from the inside to the outside with a minimum of three revolutions without lifting the pen from the paper. Participants drew three spirals with both their dominant and non-dominant hand during each session, resulting in a total of 555 spirals (spirals with <3 revolutions were excluded).

From the recorded timestamps, tip pressures, and spatial coordinates, 14 parameters were calculated (Table S1) and roughly grouped into the following categories: velocity, deviation of the actual from the "ideal" spiral, and frequency domain-based parameters developed and implemented by the authors.

Because each subject drew a maximum of three spirals per hand in each of four sessions, we derived summary

statistics over the maximum of 12 repetitions per hand and per parameter for each participant (participants with <6 spirals per hand were not analyzed).

Speech assessment

Speech tests were recorded remotely on a computer or smartphone once in the clinic and three times within 1 week at home (2 attempts per test). Results were analyzed for 10 parameters (Table S2) with the NeuroVocalix system (Cambridge Cognition, Cambridge, UK; referred to as the "PA-TA-KA" app) via a standard web browser. During each attempt, the participant repeated the syllables /pa/, /ta/, and /ka/, as well as the combined syllables /pataka/, as fast as possible for 10 s.[22,23] Summary statistics were calculated per participant and across the four sessions for syllable rates (average number of syllables per second) and other parameters. The summarized speech parameters were correlated with the clinical parameters, mFARS total scores, and ADL total scores for participants with FA with at least two valid tests per syllable (12 out of 13 participants with FA).

Gait and balance assessment

Participants were equipped with five wearable sensors: one on each foot, one on each wrist, and one on the trunk (e.g., shirt). Each sensor had a triaxial accelerometer and gyroscope sampling at 128 Hz (GaitUp Physilog 5, Lausanne, Switzerland). Participants were asked to wear the sensors during waking hours to record walking, balance, and arm movements during six days. Sensors were recharged every night (additional technical methods are in Supplemental Technical Methods Details).

Sensor acceleration for activity analysis was summarized as the vector magnitude over all three axes per minute and was reported as total acceleration in gravity recorded per minute. Fifteen gait parameters (Table S3) per gait cycle were derived from the foot-worn sensors with proprietary software (GaitUp).

Participant feedback

Each study participant or parent filled in a feedback form (Supplemental Patient Acceptability and Satisfaction Survey) that graded different aspects of each test and device (i.e., ease of use, engagement, satisfaction, and future use) on a scale of 1 (bad) to 7 (good). Free-text feedback was collected with open-ended questions and summarized.

Statistical analysis

We used the Pearson method for correlation and a two-sided Wilcoxon test or linear mixed-effects models for between-group comparisons and assumed significance at $P \leq 0.05$. Summary statistics for the digital parameters were correlated with the in-clinic mFARS total score, ADL total score, and 9HPT time (each recorded once at screening). Additional methods for calculating P values and correlations are available in Supplemental Statistical Method Details.

Results

Participants

Participants in the FA and control groups were of similar age (median [interquartile range (IQR)], 13 [2], and 12 [4] years, respectively; Table 1). Genetic indicators (median shorter GAA repeat length = 766) and the relatively young age at diagnosis (median, 8 years) demonstrated that this group had early onset, genetically severe disease. For patients with early onset FA, the typical time from onset to diagnosis is 2 to 3 years.[4]

All the participants with FA were ambulatory (functional disease staging [FDS] <5). Eight (62%) participants had early stage disease with minimal disability (FDS 1 and 2) and 5 (38%) had stage 3 and 4 disease with mild to moderate severity. Patients with stage 4 FA often use a walking device, and loss of ambulation usually follows within 2 to 3 years.[5]

Adherence and participant feedback

Hand-drawing tests

Five hundred and eighty-two of the expected 600 spirals (25 participants, 4 sessions, 6 spirals per session) were received. One of the spirals was excluded because it was not a spiral (and could not be analyzed), and six were

Table 1. Participant demographics.

	Control group	FA group
Participants, n	12	13
Sex, male (%)	7 (58)	3 (23)
Age at test	12 [4]	13 [2]
Years since diagnosis	–	4 [3]
Age at diagnosis	–	8 [2]
Repeat length (shorter)[a]	–	766 [163]
Functional disease stage	0 [0]	2 [2]
Total mFARS score	0 [1]	41 [10]

Data are shown as n (%) or median [IQR].
FA, Friedreich's ataxia; IQR, interquartile range; mFARS, modified Friedreich's Ataxia Rating Scale.
[a]Guanine-adenine-adenine in intron 1 repeat expansion if applicable. There were no identified point mutations in the FA group.

excluded because the spirals were drawn on the same paper with two different pens (mixing participants, possibly siblings). Of the remaining 575 spirals, 555 had at least three revolutions (25 participants). The 20 spirals with fewer than three revolutions were largely from a participant with severe FA (aged 16 years) with a relatively high mFARS total score of 42. The mean intraparticipant standard deviations of the time the test were conducted (during the day) were 3 h (without FA) and 3.5 h (with FA).

Study participants found the pen easy to use because it was "just like a regular pen" (quote from a participant). However, some participants requested better instructions and feedback from the system because there was no contemporaneous indicator of whether the test was technically successful.

Overall, the drawing exercise received the most positive feedback in all applicable categories (Table S4).

Speech tests

Of the 25 participants, 12 recorded fewer than the expected number of tests (16 total, 4 for each of the 4 syllables) but only 2 participants recorded less than 50% of the expected number of tests (these were excluded from correlation analysis). The mean intraparticipant standard deviations of the time the test were conducted (during the day) were 2 h (without FA) and 4 h (with FA).

Most participants found the "PA-TA-KA" app test easy and quick to complete and the instructions clear, but some participants found the test duration to be too long. The test required two attempts and encouraged participants to try harder the second time, which some participants found frustrating. Also, the system required a stable internet connection, which created technical challenges in some instances.

Gait and balance tests

During the 6-day observation period, 11 out of 12 unaffected control participants and 9 of 13 children with FA had steps recorded, but with varying and overall low step counts. For the control group, the average number of steps was 13,521 (SD, 8,182), and for the FA group, the average number of steps was 7,092 (SD 6,394). One of the nine participants in the FA group was excluded from the gait analysis because steps did not compose longer walking periods. The foot sensors were worn for 34 and 26 h of the expected 72 h (i.e., approximate waking hours in 6 days) for the control and FA groups, respectively (Figure S2). The gait and balance system received an average overall satisfaction score of 3.9 out of 7 points. Participant feedback was mixed due to complexity of use.

Analysis of results

Hand-drawing analysis

Transformation of spiral data for analysis from the hand-drawing tests is shown in Figure 1. An ideal Archimedean spiral (Figure 1A), if plotted in polar coordinates, is displayed as a diagonal (Figure 1B, black line) with the theta angle and radius. However, a hand-drawn spiral usually deviates from this ideal spiral with a systematic and a random component, as illustrated in the deviations from the ideal diagonal line (straight line fit). The systematic error can be explained conceptually by the choice of the center of the spiral.[24] In our analysis, this was the first point drawn.

Spirals drawn with the non-dominant hand appeared wider and less regular than those drawn with the dominant hand. These features also appeared to differ in spirals drawn by participants with FA and control participants, the latter whose spirals were more uniform (Figure 1C). We calculated 14 parameters and compared the distribution between controls and participants with FA (Figure S3). Of the parameters found to differ significantly between these groups, mean angular velocity and mean squared error were distinct between the two groups for both dominant and non-dominant hands (Figure 1D). The mean squared error of the residuals of a linear fit to the actual data in polar coordinates (illustrated in Figure 1B) is the overall deviation from the ideal spiral. Mean angular velocity is an estimate of drawing speed for a curvilinear line.

A strong correlation between aggregated hand-drawing parameters and common clinical endpoints was reported for participants with FA (Figure 2). Additionally, the drawing parameters that best discriminated between participants with FA and controls were not those with the strongest correlation with clinical parameters. The 9HPT is the clinical test usually used for quantifying hand function in FA.[21,25] In particular, for the non-dominant hand, the 9HPT correlated strongly negatively with the median intra-spiral SD of linear drawing speed (Figure 2B). Also, digitally quantified drawing speed aspects reflected the complex and composite mFARS and ADL scores (Figure 2C,D). Additional correlations between clinical scores are found in the Supplementary material.

Speech analysis

Syllable rate distribution counts for all syllable types were significantly different between groups when using the "PA-TA-KA" app (Figure 3A). Syllables for the PATA test are counted by the physician during the in-clinic FA

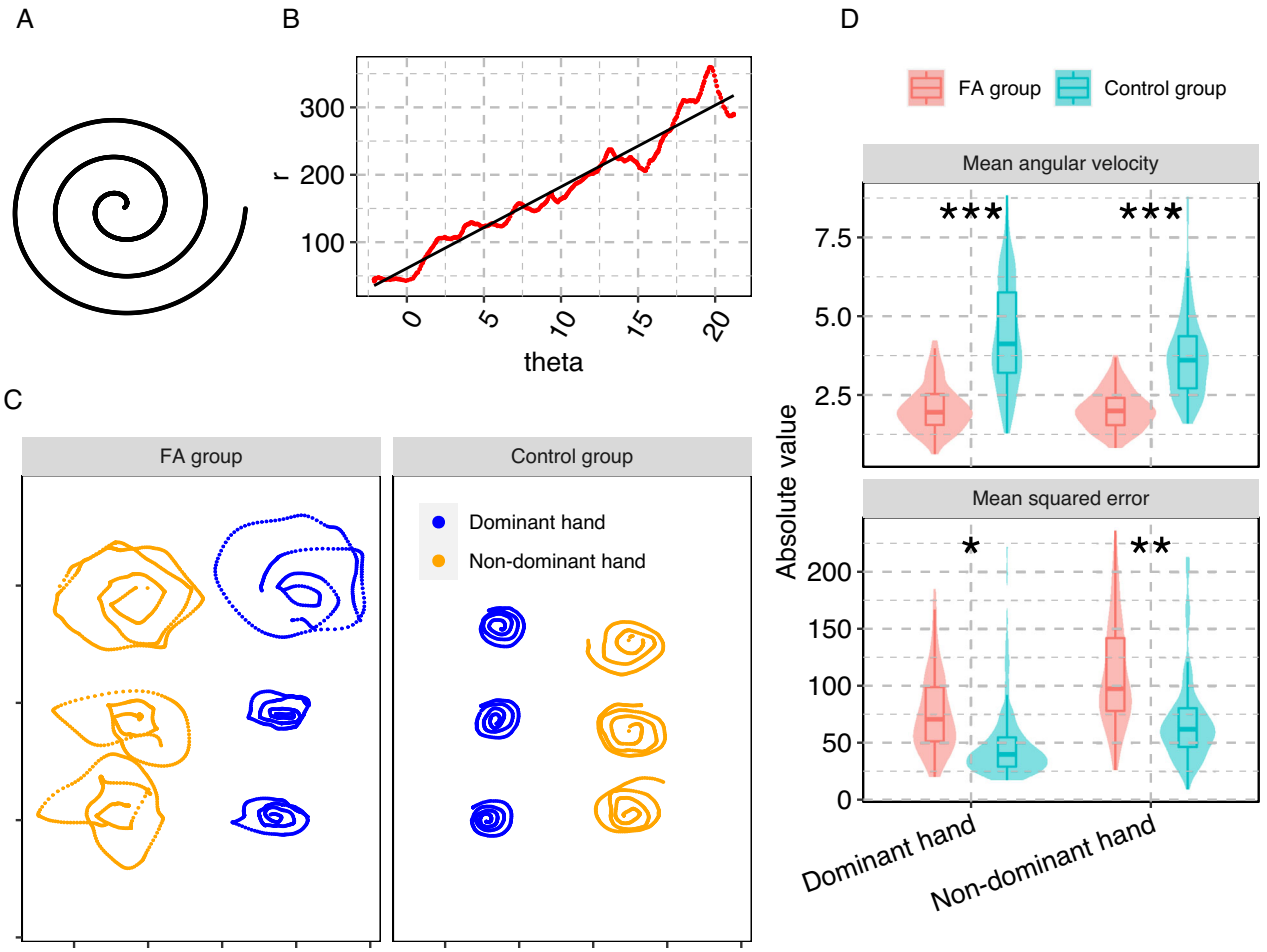

Figure 1. Hand-drawing exercise. (A) An ideal Archimedes spiral in Cartesian coordinates. (B) In polar coordinates, this spiral is represented as theta angle and radius (r), and is a diagonal (straight line), while a hand-drawn spiral oscillates on the diagonal. (C) Six example spirals are shown (left spirals with left hand, right spirals with right hand) from a 10-year-old child with FA (diagnosed 7 years ago, right-handed) and a 10-year-old control child (left-handed). (D) Two of the 14 parameters derived from all spirals of all participants per hand. Mean squared error of the residuals of a linear fit (overall deviation of the black line from the red line as in B), units are points on the paper and angular velocity in degrees/second. *$P \leq 0.001$; **$P \leq 0.05$; ***$P \leq 0.01$. FA, Friedreich's ataxia.

assessment (1 test result per participant) and are shown for reference (/pata*/ data). The PATA test easily discriminates between participants with and without FA but is not automated and therefore not suitable for frequent home-based assessments.

Each test in Figure 3A is color coded by the mFARS bulbar score, which ranges from 0 (no bulbar impairment) to 5 (no useful speech). The participants with FA in this study scored only mild bulbar impairment with a maximum of 1 out of 5 possible points. There was no clear separation of greater bulbar scores (0.5 and 1.0) and low score (0) by syllable rate in the FA group. Nevertheless, the FA group with a bulbar score of 0 had a significantly lower mean syllable count than the participants in the control group ($P = 0.001$ from a mixed-effects analysis of variance), indicating a greater resolution of syllable counting compared with the bulbar score. Note, only a few participants scored above 0 (4 out of 13 in the FA group). Speech parameters per syllable test scores are in Figure S4.

Although the /pa/ syllable rate was the weakest discriminator between participants with FA and controls, other aggregated parameters derived from this syllable correlated well with mFARS and ADL total scores (Figure 3B, C), indicating that the variability of the syllable duration increased between tests for participants with more severe disease. The ADL total score correlated strongly with between- and within-test variability of the syllable

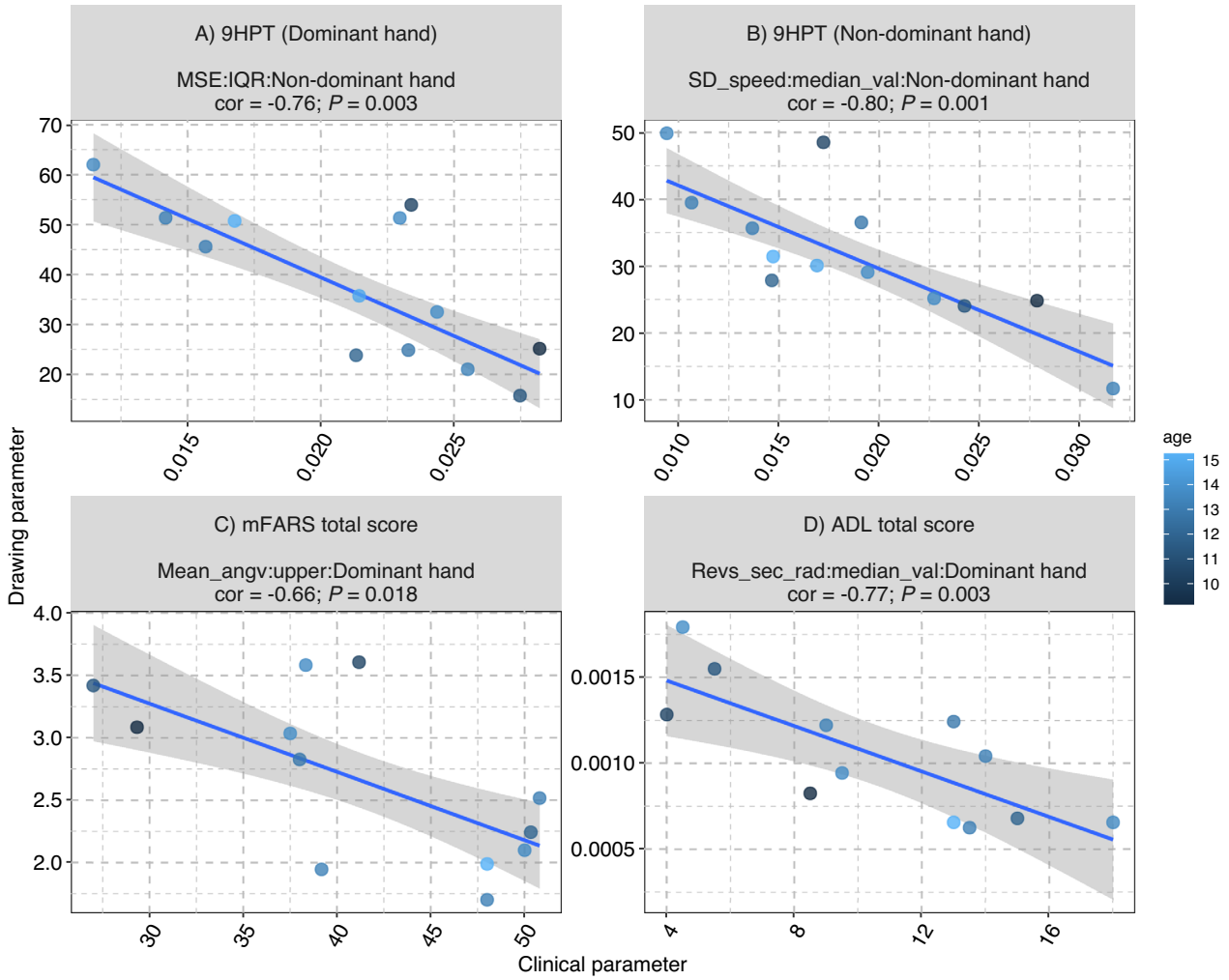

Figure 2. Correlation of hand-drawing parameters with clinical reference scores for participants with FA. The drawing parameter for each panel is a summary statistic S1 within spiral (e.g., mean angular velocity) and across spirals (e.g., upper 90th percentile of S1) giving second-level summary statistics S2 per hand. (A) 9HPT speed versus IQR of MSE for the dominant hand. (B) Median SD of linear drawing speed versus 9HPT for the non-dominant hand. (C) mFARS total score versus 90th percentile of mean angular velocity. (D) ADL total score versus median revolutions per second normalized by revolution radius. Each participant with FA is color coded by age, and the Pearson correlation coefficient and its *P* value are given per panel. 9HPT, 9-hole peg test; ADL, activities of daily living; IQR, interquartile range; Mean_angv, mean angular velocity (change in angle per second); mFARS, modified Friedreich's Ataxia Rating Scale; MSE, mean squared error; Rev_sec_rad, revolutions per second per median radius (in points on paper); SD_speed, SD of linear drawing speed in mm/second.

duration. These results were broadly in agreement with previous speech analyses comparing adults with FA and an unaffected population.[26]

Speech and hand-drawing both require fine motor skills for timely synchronized muscle control driven by the cerebellum, and we observed that variation in syllable duration correlated negatively with drawing velocity ($r = -0.7$, $P < 0.016$; Figure S5). Significant correlation between speech parameters and GAA repeat length of the short allele also was noted, in particular the instability of the gaps between /ka/ syllables ($r = 0.71$, $P < 0.01$; Figure S6).

Analysis of gait and physical activity

Greater activity levels were measured for the foot sensors, followed by wrist and trunk sensors. The control group was significantly more active than the FA group with foot and wrist movements (both $P < 0.01$) but not with trunk movements (Figure S7).

The interpretation of activity levels from wrist and trunk sensors was challenging because of unknown context (i.e., the movements cannot be classified into ADL categories). The main activity from foot-worn sensors was likely walking, which can be further characterized by gait

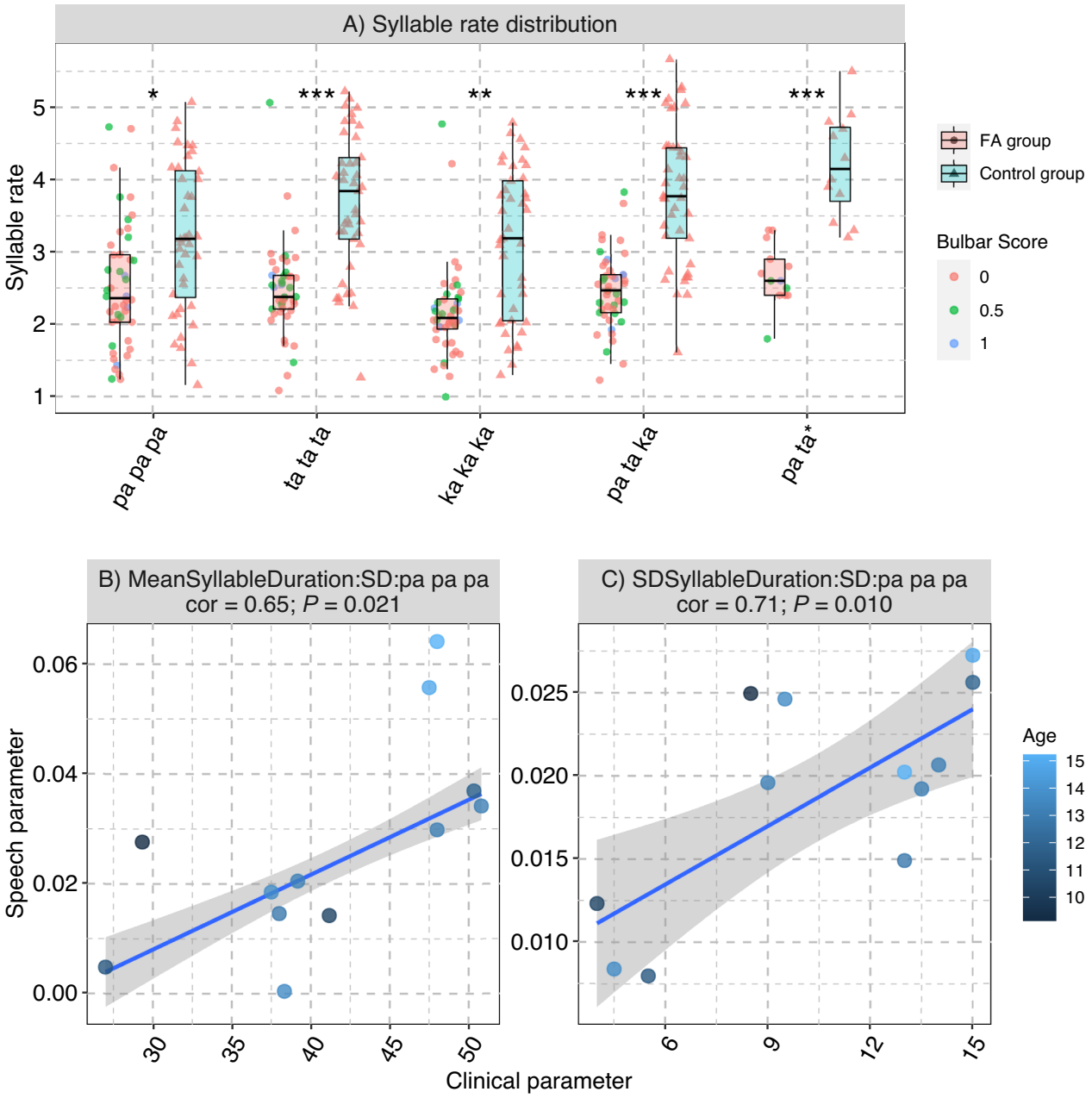

Figure 3. Speech parameters. (A) Syllables per second (rate) per group, color coded by mFARS bulbar score. Each distribution is from all four sessions for all participants. The /pata/* is for the PATA test as part of the clinical FA assessment. (B) Correlation of SD of the mean syllable duration per subject for the /pa/ syllable (within subject syllable duration variability) and the total mFARS score. (C) Correlation of the SD of the syllable duration within each test and between tests within subject with ADL total score. (B) and (C) aggregate speech parameters per participant versus clinical parameters. *$P \leq 0.05$; **$P \leq 0.001$; ***$P \leq 0.01$. ADL, activities of daily living; FA, Friedreich's ataxia; mFARS, modified Friedreich's Ataxia Rating Scale.

parameters (e.g., walking speed). Peak swing and stance period were the most discriminatory parameters between groups (Figure 4A). The FA group spent more time in stance phase and had high peak acceleration during swing time (all parameters are in Figure S8). Similar results have been reported from gait-lab experiments for patients with cerebellar ataxia.[27–29]

Figure 4B presents the clinical test scores and gait parameters with the strongest correlation in the FA group. There was correlation between the 25-foot walk

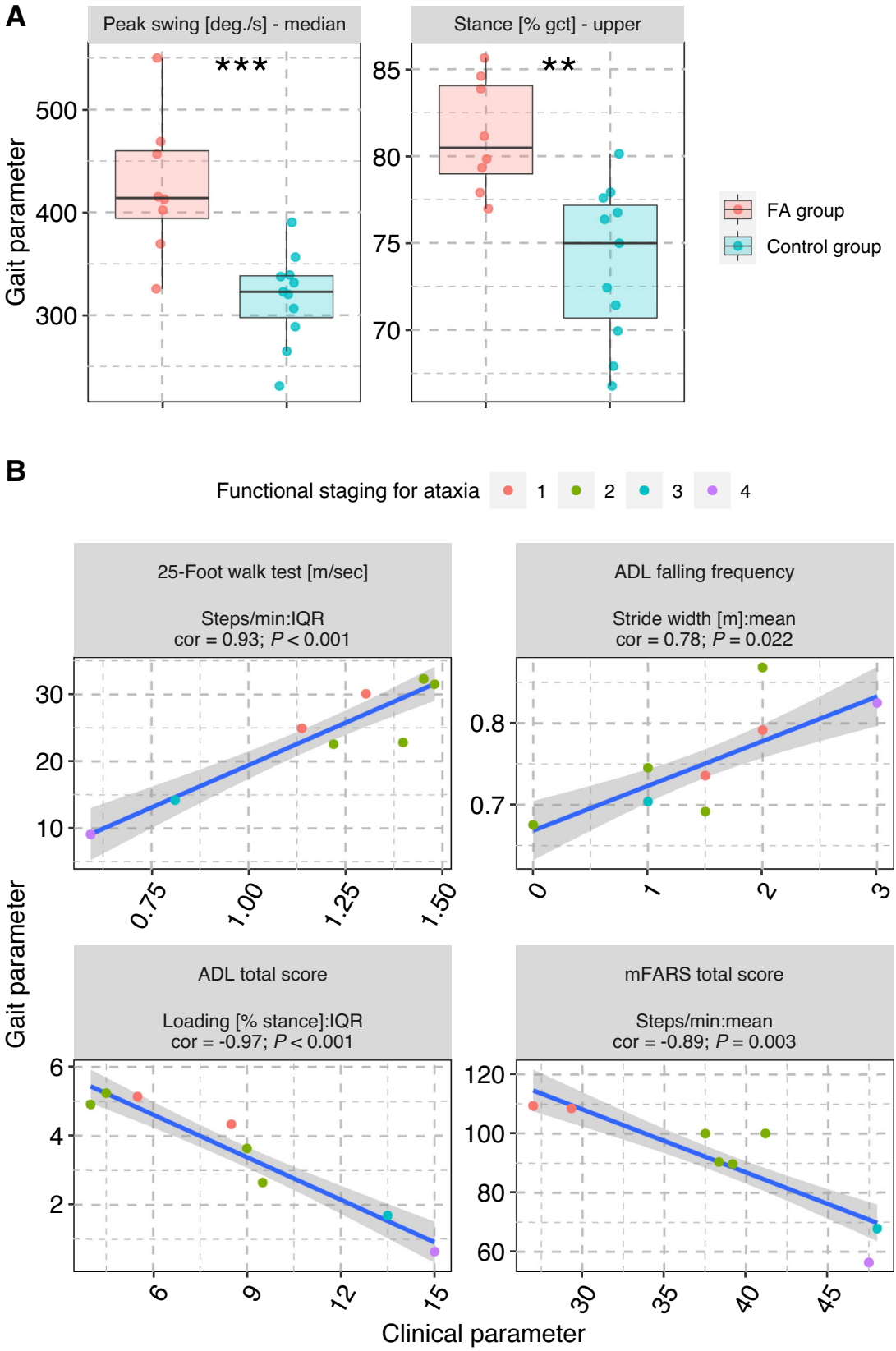

Figure 4. Gait parameters. (A) Distribution of the most discriminating parameters: peak swing (fastest angular velocity during swing period) and stance period (time between heel strike to toe-off as percent of gct) for the two groups. Each participant is shown as a dot and was aggregated over all gait cycles (median, mean, upper = 90th percentile, IQR). (B) For the clinical parameters (reciprocal) 25-foot walk test, ADL falling history, ADL total score, and mFARS total score, the strongest gait correlates are shown: steps per minute is a measure of cadence, mean stride width (i.e., the distance between heel strikes of the same foot), and the variability of loading (i.e., the time between heel strike and foot fully flat on the ground). Parameters were aggregated per participant as in (A). The color code by functional disease stage was chosen because it mainly describes limitations in mobility. $*P \leq 0.001$; $**P \leq 0.01$. ADL, activities of daily living; FA, Friedreich's ataxia; gct, gait cycle time; IQR, interquartile range; mFARS, modified Friedreich's Ataxia Rating Scale.

test and the cadence as estimated from real-world walking. These correlations remained strong and significant even when accounting for FDS as a covariate, except for the correlation between mean steps per minute and mFARS ($r = 0.05$, $P = 0.9$), indicating that the mFARS total score was driven by FDS.

The risk of falling is an important safety consideration, and changes in gait parameters have been associated with fall risk for patients with cerebellar ataxia.[30] Digitally derived stride width strongly reflected the risk of falling ($n = 8$, $r = 0.78$, $P = 0.022$), whereas the clinical 25-foot walk test from the mFARS had no significant correlation with ADL falling history ($n = 8$, $r = 0.59$, $P > 0.1$) in the same participants with FA.

An association between FA genetics and gait was also discovered. Mean stride width from the real-world gait analysis dropped (Pearson $r = 0.79$, $P = 0.021$) with GAA repeat length of the short allele (Figure S9) in participants with FA, whereas there was no significant correlation between 25-foot walk time and allele length in the same participants. For the young patients in our study with early disease onset, stride and other gait parameters related differently to GAA repeat length than previously reported in an older, genetically more diverse FA population.[31]

Discussion

This cross-sectional study of children with and without FA evaluated the usability and discriminative power of home-based digital endpoints for hand motor function, automated analysis of speech, and gait and balance between groups and compared them with traditional clinical scales for FA.

The digital endpoints for hand motor function and automated analysis of speech received good feedback and delivered high-quality data with strong discriminative power between the groups. They also showed correlation with clinical parameters (such as mFARS and ADL total scores) in the FA group. The assessments were quick to conduct and with a low burden for participants. These findings suggest that home-based endpoints could be reliably deployed in future interventional studies to measure longitudinal disease progression and the impact of a therapeutic intervention. Conversely, the complexity of the wearable sensor network led to scant analyzable data. A single button per sensor started and ended recordings as well as marking events (gait and balance tests), depending on how long it was pressed. Difficulties with this mode of operation are a likely cause for missing or non-synchronized recordings. This highlights the need for simple and robust systems, planning, validation, and training when deploying digital wearable devices (e.g., for gait analysis over longer periods of time, foot-worn sensors must be mounted comfortably but securely to avoid loss of data). The data for the hand-drawing and gait assessments were downloaded only at the end of study. For longer studies, we recommend uploading data in a timely manner (e.g., via a connected app) to detect adherence or quality issues in close to real time.

The digital assessments in this study encompassed many derived parameters, and some of these parameters may not be easy to interpret clinically. Simple parameters such as velocity were easier to interpret and are more likely to be of use in clinical practice than frequency domain parameters. However, more complex parameters could provide added value to existing clinical test scores. Parameters discriminating best between the groups do not necessarily track disease severity, and we observed this for all digital endpoints in this study. The difference in functional performance between the groups was large compared with differences between disease states in the FA group, for which parameters must be able to detect a relatively small change. The dynamic range from healthy to severe FA may not be linear for any single parameter.

Test–retest reliability varied strongly between the different digital parameters and between groups with a maximum ICC of 0.77 for hand-drawing (Figures S10–S12) compared with 0.95 for mFARS.[11] However, the digital assessments can be self-administered at home over long periods of time, potentially generating many more data points over time than in-clinic mFARS assessments. Also, in this study, some parameters with low ICC demonstrated a greater correlation with disease status based on clinical scores. This indicates that disease severity in FA is reflected by the intra-participant variability itself for some of the digital parameters.

Patient safety must be considered in home-based assessments and with digital, wearable equipment. Some tests such as stance, balance, or walk tests can pose a risk of falling and should only be performed with a caregiver. Continuous data collection with wearable devices in daily living that do not require specific tests may reduce this risk. Furthermore, care must be taken to deploy systems that are compliant with data privacy regulations, which may exclude certain commercial products.

There are limitations in this study. The lack of longitudinal data did not allow for analysis of disease progression for any given participant or for the FA group in aggregate, longer-term learning effects, or day-to-day variability. Also, the participant feedback and adherence analysis for 1 week may not be representative of longer studies.

Genetic severity and age of the study cohort were relatively homogeneous, although we observed substantial between-participant variability in most digital readouts, and the small number of participants limited the statistical inference. Some of the reported correlations may turn out to be spurious when studying a more diverse population. Dependencies or confounding between the digital and clinical parameters, such as age or years since diagnosis, may be observed. We applied partial correlation to evaluate the effect of these co-factors and most appeared to be relatively independent. Specifically, speech parameters can be age dependent for children. Interestingly, we observed that /pataka/ syllables were spoken more regularly with increasing age in the control children, whereas the group with FA revealed speech irregularity in all ages (Figure S13).

The gait analysis was based on proprietary algorithms (GaitUp) that were not designed specifically for an ataxic population. Thus, its performance and possible bias in the pediatric group with FA compared with the general population is unknown. Furthermore, participants must be ambulatory for gait analysis, which may not be a given for later stage FA.

If tests are conducted frequently, learning effects might impact test performance. Learning effects could be addressed with statistical methods or longer intervals between tests, but this short study did not allow for thorough analysis of learning effects. However, only weak trends toward better performance were observed over the four sessions within 1 week for some parameters.

Robust home-based continuous or task-based digital assessments could be particularly useful when travel is reduced or not feasible, such as during the COVID-19 pandemic, when patients cannot or prefer not to attend clinic visits. We conducted this short cross-sectional study in summer 2019 in the United States and planned to expand this into a longitudinal study by including an additional clinic visit for standard FA assessment at 1 year to compare disease progression with the different endpoints. However, this second planned observation period was cancelled because of the COVID-19 pandemic.

The digital endpoints presented here may complement the established clinical scales because they are relatively easy to conduct at high frequency without a specialist and they provide a breadth of objective data. Composite clinical outcomes (i.e., mFARS and ADL) are designed to describe overall disease severity but may not offer the granularity to correlate sufficiently with measures of individual subdomains.

The work presented here encourages refinement and deployment of these or similar digital endpoints in interventional studies, not only in FA, but also in other neurologic and neuromuscular disorders.

Acknowledgments

The authors acknowledge all participants in this cohort study and express their gratitude for their commitment to research, as well as to the clinical research teams for the organization and data collection. The authors thank Dr. Jang-Ho Cha from Novartis Neuroscience Translational Medicine for supporting this study.

Author Contributions

C.J.M. and M.S. contributed to the conception of the research project. E.P., A.M., M.B., and C.J.M. contributed to the organization of the research project. E.P., M.B., D.R.L., M.W., and C.J.M. contributed to the execution of the project. A.M., H.H., C.R., C.J.M., A. Mc., and J.P. designed the statistical analysis. A.M., A. Mc., and J.P. executed the statistical analysis. The first draft was created by A.M., A. Mc., C.R., and M.L.K. All authors reviewed and critiqued the manuscript.

Funding Information

This work was funded by Novartis Pharma AG.

Conflicts of Interest

A.M., E.P., A. Mc., J.P., H.H., M.L.K., M.S., and C.J. M. are or were employees of Novartis and hold stock in Novartis. D.R.L. has received research funding for the current work. C.R. has received research funding and consultancy fees from the Friedreich's Ataxia Research Alliance, Voyager Therapeutics, Novartis, Larimar Therapeutics, and PTC Therapeutics. M.B. and M.W. have nothing to disclose. Editorial support, including copyediting and formatting of the manuscript and supplement,

was provided by Laura Sitler, ELS, Kay Square Scientific, Newtown Square, PA, USA. This support was funded by Novartis Gene Therapies, Inc., Bannockburn, IL, USA.

REFERENCES

1. Campuzano V, Montermini L, Molto MD, et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 1996;271 (5254):1423–1427.
2. Durr A, Cossee M, Agid Y, et al. Clinical and genetic abnormalities in patients with Friedreich's ataxia. N Engl J Med 1996;335(16):1169–1175.
3. Koeppen AH. Friedreich's ataxia: pathology, pathogenesis, and molecular genetics. J Neurol Sci 2011;303(1–2):1–12.
4. Indelicato E, Nachbauer W, Eigentler A, et al. Onset features and time to diagnosis in Friedreich's ataxia. Orphanet J Rare Dis 2020;15(1):198.
5. Rummey C, Farmer JM, Lynch DR. Predictors of loss of ambulation in Friedreich's ataxia. EClinicalMedicine 2020;18:100213.
6. Trouillas P, Takayanagi T, Hallett M, et al. International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. The Ataxia Neuropharmacology Committee of the World Federation of Neurology. J Neurol Sci 1997;145(2):205–211.
7. Cano SJ, Hobart JC, Hart PE, et al. International cooperative ataxia rating scale (ICARS): appropriate for studies of Friedreich's ataxia? Mov Disord 2005;20 (12):1585–1591.
8. Schmitz-Hubsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 2006;66(11):1717–1720.
9. Subramony SH, May W, Lynch D, et al. Measuring Friedreich ataxia: interrater reliability of a neurologic rating scale. Neurology 2005;64(7):1261–1262.
10. Rummey C, Corben LA, Delatycki MB, et al. Psychometric properties of the Friedreich ataxia rating scale. Neurol Genet 2019;5(6):371.
11. Rummey C, Zesiewicz TA, Perez-Lloret S, et al. Test-retest reliability of the Friedreich's ataxia rating scale. Ann Clin Transl Neurol 2020;7(9):1708–1712.
12. Patel M, Isaacs CJ, Seyer L, et al. Progression of Friedreich ataxia: quantitative characterization over 5 years. Ann Clin Transl Neurol 2016;3(9):684–694.
13. Reetz K, Dogan I, Costa AS, et al. Biological and clinical characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) cohort: a cross-sectional analysis of baseline data. Lancet Neurol 2015;14(2):174–182.
14. Grobe-Einsler M, Taheri Amin A, Faber J, et al. Development of SARA^{home}, a new video-based tool for the assessment of ataxia at home. Mov Disord 2021; https://doi.org/10.1002/mds.28478. [e-pub ahead of print.].
15. Tai G, Corben LA, Woodcock IR, et al. Determining the validity of conducting the modified Friedreich ataxia rating scale and the scale for the assessment and rating of ataxia through video. Mov Disord Clin Pract 2021; https://doi.org/10.1002/mdc3.13204. [e-pub ahead of print.]
16. Rutkove SB, Narayanaswami P, Berisha V, et al. Improved ALS clinical trials through frequent at-home self-assessment: a proof of concept study. Ann Clin Transl Neurol 2020;7(7):1148–1157.
17. Goldsack JC, Coravos A, Bakker JP, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). NPJ Digit Med 2020;3:55.
18. Mehta N, Wang J, Wang Y, et al. The use of mobile technology in drug development. Clin Pharmacol Ther 2020;108(4):706–709.
19. Haberkamp M, Moseley J, Athanasiou D, et al. European regulators' views on a wearable-derived performance measurement of ambulation for Duchenne muscular dystrophy regulatory trials. Neuromuscul Disord 2019;29 (7):514–516.
20. Lynch DR, Farmer JM, Tsou AY, et al. Measuring Friedreich ataxia: complementary features of examination and performance measures. Neurology 2006;66(11):1711–1716.
21. Lynch DR, Farmer JM, Wilson RL, Balcer LJ. Performance measures in Friedreich ataxia: potential utility as clinical outcome tools. Mov Disord 2005;20(7):777–782.
22. Diepeveen S, van Haaften L, Terband H, et al. A standardized protocol for maximum repetition rate assessment in children. Folia Phoniatr Logop 2019;71(5–6):238–250.
23. Singh A, Epstein E, Myers LM, Farmer JM, Lynch DR. Clinical measures of dysarthria in Friedreich ataxia. Mov Disord 2010;25(1):108–111.
24. Wang H, Yu Q, Kurtis MM, et al. Spiral analysis-improved clinical utility with center detection. J Neurosci Methods 2008;171(2):264–270.
25. Corben LA, Tai G, Wilson C, et al. A comparison of three measures of upper limb function in Friedreich ataxia. J Neurol 2010;257(4):518–523.
26. Vogel AP, Wardrop MI, Folker JE, et al. Voice in Friedreich ataxia. J Voice 2017;31(2):243.e9–243.e19.
27. Buckley E, Mazza C, McNeill A. A systematic review of the gait characteristics associated with cerebellar ataxia. Gait Posture 2018;60:154–163.
28. Velazquez-Perez L, Rodriguez-Labrada R, Gonzalez-Garces Y, et al. Prodromal spinocerebellar ataxia type 2 subjects have quantifiable gait and postural sway deficits. Mov Disord 2021;36(2):471–480.
29. Milne SC, Murphy A, Georgiou-Karistianis N, et al. Psychometric properties of outcome measures evaluating decline in gait in cerebellar ataxia: a systematic review. Gait Posture 2018;61:149–162.

30. Schniepp R, Wuehr M, Schlick C, et al. Increased gait variability is associated with the history of falls in patients with cerebellar ataxia. J Neurol 2014;261(1):213–223.
31. Milne SC, Hocking DR, Georgiou-Karistianis N, et al. Sensitivity of spatiotemporal gait parameters in measuring disease severity in Friedreich ataxia. Cerebellum 2014;13 (6):677–688.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Visit and assessment schedule.
Table S1. Parameters derived for spiral drawings.
Table S2. Speech parameter description.
Table S3. Gait parameters.
Table S4. Participant satisfaction per test and device.
Figure S2. Analysis of gait from physical activity and real-world walking.
Figure S3. Spiral drawing parameter distributions per group and hand for all recorded spirals.
Figure S4. Speech parameters per syllable test.
Figure S5. Comparison of two digital endpoint parameters for speech (X-axis) and drawing (Y-axis) for participants with FA.
Figure S6. Correlation between speech instability variability (IQR) for the /ka/ syllable summary and the GAA repeat expansion.
Figure S7. Sensor activity levels measured per body location (mean per left/right for feet and wrists).
Figure S8. Gait parameters summarized with different statistics per participant from all steps in all walks.
Figure S9. Correlation between different stride width summary statistics and the GAA repeat expansion.
Figure S10. ICC for drawing parameters calculated from per-participant mean values per session.
Figure S11. ICC for speech parameters calculated from per-participant mean values per session.
Figure S12. ICCs for gait parameters calculated from per-participant mean values from continuous walking bouts in 3 days.
Figure S13. Correlation between age of participants and a speech parameter (parameter of strongest correlation).
Supplementary Material

