

## Capturing longitudinal change in cerebellar ataxia: Context-sensitive analysis of real-life walking increases patient relevance and effect size

Jens Seemann MSc<sup>1,2</sup>, Theresa Beyme<sup>1,3</sup>, Natalie John<sup>1,3</sup>, Florian Harmuth MSc<sup>4</sup>, Martin Giese PhD<sup>1,2</sup>,  
Ludger Schöls MD<sup>5,6</sup>, Dagmar Timmann MD<sup>7</sup>, Matthis Synofzik, MD<sup>3,5\*</sup>, Winfried Ilg PhD<sup>1,2\*</sup>

<sup>1</sup>Section Computational Sensomotrics, Hertie Institute for Clinical Brain Research, Tübingen, Germany

<sup>2</sup>Centre for Integrative Neuroscience (CIN), Tübingen, Germany

<sup>3</sup> Division Translational Genomics of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research and Center for Neurology, University of Tübingen, Tübingen, Germany

<sup>4</sup>Institute of Medical Genetics and Applied Genomics, University Hospital Tübingen, University of Tübingen, Germany

<sup>5</sup>German Research Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany

<sup>6</sup>Department of Neurodegeneration, Hertie Institute for Clinical Brain Research and Centre of Neurology, Tübingen, Germany

<sup>7</sup>Department of Neurology and Center for Translational and Behavioral Neurosciences (C-TNBS), Essen University Hospital, University of Duisburg-Essen, Essen, Germany

\*authors contributed equally

Corresponding author: Winfried Ilg  
Section Computational Sensomotrics,  
Hertie Institute for Clinical Brain Research,  
Otfried-Müller-Straße 25, 72076 Tübingen, Germany  
phone: ++49 7071 29 89125  
email: winfried.ilg@uni-tuebingen.de

Number of characters in title: 147

Number of words in the abstract: 351

Number of words: 4207

Number of figures/tables: 5

Number of references: 59

Running title: Longitudinal analysis of ataxic gait characteristics in real life

*keywords:* cerebellar ataxia, wearable sensors, real life walking, biomarker, digital health

Statistical analysis was conducted by Jens Seemann and Winfried Ilg

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## Acknowledgments:

This work was supported by the International Max Planck Research School for Intelligent Systems (IMPRS-IS) (to J.S.) and the Else Kröner-Fresenius-Stiftung Medical Scientist programme ‘ClinbrAIIn’ (to W.I. and M.G.), as well as the Else Kröner-Fresenius Stiftung Clinician Scientist program "PRECISE.net" (to M.S.). In addition, this work was supported by the European Union, project European Rare Disease Research Alliance (ERDERA, # 101156595) (to M.S.).

## Disclosures:

Dr Ilg received consultancy honoraria by Ionis Pharmaceuticals, unrelated to the present work.

Mr Seemann reports no disclosures.

Mrs Beyme reports no disclosures.

Mrs John reports no disclosures.

Mr Harmuth reports no disclosures.

Prof Giese reports no disclosures.

Prof Schöls served as advisor for Alexion, Novartis and Vico. He participates as a principal investigator in clinical studies sponsored by Vigil Neuroscience (VGL101-01.001; VGL101-01.002), Vico Therapeutics (VO659-CT01), PTC Therapeutics (PTC743-NEU-003-FA) and Stealth BioTherapeutics (SPIMD-301), all unrelated to the present work.

Prof Timmann reports no disclosures.

Prof Synofzik has received consultancy honoraria from Ionis, UCB, Prevail, Orphazyme, Biogen, Servier, Reata, GenOrph, AviadoBio, Biohaven, Zevra, Lilly, and Solaxa, all unrelated to the present manuscript.

## Abstract

**OBJECTIVES:** With disease-modifying drugs for degenerative ataxias on the horizon, ecologically valid measures of motor performance that can detect patient-relevant changes in short, trial-like time frames are highly warranted.

In this 2-year longitudinal study, we aimed to unravel and evaluate measures of ataxic gait which are sensitive to longitudinal changes in patients' real life by using wearable sensors.

**METHODS:** We assessed longitudinal gait changes of 26 participants with degenerative cerebellar disease (SARA:9.4±4.1) at baseline, 1-year and 2-year follow-up assessment using 3 body-worn inertial sensors in two conditions: (1) laboratory-based walking (LBW); (2) real-life walking (RLW) during everyday living. In the RLW condition, a context-sensitive analysis was performed by selecting comparable walking bouts according to macroscopic gait characteristics, namely bout length and number of turns within a two-minute time interval. Movement analysis focussed on measures of spatio-temporal variability, in particular stride length variability, lateral step deviation, and a compound measure of spatial variability (*SPCmp*).

**RESULTS:** Gait variability measures showed high test-retest reliability in both walking conditions (ICC > 0.82). Cross-sectional analyses revealed high correlations of gait measures with ataxia severity (SARA, effect size  $\rho \geq 0.75$ ); and in particular with patients' subjective balance confidence (ABC score,  $\rho \geq 0.71$ ), here achieving higher effect sizes for real-life than lab-based gait measures (e.g. *SPCmp*: RLW  $\rho = 0.81$  vs LBW  $\rho = 0.71$ ).

While the clinician-reported outcome SARA showed longitudinal changes only after two years, the gait measure *SPCmp* revealed changes already after one year with high effect size ( $r_{prb} = 0.80$ ). In the subgroup with spinocerebellar ataxia type 1, 2 or 3 (SCA<sub>1/2/3</sub>), the effect size was even higher ( $r_{prb} = 0.86$ ). Based on these effect sizes, sample size estimation for the gait measure *SPCmp* showed a required cohort size of n=42 participants (n=38 for SCA<sub>1/2/3</sub>).

subgroup) for detecting a 50% reduction of natural progression after one year by a hypothetical intervention, compared to n=254 for the SARA.

CONCLUSIONS: Gait variability measures revealed high reliability and sensitivity to longitudinal change in both laboratory-based constrained walking as well as in real-life walking. Due to their ecological validity and larger effect sizes, characteristics of real-life gait recordings are promising motor performance measures as outcomes for future treatment trials.

## Introduction

Gait measures constitute promising candidates for motor performance outcome measures in upcoming therapeutic intervention studies in ataxias<sup>1,2</sup>, since gait disturbances often present as the first signs of degenerative cerebellar disease (DCD)<sup>3-5</sup> and represent one of the most patient-relevant disabling features throughout the disease course<sup>6,7</sup>. It has been shown in laboratory-based assessments by different movement capture technologies that measures of spatio-temporal variability allow to characterize the specificities of ataxic gait<sup>8-14</sup> (for reviews, see<sup>15-18</sup>) with high sensitivity to ataxia severity in cross-sectional and recently also first longitudinal studies<sup>19,20</sup>.

With progress in wearable sensor technology enabling gait recordings in patients' real life, it was hypothesized that those real-life gait measures could be potentially even more sensitive to disease-specific signatures of ataxic gait impairment compared to clinical and lab settings, due to the complexity and challenges of the environments<sup>21-25</sup>, but also due to the larger amount of available walking bouts<sup>26</sup>. In a first cross-sectional study on real-life gait in degenerative cerebellar ataxia, we have shown that ataxic-sensitive gait measures indeed allow not only to capture the gait variability inherent in ataxic gait in real life; but also to demonstrate high sensitivity to small cross-sectional differences in disease severity, with higher effect sizes in real-life walking compared to clinical gait assessment<sup>27</sup>. However, to serve as ecologically valid patient-focused progression and therapy response outcomes, these gait measures have to prove

(i) their sensitivity to individual longitudinal change in a sufficiently short time span realistic for intervention trials<sup>1,28</sup>, as well (ii) their meaningfulness by anchoring with patient-reported, – as required by the FDA<sup>29</sup>.

Noteworthy, to compare patients' real-life gait behavior at two measurement time points, the influence of context and environment on gait measures have to be considered. These contextual and environmental factors have been shown to have a significant impact on macroscopic gait characteristics, such as average speed, length of walking bouts, and number of turning movements<sup>25, 30-32</sup>. These gait characteristics will differ for indoor (e.g. in a small apartment) vs. outdoor walking, and in turn will influence several gait measures<sup>33, 34</sup>, as it has been shown in healthy participants as well as for different patient populations (Parkinson's disease, dementia, multiple sclerosis, cerebral palsy)<sup>34-36</sup>. This holds for general performance measures like mean gait speed and even more for variability measures<sup>35-37</sup> (e.g. stride length variability, stride duration variability). The analysis of shorter walking bouts for indoor walking - compared to longer outdoor walking - inherently delivers increased variability measures for both healthy participants and patients<sup>36, 38</sup>.

Thus, if one compares measures of a patient's gait variability at two time points one year apart, increased variability could potentially be caused not by increased balance disturbances, but rather by mere differences in the context and environmental factors of the recorded gait behavior. Thus, to identify disease progression or treatment-induced changes in longitudinal analyses of real-life gait behaviour, one needs to take into account these contextual and environmental factors, and in particular their influence on macroscopic gait characteristics.

In this study, we performed a longitudinal analysis of baseline, one- and two-year follow-up gait recordings in lab-based gait assessment as well as in patients' individual real life. Matching of longitudinal walking bouts (baseline and follow-up assessments) was performed according to macroscopic characteristics of walking behavior, namely the bout length and number of

turns. We hypothesized that gait measures capturing longitudinal change in patients' real life could be (i) more sensitive to progression in short, trial-like time-frames (e.g. 1 year) compared to lab-based gait assessments and clinical rating scales; and (ii) more patient-relevant in terms of correlation with patient-reported outcomes of balance confidence in important activities of everyday living. This would be key for future treatment trials, as the targeted primary outcome is usually slowing of disease progression in a limited study period, ideally capturable within one year; and by outcomes reflecting patient relevance, as emphasized by the FDA<sup>29</sup>.

## Methods

### Participants and clinical outcome assessments

*Study participants.* 26 participants at an ataxic or preataxic stage of degenerative cerebellar disease (DCD, age:  $48 \pm 9.5$  years) were recruited from the Ataxia Clinics of the University Hospitals Tübingen and Essen. They consisted of 21 participants in the ataxic stage of DCD as defined by a Scale for the Assessment and Rating of Ataxia [SARA] score of  $\geq 3$  (group ATX; SARA:  $9.4 \pm 3.2$  points), and 5 participants with repeat-expansions in SCA2, SCA3, or SCA6 at the preataxic stage of DCD (SARA score  $< 3$ ) (group PRE; SARA:  $1.6 \pm 0.65$  points)<sup>39</sup>. A total of 18/26 DCD participants carried a repeat expansion in SCA1,2, or 3 (SCA<sub>1/2/3</sub> subgroup). All main analyses were additionally performed in this subgroup, as these fast-progressing polyQ SCA types are a common promising target in many upcoming intervention trials<sup>1,2</sup>. Details of patient characteristics are shown in Table 1.

Patients were included based on the following inclusion criteria: 1.) degenerative cerebellar ataxia in the absence of any signs of secondary CNS disease; 2.) age between 18 and 75 years; 3.) able to walk without walking aids. Exclusion criteria were: severe visual or hearing impairment, cognitive impairment (as assessed by the clinician global impression as standard part of the INAS<sup>40</sup>, particularly in relation to limitations in understanding instructions or performing the gait task), or orthopaedic limitations (e.g., severe arthrosis, or previous lower

limb fractures or hip/knee replacements) that functionally affect gait. In addition, we recruited 34 healthy controls (HC, age:44.08±14.78 years). Healthy participants had no history of any neurological or psychiatric disease, no family history of neurodegenerative disease, and did not show any neurological signs upon clinical examination. Participants were analyzed cross-sectionally at baseline and, where available, longitudinally at one-year and two-year follow-ups.

*Clinical outcome assessments and patient-reported outcomes:* The severity of ataxia was rated using the SARA<sup>39</sup>. The 3 items rating gait and posture are grouped by the subscore SARA posture & gait (SARA<sub>p&g</sub>)<sup>14, 41</sup>. SARA assessments were performed by expert ataxia neurologists (MS, LS, DT). To capture the impact of disease on subjective confidence in daily living, DCD participants were asked to self-report their balance confidence in activities important in daily living using the Activity-specific Balance Confidence Scale (ABC)<sup>42</sup>.

## Standard protocol approvals, registration, and patient consent

The experimental procedure was approved by the local ethics committee (598/2011BO1; 303/2008BO2). All participants gave their informed consent prior to participation.

## Gait Conditions

Walking movements were recorded in two different conditions, namely: (i) **Lab-Based Walking** (LBW condition): Participants walked 60m straight on a 30m indoor floor (i.e. including one turn) at their preferred self-selected speed on a pre-specified straight route in an institutional setting, supervised without any distractions. The turn is excluded from the analysis.

(ii) **Real-Life Walking** (RLW condition): unconstrained walking during participants' usual individual everyday living, where participants were free to move how they wanted and were used to in their individual daily life, without supervision by any study personnel (total recording time: 4-6 hours). Participants were instructed to wear the sensors inside and outside their home, and include at least a half-hour walk. Participants were instructed to wear the sensors in

consecutive recording sessions, each with a duration of maximal 2 hours. Participants documented their recorded walking movements in an activity protocol.

## Movement recordings and gait measures

Three Opal inertial sensors (APDM, Inc., Portland, US) were attached on both feet and posterior trunk at the level of L5 with elastic Velcro bands. Inertial sensor data was collected and wirelessly streamed to a laptop for the automatic generation of gait and balance metrics by Mobility Lab software (APDM, Inc., Portland, US). For the real-life condition (RLW), data was logged on board of each OPAL sensor and downloaded after the session. Step events, as well as spatio-temporal gait measures from the IMU sensors, were extracted using APDM's mobility lab software (Version 2)<sup>43</sup>, which has been shown to deliver good-to-excellent accuracy and repeatability<sup>44, 45</sup>.

From the rich source of gait measures, we here adopted a hypothesis-driven approach, focusing on those measures that have been considered promising candidate gait measures in previous work. Recent longitudinal studies from our group in laboratory-based gait analysis identified one-year longitudinal gait change in a SCA3 cohort<sup>19</sup> and an early SCA2 cohort<sup>20</sup>, revealing stride length variability (*StrideLcv*), lateral body sway, lateral step deviation (*LatStepDev*) and a compound measure of *spatial step variability* (*SPCmp*, combining *StrideLcv* and *LatStepDev*)<sup>27</sup> as most sensitive to longitudinal change in ataxia severity.

*SPCmp* was determined in two steps: *step one* determines for each of the two parameters (*StrideLcv*) and (*LatStepDev*) separately the relative value of an individual subject in comparison to the value range of all participants at baseline (resulting in values between [0-1]). In *step 2*, that measure out of these two measures was taken for final analysis where the individual's result showed a larger abnormality (shown by a value nearer to 1), see also Equation 1 and Figure S3 in Supplement E3.



$$SPcmp (DCD_i) = \max \left( \left( \frac{StrideLCV_{DCD_i - \min_{DCD+HC}}}{\max_{DCD+HC} StrideLCV - \min_{DCD+HC} StrideLCV} \right), \left( \frac{LatStepDev_{DCD_i - \min_{DCD+HC}}}{\max_{DCD+HC} LatStepDev - \min_{DCD+HC} LatStepDev} \right) \right) \quad (\text{Eq.1})$$

In the real-life walking analysis, we showed that *LatStepDev* and *SPCmp* were sensitive to the cross-sectional ataxia severity<sup>27</sup>, as well as to short-term treatment-related improvements in *SCA27B*<sup>46</sup>.

In addition, we included gait speed as a general indicator of functional mobility as well as the variability of the lateral angle of the foot during the stance phase, relative to the forward motion of the gait cycle (*ToeOutAngle<sub>SD</sub>*, inspired by<sup>12</sup>). The lateral sway was determined by the coronal range of motion measured by the lumbar sensor (*CorROM<sub>SD</sub>*).

Several variability measures were calculated using the coefficient of variation  $CV = \sigma/\mu$ , normalizing the standard deviation with the mean value. On this basis, stride length CV (*StrideLCV*) was determined. See detailed description of gait measures in Supplements E2 and E3.

## Selected walking bouts and matching of real-life walking behavior

The analysis focused on walking bouts > 15 strides, as gait variability measures in too short walking bouts are often estimated inaccurately<sup>26</sup>. The first two and last two strides of each bout were removed to reduce the effects of gait initiation and goal-directed deceleration on the variability measures<sup>47</sup>. A walking bout is here defined as a sequence of strides that is not interrupted by a detected turn or a complete halt of at least two median gait cycle durations of the individual. Finally, bouts that showed a jump in gait speed of at least 0.5 m/s between two strides were discarded.

## Matching of real-life walking behavior based on macroscopic gait characteristics

Due to the described influence of contextual and environmental factors on macroscopic gait characteristics - and thus in turn on the examined gait measures-, we introduced a matching

procedure for identifying for each subject comparable walking bouts in the longitudinal assessments. This longitudinal matching procedure was performed based on two macroscopic gait descriptors: bout length and number of turns within 60 s before and after the bout (#turns) (see Supplement E1 for details).

To determine test-retest reliability in everyday life, we used the matching procedure described above. Here, the two baseline measurement days with the most bouts were selected and matched. For the lab-based walking task (LBW), we divided the 60m task into two 30m segments (before and after the turn) and calculated the split-half reliability of gait measures (see Statistics). The test-retest reliability of gait measures was calculated using the Intraclass correlation ICC(2, 1)<sup>48</sup>. ICC values <0.5, between 0.5 and 0.75, between 0.75 and 0.9, and >0.90 were considered as poor, moderate, good, and excellent reliability, respectively<sup>48</sup>. Based on the ICC the minimum detectable change (MDC) was calculated, which is critical in determining whether a treatment-related slowing of disease progression can be reliably detected or is lost in the measurement noise<sup>49</sup>.

$$MDC_{90} = 1.65 \times SD_{baseline} \times (\sqrt{2[1-ICC]})$$

With 1.65 is the z-score of 90 % level of confidence.

## Statistics

Between-group differences (DCD vs. HC group) of movement features were determined by the non-parametric Kruskal-Wallis-test. When the Kruskal-Wallis-test yielded a significant effect ( $p < 0.05$ ), post-hoc analysis was performed using a Mann-Whitney U-test. Effects sizes were determined by Cliff's delta<sup>50</sup>. Repeated measurement analyses were performed for longitudinal analyses using the non-parametric Friedman test to determine within-group differences between assessments. When the Friedman test yielded an effect ( $p < 0.1$ ), post-hoc analysis was performed using a Wilcoxon-signed-rank-test for pairwise comparisons. Effect sizes  $r_{prb}$  for the repeated measurements analyses were determined by matched-pairs rank biserial correlation<sup>51</sup> and were given with 95% confidence intervals.

We report three significance levels: (i) uncorrected \*:  $p < 0.05$ , (ii) Bonferroni-corrected for multiple comparisons \*\*:  $p < 0.05/n=6$ : number of analyzed features, (iii) \*\*\* $p < 0.001$ . Spearman's  $\rho$  was used to examine the correlation between movement measures and SARA scores as well as between measures for different walking conditions. Effect sizes  $\rho$  were given with 95% confidence intervals and were classified as  $\rho$ : 0.1 small effect, 0.3 medium effect, 0.5 large effect, 0.7 very large effect<sup>52</sup>. Statistical analysis was performed using MATLAB (Version R2024A). Based on the longitudinal changes of the gait measures and the SARA score, a sample size estimation was performed using G\*power 3.1<sup>53</sup> to determine the required cohort size for detecting a 50% reduction of progression by a hypothetical intervention.

## Data Availability

Data will be made available upon reasonable request. The authors confirm that the data supporting the findings of this study are available within the article and its Supplementary material. Raw data regarding human participants (e.g. clinical data) are not shared freely to protect the privacy of the human participants involved in this study; no consent for open sharing has been obtained.

## Results

### Sensitivity of gait measures to ataxia severity: cross-sectional results

Cross-sectional analyses of the baseline assessments revealed group differences between DCD vs HC for several examined gait variability measures in both walking conditions, constrained lab-based walking (LBW) (e.g. *StrideLCV*:  $p = 0.002^{**}$ ; *LatStepDev*:  $p < 0.001^{***}$ ; *SPCmp*:  $p < 0.001^{***}$ ) as well as real-life walking (RLW) (e.g. *LatStepDev*:  $p < 0.001^{***}$ ; *SPCmp*:  $p < 0.001^{***}$ , for an overview of all results see Table 2, Figure 1A). High sensitivity of gait measures *StrideLCV*, *LatStepDev*, and *SPCmp* to cross-sectional ataxia severity was indicated by significant correlations with the SARA total score and the SARA<sub>p&g</sub> subscore, with large

effect in both conditions  $\rho > 0.75$ , see Table 2. In addition, gait measures revealed high correlations with the patient-reported balance confidence in important activities of daily living, assessed by the ABC score (*StrideLCV*, *LatStepDev*, *SPCmp*:  $p < 0.001^{***}$ , Table 2). Correlations with the ABC score reveal higher effect sizes for the gait measures in real-life walking (RLW) than lab-based walking (LBW) (e.g. *SPCmp*:  $\rho = -0.81$  in RLW vs  $\rho = -0.71$  in LBW).

Most measures of spatial and temporal variability were highly correlated across the conditions LBW and RLW in cerebellar patients ( $p < 0.003$ , Supplement E8).

## Gait measures for matched walking bouts show good-to-excellent test-retest-reliability

To identify suitable macroscopic gait characteristics for the matching of walking bouts, linear regression analysis showed significant contributions of both, *bout length* and *number of turns* in explaining the variability for walking bouts of healthy controls (see Supplement E3). Performing the presented matching procedure on the real-life baseline assessment (see Methods), revealed good to excellent test-retest-reliability (ICC) for several gait measures like *LatStepDev*, *StrideLCV*, and *SPCmp* ( $ICC(2,1) \geq 0.82$  (see Table 2)).

## Sensitivity of gait measures to longitudinal change

We next analysed whether the gait measures allow to detect longitudinal changes in real life at one-year follow-up (duration:  $380 \pm 52$  days) and at a two-year follow-up assessment (duration:  $816 \pm 106$  days). Longitudinal gait data were available from 23 DCD participants for the first (FU1) and 22 DCD for the second follow-up assessment (FU2). In addition, 34 healthy controls were available for FU1 and 17 for FU2.

Longitudinal analyses revealed for the clinical ataxia score SARA significant changes only in the second follow-up assessment (see Table 3). In contrast, several gait measures - in particular *StrideLCV* and the compound measure of spatial variability (*SPCmp*) - revealed significant

changes already after one year in the real-life walking condition RLW (*StrideLCV*: FU1:  $p=0.0047^{**}$ ,  $r_{prb}=0.67$ ; *SPCmp*: FU1:  $p=0.0007^{***}$ ,  $r_{prb}=0.80$ ). The robustness of these results is supported by stable effect sizes in the second follow-up assessment (*SPCmp*: FU2:  $p=0.016^*$ ,  $r_{prb}=0.77$ ) (Figure 2A, Table 3).

Moreover, these longitudinal changes already after one year for the gait measures *StrideLCV* and *SPCmp* can also be observed and validated in the subgroup  $SCA_{1/2/3}$  (FU1:  $n=18$ ; FU2:  $n=18$ ) (*SPCmp*: FU1:  $p=0.0023^{**}$ ,  $r_{prb}=0.86$ ; FU2:  $p=0.0067^{**}$ ,  $r_{prb}=0.76$ ) – in fact with higher effect size, despite smaller sample size of participants (Supplement E6 ). For both the patient group DCD as well as for the subgroup  $SCA_{1/2/3}$ , these annual changes were larger than the minimum detectable change (MDC, see Tables 3 and supplement E6, e.g. *SPCmp*:  $\Delta_{BL \text{ vs. FU1}} = 0.12 > MDC_{90} = 0.02$ ). In the HC group, longitudinal changes were observed in gait speed, but not in the ataxic-specific measures of spatio-temporal variability (see Figure 1).

Based on observed effect sizes in the DCD population, sample size estimation for the compound spatial variability measure (*SPCmp*) showed a required cohort size of  $n=42$  for detecting a 50% reduction of natural progression after one year by a hypothetical intervention (90% power and one-sided 5% type I error) in comparison to  $n=254$  for the SARA score (Figure 2 B). Required cohort sizes were even smaller for the  $SCA_{1/2/3}$  subgroup ( $n=38$ ), given the larger effect size (Figure 2B).

Compared to the matched walking trials, analysis of the non-matched walking trials (bout length  $>15$ ) also showed significant longitudinal differences, yet with smaller effect sizes (e.g. *SPCmp*: FU1:  $p=0.0089^*$ ,  $r_{prb}=0.62$ ; FU2:  $p=0.01^*$ ,  $r_{prb}=0.58$ , see Supplement E5). Moreover, the analysis of the non-matched walking bouts revealed a correlation between longitudinal differences in variability measures like *StrideLCV* and longitudinal differences in mean bout length ( $r=0.59$ ,  $p=0.0034^{**}$ ) (Supplement E1), potentially influencing the observed changes. Such correlations were not observed in the matched analysis (Supplement E1).

## Discussion

This study aimed to test the hypothesis that the longitudinal progression of ataxic-related gait impairments can be reliably and sensitively captured in real-life walking by analysis of walking bouts which are matched according to macroscopic gait characteristics reflecting environmental factors. We showed that real-life gait measures can capture longitudinal change within short trial-like time frames like one year with high effect size, thus outperforming in their sensitivity both clinician-reported outcomes like the SARA as well as lab-based gait measures which show longitudinal change only after two years. These results thus indicate an increased sensitivity as well as ecological validity for real-life gait measures as promising motor performance outcomes in upcoming treatment trials.

### Increased spatio-temporal gait variability as a consistent feature of ataxic gait in laboratory-based assessments

Our findings in the constrained walking condition LBW confirm the results of previous studies from our and other groups with different movement capture technologies<sup>8, 9, 12, 13, 19</sup>. These studies showed that spatio-temporal variability measures like stride length variability, stride time variability, and lateral step deviation in constrained walking conditions serve as reliable and valid measures for cerebellar ataxia and - as demonstrated here for wearable sensors - correlate with gait and posture ataxia severity as well as with patients' subjective balance confidence in important activities of daily living (ABC-score).

### Measures of ataxic gait in real life: cross-sectional sensitivity to ataxia severity

Although real-life gait is inherently far more variable in both healthy controls and cerebellar patients<sup>38</sup>, several of our gait variability measures (e.g. *StrideLCV*, *LatStepDev*, and *SPCmp*), allow to capture the specific gait variability inherent in ataxic gait in real life<sup>27</sup>. The compound

measure *SPCmp* -integrating variability in the anterior-posterior as well as in the mediolateral dimension - hereby seems to benefit from capturing different compensation strategies used individually and in diverse stages of degenerative cerebellar disease, and might allow capturing gait ataxia in particular in more advanced disease stages (see Supplement E3)<sup>27</sup>.

Consistent with and validating our previous study<sup>27</sup> cross-sectional sensitivity of real-life gait measures (e.g. *StrideLcv*, *LatStepDev*, and *SPCmp*) for ataxia severity was shown by the high correlation with clinical severity of ataxia ( $p < 0.002^{**}$ , see Table 2). Even more meaningful for patient relevance and ecological validity, however, is the high correlation between gait measures and patients' subjective balance confidence in important activities of daily living, as assessed by the ABC score (see Table 2). The high effect sizes of our selected gait measures in their correlation with the ABC score – confirmed and numerically even higher for real-life than lab-based gait assessments (*SPCmp*: RLW  $\rho = 0.81$ , LBW  $\rho = 0.71$ ) - emphasize the relevance of ataxia-related gait impairments, as captured by body-worn sensors, to patients' everyday life – which is key to FDA-conform patient-focussed outcome and drug development<sup>29</sup>.

### Matching walking bouts according to bout length and number of turns

The influence of contextual and environmental factors on gait measurements during real-life walking is currently under intense investigation in various movement disorders<sup>25, 32, 35, 37</sup>. So far, matching or selection procedures based on contextual factors or macroscopic gait characteristics have been used exclusively in cross-sectional studies, e.g. to compare a group of patients with healthy controls in activity monitoring and real-life walking, to differentiate patient subgroups<sup>25</sup>, or to compare patients' real-life walking behaviour with clinical gait assessments<sup>37, 54</sup>. Moreover, in most approaches, the selection of walking sessions was based on bout length alone.

In contrast, the focus of our study was to investigate a longitudinal matching approach as a novel strategy for longitudinal change analysis, allowing to identify comparable walking bouts

at baseline and follow-up visits. As turning movements are an important component of real-life walking behavior, and also typically differ between indoor and outdoor walking<sup>30</sup>, we included the number of turning movements as well as the bout length in our macroscopic characterization of walking behavior to identify comparable walking bouts. The number of turning movements indeed contributes to the macroscopic walking characteristics in explaining step variability, as observed by linear regression analysis of the baseline recordings in healthy controls (Supplement E4). For the longitudinal analyses, the matching procedure ensures that observed effects are not predominantly caused by differences in bout length (Supplement E4). Using the proposed matching procedure, we showed good-to-excellent test-retest-reliability within the baseline assessment, including real-life recordings on different days.

Although the non-matched data showed also significant longitudinal changes (Supplement E7), the results from the matched walking bouts revealed larger effect sizes and were more reliable, given that the matching procedure allowed to substantially reduce the influence of longitudinal changes in bout length as a main confounder (Supplement E1).

Thus, in sum, this approach allows comparison of real-life gait assessments without requiring participants to perform similar gait behaviour in all home assessments or capturing home data over a long period of time (where contextual differences can be assumed to average out). At the same time, it highlights the need also for other future longitudinal gait studies to control for changes in macroscopic walking characteristics like bout length and the number of turnings, as potential confounders for longitudinal analyses.

## Gait measures in real-life walking capture longitudinal change with increased effect size

Using these matching procedures, and consistent with the cross-sectional results, measures of spatio-temporal variability in real-life walking (and in particular the compound measure *SPCmp*) show high responsiveness to change at one year. Their effect size in longitudinal



sensitivity to change outperforms the clinician-reported outcome (SARA) and laboratory gait measures.

Importantly, the high test-retest reliability, even in real-life conditions, resulted in an average longitudinal change that was higher than the minimum detectable change (MDC) (see Table 3), which is a critical requirement for reliable detection in clinical trials<sup>55</sup>.

The subgroup  $SCA_{1/2/3}$  showed even higher effect sizes and consequently smaller sample size estimates. This finding is consistent with previous studies showing faster progression in  $SCA_{1,2}$ <sup>56</sup> compared to a cross-genotype DCD population, which also includes slower progressing DCD types, e.g.  $SCA6$ <sup>57</sup> and non-PolyQ SCAs<sup>58</sup>.

The reduction in sample size inferred by these digital-motor performance outcomes could be decisive for the feasibility of a treatment trial: whereas trials with e.g. 254 SCA participants per trial arm (as required for SARA as outcome) are almost impossible, 38 SCA participants (as required for the gait performance measure  $SPCmp$  in  $SCA_{1/2/3}$ )(Figure2B) are well feasible.

## Limitations

Overall, the present study aimed to explore and longitudinally validate digital-motor sensor measures in real-life in ataxia with a cross-genotype cohort, as this approach allows validation digital-motor measures *across* DCDs (whereas validation within each single genotype would take many additional years and effort). This approach was based on the assumption that our digital-motor measures would capture functional impairment generically across DCDs, given that they qualitatively affect the same ataxia-related functions. This assumption is being corroborated by our previous work demonstrating validity of ataxia-specific gait measures (e.g.  $StrideLcv$ ,  $LatStepDev$ , and  $SPCmp$ ) across various DCDs<sup>20, 46, 59</sup>; as well as by the comparable longitudinal results for the overall DCD cohort and  $SCA_{1/2/3}$  subgroup observed in the current study. However, given that different DCD genotypes gradually differ in their specific progression rates, the validity of the sensitivity of digital motor measures to detect longitudinal

changes formally remains to be demonstrated within specific genotypes as well as pre-ataxic populations only. Larger multi-centre studies focussing on real-life behaviour are thus warranted to confirm our results in larger cohorts sufficiently powered for genotype-specific analyses and including a higher number of preataxic participants. In addition, further work could ideally include longer real-life assessments (e.g. two weeks) to quantify day-to-day variability in patients' motor behaviour.

## Conclusion

This study unravels methods and measures that allow to quantify longitudinal changes in real-life ataxic gait with high effect size and high correlations with patient-reported outcomes of daily living, thus yielding promising ecologically valid, patient-focussed outcome measure candidates for future natural history and treatment trials in degenerative cerebellar ataxias. In addition to the higher effect sizes gained from real-life assessments, these measures allow for the objective quantification of patients' real-life motor performance - instead of clinical outcome assessment in partly "artificial" settings and tasks, e.g. by motor tasks as part of clinical scores (=SARA) or under lab conditions (=standard digital-motor assessments), which serve as surrogate parameters at best. Thus, measures of real-life motor performance add ecological validity and thus help to inform upcoming treatment trials in degenerative cerebellar ataxias and FDA-compatible development of patient-focused outcomes and approval of novel treatments<sup>21, 28</sup>.

## Appendix I Authors

Name	Location	Role	Contribution
Jens Seemann, MSc	Hertie Institute for Clinical Brain Research, Tübingen, Germany	Author	Execution of the study, analysis and interpretation of the data, revising the manuscript
Theresa Beyme	Hertie Institute for Clinical Brain Research, Tübingen, Germany	Author	Execution of the study, analysis of the data, revising the manuscript
Natalie John	Hertie Institute for Clinical Brain Research, Tübingen, Germany	Author	Execution of the study, analysis of the data, revising the manuscript
Florian Harmuth, MSc	Institute of Medical Genetics and Applied Genomics, University Hospital Tübingen	Author	Execution of the study, analysis of the data, revising the manuscript
Martin Giese, PhD	Hertie Institute for Clinical Brain Research, Tübingen, Germany	Author	interpretation of the data, revising the manuscript
Ludger Schöls, MD	Hertie Institute for Clinical Brain Research, Tübingen, Germany	Author	interpretation of the data, revising the manuscript
Dagmar Timmann, MD	Essen University Hospital University of Duisburg-Essen, Germany	Author	Organisation of the study, interpretation of the data, revising the manuscript
Matthis Synofzik, MD	Hertie Institute for Clinical Brain Research, Tübingen, Germany	Author	design and conceptualization of the study, interpretation of the data, drafting the manuscript.
Winfried Ilg, PhD	Hertie Institute for Clinical Brain Research, Tübingen, Germany	Author	design and conceptualization of the study, analysis and interpretation of the data, drafting the manuscript.

## References

1. 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.
2. Ashizawa T, Oz G, Paulson HL. Spinocerebellar ataxias: prospects and challenges for therapy development. *Nat Rev Neurol* 2018;14:590-605.
3. Globas C, du Montcel ST, Baliko L, et al. Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. *Mov Disord* 2008;23:2232-2238.
4. Ilg W, Branscheidt M, Butala A, et al. Consensus Paper: Neurophysiological Assessments of Ataxias in Daily Practice. *Cerebellum* 2018;17:628-653.
5. Luo L, Wang J, Lo RY, et al. The Initial Symptom and Motor Progression in Spinocerebellar Ataxias. *Cerebellum* 2017;16:615-622.
6. NAF. The voice of the patient: Friedreich's Ataxia. Summary report resulting from an externally led patient-focused drug development meeting. <https://www.ataxia.org/ataxiapfd/2017>.
7. Rosen A, Hagen S, Trace K, Compton A. The Voice of the Patient: Living with Polglutamine Spinocerebellar Ataxias (SCA) and Dentatorubal-Pallidoluyision Atrophy (DRPLA). <https://www.ataxia.org/ataxiapfd/>: National Ataxia Foundation 2021 January 2021.
8. Ilg W, Golla H, Thier P, Giese MA. Specific influences of cerebellar dysfunctions on gait. *Brain* 2007;130:786-798.
9. Serrao M, Pierelli F, Ranavolo A, et al. Gait pattern in inherited cerebellar ataxias. *Cerebellum* 2012;11:194-211.
10. Schniepp R, Wuehr M, Schlick C, et al. Increased gait variability is associated with the history of falls in patients with cerebellar ataxia. *Journal of neurology* 2014;261:213-223.

11. Zhou H, Nguyen H, Enriquez A, et al. Assessment of gait and balance impairment in people with spinocerebellar ataxia using wearable sensors. *Neurol Sci* 2022;43:2589-2599.
12. Shah VV, Rodriguez-Labrada R, Horak FB, et al. Gait Variability in Spinocerebellar Ataxia Assessed Using Wearable Inertial Sensors. *Mov Disord* 2021;36:2922-2931.
13. Rochester L, Galna B, Lord S, Mhiripiri D, Eglon G, Chinnery PF. Gait impairment precedes clinical symptoms in spinocerebellar ataxia type 6. *Mov Disord* 2014;29:252-255.
14. Ilg W, Fleszar Z, Schatton C, et al. Individual changes in preclinical spinocerebellar ataxia identified via increased motor complexity. *Mov Disord* 2016;31:1891-1900.
15. Buckley E, Buckley C, Storm F, Mazza C, McNeill A. Spatiotemporal and Upper Body Variables to Quantify Gait Characteristics in Cerebellar Ataxia 8th World Congress of Biomechanics 2018.
16. Milne SC, Murphy A, Georgiou-Karistianis N, Yiu EM, Delatycki MB, Corben LA. Psychometric properties of outcome measures evaluating decline in gait in cerebellar ataxia: A systematic review. *Gait & posture* 2018;61:149-162.
17. Ilg W, Timmann D. Gait ataxia--specific cerebellar influences and their rehabilitation. *Mov Disord* 2013;28:1566-1575.
18. Cabaraux P, Agrawal SK, Cai H, et al. Consensus Paper: Ataxic Gait. *Cerebellum* 2023;22:394-430.
19. Ilg W, Muller 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.
20. Seemann J, Daghzen L, Cazier M, et al. Digital gait measures capture 1-year progression in early-stage spinocerebellar ataxia type 2. *medRxiv* 2023:2023.2010.2008.23296692.
21. Manta C, Patrick-Lake B, Goldsack JC. Digital Measures That Matter to Patients: A Framework to Guide the Selection and Development of Digital Measures of Health. *Digit Biomark* 2020;4:69-77.
22. Orendurff MS, Schoen JA, Bernatz GC, Segal AD, Klute GK. How humans walk: bout duration, steps per bout, and rest duration. *J Rehabil Res Dev* 2008;45:1077-1089.
23. Thierfelder A, Seemann J, John N, et al. Real-Life Turning Movements Capture Subtle Longitudinal and Preataxic Changes in Cerebellar Ataxia. *Mov Disord* 2022;37:1047-1058.
24. Buckley C, Alcock L, McArdle R, et al. The Role of Movement Analysis in Diagnosing and Monitoring Neurodegenerative Conditions: Insights from Gait and Postural Control. *Brain Sci* 2019;9.
25. Mc Ardle R, Del Din S, Donaghy P, Galna B, Thomas AJ, Rochester L. The Impact of Environment on Gait Assessment: Considerations from Real-World Gait Analysis in Dementia Subtypes. *Sensors* 2021;21.
26. Mico-Amigo ME, Bonci T, Paraschiv-Ionescu A, et al. Correction: Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. *J Neuroeng Rehabil* 2024;21:71.
27. Ilg W, Seemann J, Giese M, et al. Real-life gait assessment in degenerative cerebellar ataxia: Toward ecologically valid biomarkers. *Neurology* 2020;95:e1199-e1210.
28. Byrom B, Watson C, Doll H, et al. Selection of and Evidentiary Considerations for Wearable Devices and Their Measurements for Use in Regulatory Decision Making: Recommendations from the ePRO Consortium. *Value Health* 2018;21:631-639.
29. FDA. Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making [online]. Accessed 18.04.
30. Glaister BC, Bernatz GC, Klute GK, Orendurff MS. Video task analysis of turning during activities of daily living. *Gait & posture* 2007;25:289-294.
31. Rehman RZU, Guan Y, Shi JQ, et al. Investigating the Impact of Environment and Data Aggregation by Walking Bout Duration on Parkinson's Disease Classification Using Machine Learning. *Front Aging Neurosci* 2022;14:808518.
32. Roth N, Wieland GP, Kuderle A, et al. Do We Walk Differently at Home? A Context-Aware Gait Analysis System in Continuous Real-World Environments. *Annu Int Conf IEEE Eng Med Biol Soc* 2021;2021:1932-1935.
33. Mc Ardle R, Morris R, Hickey A, et al. Gait in Mild Alzheimer's Disease: Feasibility of Multi-Center Measurement in the Clinic and Home with Body-Worn Sensors: A Pilot Study. *J Alzheimers Dis* 2018;63:331-341.
34. Del Din S, Godfrey A, Galna B, Lord S, Rochester L. Free-living gait characteristics in ageing and Parkinson's disease: impact of environment and ambulatory bout length. *J Neuroeng Rehabil* 2016;13:46.
35. Shah VV, McNames J, Harker G, et al. Effect of Bout Length on Gait Measures in People with and without Parkinson's Disease during Daily Life. *Sensors* 2020;20.
36. Storm FA, Nair KPS, Clarke AJ, Van der Meulen JM, Mazza C. Free-living and laboratory gait characteristics in patients with multiple sclerosis. *PLoS one* 2018;13:e0196463.
37. Carcreff L, Gerber CN, Paraschiv-Ionescu A, et al. Comparison of gait characteristics between clinical and daily life settings in children with cerebral palsy. *Scientific reports* 2020;10:2091.
38. Tamburini P, Storm F, Buckley C, Bisi MC, Stagni R, Mazza C. Moving from laboratory to real life conditions: Influence on the assessment of variability and stability of gait. *Gait & posture* 2018;59:248-252.

39. Schmitz-Hubsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* 2006;66:1717-1720.
40. Schmitz-Hubsch T, Coudert M, Bauer P, et al. Spinocerebellar ataxia types 1, 2, 3, and 6: disease severity and nonataxia symptoms. *Neurology* 2008;71:982-989.
41. Lawerman TF, Brandsma R, Verbeek RJ, et al. Construct Validity and Reliability of the SARA Gait and Posture Sub-scale in Early Onset Ataxia. *Front Hum Neurosci* 2017;11:605.
42. Powell LE, Myers AM. The Activities-specific Balance Confidence (ABC) Scale. *The journals of gerontology Series A, Biological sciences and medical sciences* 1995;50A:M28-34.
43. Mancini M, King L, Salarian A, Holmstrom L, McNames J, Horak FB. Mobility Lab to Assess Balance and Gait with Synchronized Body-worn Sensors. *J Bioeng Biomed Sci* 2011;Suppl 1:007.
44. Washabaugh EP, Kalyanaraman T, Adamczyk PG, Claflin ES, Krishnan C. Validity and repeatability of inertial measurement units for measuring gait parameters. *Gait & posture* 2017;55:87-93.
45. Morris R, Stuart S, McBarron G, Fino PC, Mancini M, Curtze C. Validity of Mobility Lab (version 2) for gait assessment in young adults, older adults and Parkinson's disease. *Physiol Meas* 2019;40:095003.
46. Seemann J, Traschutz A, Ilg W, Synofzik M. 4-Aminopyridine improves real-life gait performance in SCA27B on a single-subject level: a prospective n-of-1 treatment experience. *Journal of neurology* 2023;270:5629-5634.
47. Kroneberg D, Elshehabi M, Meyer AC, et al. Less Is More - Estimation of the Number of Strides Required to Assess Gait Variability in Spatially Confined Settings. *Front Aging Neurosci* 2018;10:435.
48. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15:155-163.
49. Beckerman H, Roebroek ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res* 2001;10:571-578.
50. Cliff N. Answering Ordinal Questions with Ordinal Data Using Ordinal Statistics. *Multivariate behavioral research* 1996;31:331-350.
51. Kerby DS. The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation. *Comprehensive Psychology* 2014;3:11.IT.13.11.
52. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2. Edition ed. Hillsdale: Lawrence Erlbaum Associates, 1988.
53. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* 2009;41:1149-1160.
54. Shah VV, McNames J, Mancini M, et al. Quantity and quality of gait and turning in people with multiple sclerosis, Parkinson's disease and matched controls during daily living. *Journal of neurology* 2020;267:1188-1196.
55. Ilg W, Milne S, Schmitz-Hubsch T, et al. Quantitative Gait and Balance Outcomes for Ataxia Trials: Consensus Recommendations by the Ataxia Global Initiative Working Group on Digital-Motor Biomarkers. *Cerebellum* 2023;23:1566-1592.
56. Diallo A, Jacobi H, Tezenas du Montcel S, Klockgether T. Natural history of most common spinocerebellar ataxia: a systematic review and meta-analysis. *Journal of neurology* 2021;268:2749-2756.
57. 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. *The Lancet Neurology* 2015;14:1101-1108.
58. Durr A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *The Lancet Neurology* 2010;9:885-894.
59. Beichert L, Seemann J, Kessler C, et al. Towards patient-relevant, trial-ready digital motor outcomes for SPG7: a cross-sectional prospective multi-center study (PROSPAX). Accepted for *Neurology* 2024:2024.2001.2009.24301064.

## Tables

Table 1 **Patient characteristics at baseline assessment.** Clinical ataxia severity was determined by the SARA<sup>39</sup>. ADCD: autosomal dominant ataxia of still undefined genetic cause; SCA: autosomal-dominant spinocerebellar ataxia of defined genetic type. The following diagnosis denotes the gene underlying the respective ataxia type: ADCK3 (=ARCA 2, Autosomal-recessive cerebellar ataxia type 2); PNPLA6.

Patient	Diagnosis	SARA <sup>BL</sup>
PRE1	SCA6	2.5
PRE2	SCA2	2
PRE3	SCA3	1
PRE4	SCA3	1
PRE5	SCA3	1.5
ATX1	SCA6	3
ATX2	SCA3	8.5
ATX3	ADCD	4.5
ATX4	ADCK3	8.5
ATX5	SCA3	13
ATX6	SCA2	4.5
ATX7	SCA14	10
ATX8	ADCK3	10
ATX9	PNPLA6	9.5
ATX10	SCA1	12
ATX11	SCA2	12.5
ATX12	SCA3	13
ATX13	SCA3	9
ATX14	SCA6	8.5
ATX15	SCA1	17.5
ATX16	SCA1	10.5
ATX17	SCA3	6.5
ATX18	SCA1	9
ATX19	SCA3	7.5
ATX20	SCA3	9.5
ATX21	SCA3	10
Subgroup	PRE #5	age Ø 45±10.1 SARA Ø 1.6±0.65
Subgroup	ATX #21	age Ø 49.8±9.1 SARA Ø 9.38±3.2
Group	DCD #26	age Ø 48.88±9.5 SARA Ø 7.9±4.2

*Table 2 Cross-sectional analyses: Between-group differences of healthy controls (HC) and DCD participants for clinician reported outcomes, patient-reported outcomes, and gait measures in the lab-based (LBW) and real-life (RLW) walking conditions. Stars indicate significant differences between groups (\* $\equiv p < 0.05$ , \*\* $\equiv p < 0.0083$  Bonferroni-corrected, \*\*\* $\equiv p < 0.001$ ).  $\delta$  denotes the effect sizes determined by Cliff's delta. Correlations between gait measures and clinician-reported ataxia severity (SARA, SARA<sub>p&g</sub>) and patient-reported outcomes (ABC) are given for the DCD group. Effect sizes of correlations are given using Spearman's  $\rho$ . CI<sub>p</sub> denotes the 95% confidence intervals for  $\rho$ . *m*: mean; *sd*: standard deviation. Test-retest reliability is analyzed by determining the intraclass correlation coefficient (ICC) (see Methods).*

	Measure	group difference DCD vs. HC		correlations DCD				ICC
		p	$\delta$	SARA		ABC		
				$\rho$ /CI <sub>p</sub>	p	$\rho$ /CI <sub>p</sub>	p	
Clinician-reported outcomes	SARA	<0.001***	0.99	-	-	0.76	0.001**	-
	SARA <sub>p&amp;g</sub>	<0.001***	0.75	-	-	0.73	0.002**	-
Patient-reported outcome	ABC	<0.001***	0.70	-	-	-	-	-
Gait – LBW	Speed	0.67	0.06	0.03 [-0.4;0.42]	0.88	-0.22 [-0.87;0.34]	0.45	0.94
	StrideL <sub>CV</sub>	0.002**	0.47	0.79 [0.59;0.90]	<0.001***	-0.55 [-0.77;0.59]	0.04*	0.73
	LatStepDev	<0.001***	0.64	0.81 [0.54;0.93]	<0.001***	-0.71 [-0.94;-0.03]	0.005**	0.91
	SPCmp	<0.001***	0.60	0.81 [0.53;0.94]	<0.001***	-0.71 [-0.91;0.09]	0.005**	0.83
	CorRoM <sub>SD</sub>	<0.001**	0.54	0.74 [0.29;0.83]	<0.001***	-0.72 [-0.85;0.3]	0.004**	0.75
	ToeOutAng <sub>SD</sub>	<0.001**	0.62	0.64 [0.41;0.87]	<0.001***	-0.70 [-0.95;-0.48]	0.006*	0.49
Gait – RLW	Speed	0.19	0.2	-0.47 [-0.79;-0.2]	0.02*	0.05 [-0.72;0.6]	0.86	0.83
	StrideL <sub>CV</sub>	0.06	0.29	0.75 [0.67;0.94]	<0.001***	-0.35 [-0.91;0.37]	0.23	0.84
	LatStepDev	<0.001***	0.61	0.77 [0.52;0.93]	<0.001***	-0.81 [-0.93;0]	<0.001***	0.82
	SPCmp	<0.001***	0.54	0.78 [0.51;0.93]	<0.001***	-0.81 [-0.92;0.03]	<0.001***	0.85
	CorRoM <sub>SD</sub>	0.13	0.23	0.59 [0.28;0.81]	0.002**	-0.62 [-0.98;-0.17]	0.02*	0.84
	ToeOutAng <sub>SD</sub>	0.008**	0.41	0.59 [0.06;0.78]	0.002**	-0.27 [-0.94;0.33]	0.36	0.79

Table 3 Longitudinal within-subject comparison for the group of degenerative cerebellar disease (DCD). Shown are results of clinician-reported ataxia ratings (SARA score and SARA<sub>p&g</sub> posture&gait subscore<sup>41</sup>) as well as of gait measures in clinical assessment (LBW) and in real-life walking (RLW) for baseline, 1-year, and 2-year follow-up assessments. Friedman test determined within-group longitudinal differences (+,  $p < 0.1$ ). Post-hoc test  $p$ -values determined by Wilcoxon signed-rank test for both follow-up assessments relative to baseline. Stars indicate significant differences between groups (\* $\equiv p < 0.05$ , \*\* $\equiv p < 0.0083$  Bonferroni-corrected, \*\*\* $\equiv p < 0.001$ ). Effect sizes  $r_{prb}$  determined by matched-pairs rank-biserial correlation<sup>51</sup>. CI<sub>r</sub> denotes the 95% confidence intervals for effect sizes  $r_{prb}$ . Shown are analyses for the whole group of degenerative cerebellar disease (DCD), both consisting of preataxic and ataxic participants. MDC<sub>90</sub> denotes the smallest reliable detectable change (90% confidence interval). The MDC column indicates whether this is smaller than the change between baseline and 1-year Follow-up ( $<^{FU1}$ ) or/and between baseline and 2-year Follow-up ( $<^{FU1+2}$ ).

GROUP		Friedman test		Baseline	1-year Follow-Up			2-year Follow-Up			MDC <sub>90</sub>
DCD	Measure	$\chi^2$	p	m ± sd	m ± sd	p	$r_{prb}/CI_r$	m ± sd	p	$r_{prb}/CI_r$	
Clinician-reported outcomes	SARA	1.42	0.49	8.0±4.5	8.1±4.5	0.66	0.12 [-0.34;0.58]	9.1±5.5	0.02*	0.50 [0.07;0.81]	-
	SARA <sub>p&amp;g</sub>	3.04	0.21	3.1±2.1	3.0±2.1	0.79	0.17 [-0.27;0.57]	3.5±2.5	0.03*	0.53 [0.15;0.82]	-
Patient-reported outcome	ABC	1.87	0.39	75.6±20.5	67.7±23.4	0.03*	-0.71 [-1;-0.18]	68.7±31.2	0.12	[-0.98;0.08]	-
Gait - LBW	Speed	12.0	0.001 <sup>+</sup>	1.39±0.14	1.35±0.12	0.04*	-0.47[-0.84;-0.04]	1.26±0.19	0.0005***	-0.84[-0.99;-0.57]	0.019 $<^{FU1+2}$
	StrideLCV	1.9	0.38	0.02±0.007	0.023±0.016	0.30	0.24 [-0.24;0.66]	0.029±0.02	0.005**	0.68 [0.29;0.92]	0.0065 $<^{FU2}$
	LatStepDev	3.4	0.17	0.032±0.011	0.034±0.013	0.44	0.18 [-0.27;0.65]	0.036±0.015	0.04*	0.47 [0.02;0.82]	0.003 $<^{FU2}$
	SPCmp	6.0	0.04 <sup>+</sup>	0.38±0.21	0.41±0.21	0.27	0.26 [-0.21;0.69]	0.47±0.33	0.04*	0.48 [0.03;0.83]	0.04 $<^{FU2}$
	CorRoMSD	6.5	0.03 <sup>+</sup>	0.87±0.34	0.99±0.34	0.03*	0.51 [0.06;0.87]	0.91±0.33	0.7	0.11 [-0.49;0.46]	0.2
	ToeOutAngSD	0.6	0.7	1.81±0.5	1.79±0.61	0.54	-0.14 [-0.61;0.34]	2.00±0.72	0.13	0.36 [-0.13;0.75]	0.54
Gait - RLW	Speed	4.1	0.12	1.23±0.14	1.17±0.13	0.09	-0.39 [-0.78;0.05]	1.15±0.2	0.009**	-0.63 [-0.9;-0.22]	0.048 $<^{FU1+2}$
	StrideLCV	5.5	0.06 <sup>+</sup>	0.034±0.012	0.042±0.016	0.0047**	0.67 [0.27;0.96]	0.048±0.03	0.04*	0.49 [0.02;0.82]	0.0025 $<^{FU1+2}$
	LatStepDev	5.1	0.07 <sup>+</sup>	0.044±0.009	0.047±0.01	0.05	0.44 [0.02;0.83]	0.048±0.008	0.006	0.66 [0.27;0.95]	0.0021 $<^{FU1+2}$
	SPCmp	10.9	0.004 <sup>+</sup>	0.58±0.17	0.67±0.2	0.0007***	0.80 [0.51;0.97]	0.72±0.31	0.001**	0.78 [0.47;0.97]	0.02 $<^{FU1+2}$
	CorRoMSD	7.6	0.02 <sup>+</sup>	1.0±0.21	1.09±0.2	0.006**	0.65 [0.28;0.95]	1.07±0.24	0.18	0.32 [-0.16;0.74]	0.13
	ToeOutAngSD	3.37	0.18	2.35±0.57	2.56±0.59	0.04*	0.47 [0.01;0.8]	2.68±0.9	0.12	0.37 [-0.07;0.75]	0.22 $<^{FU2}$



## Figures

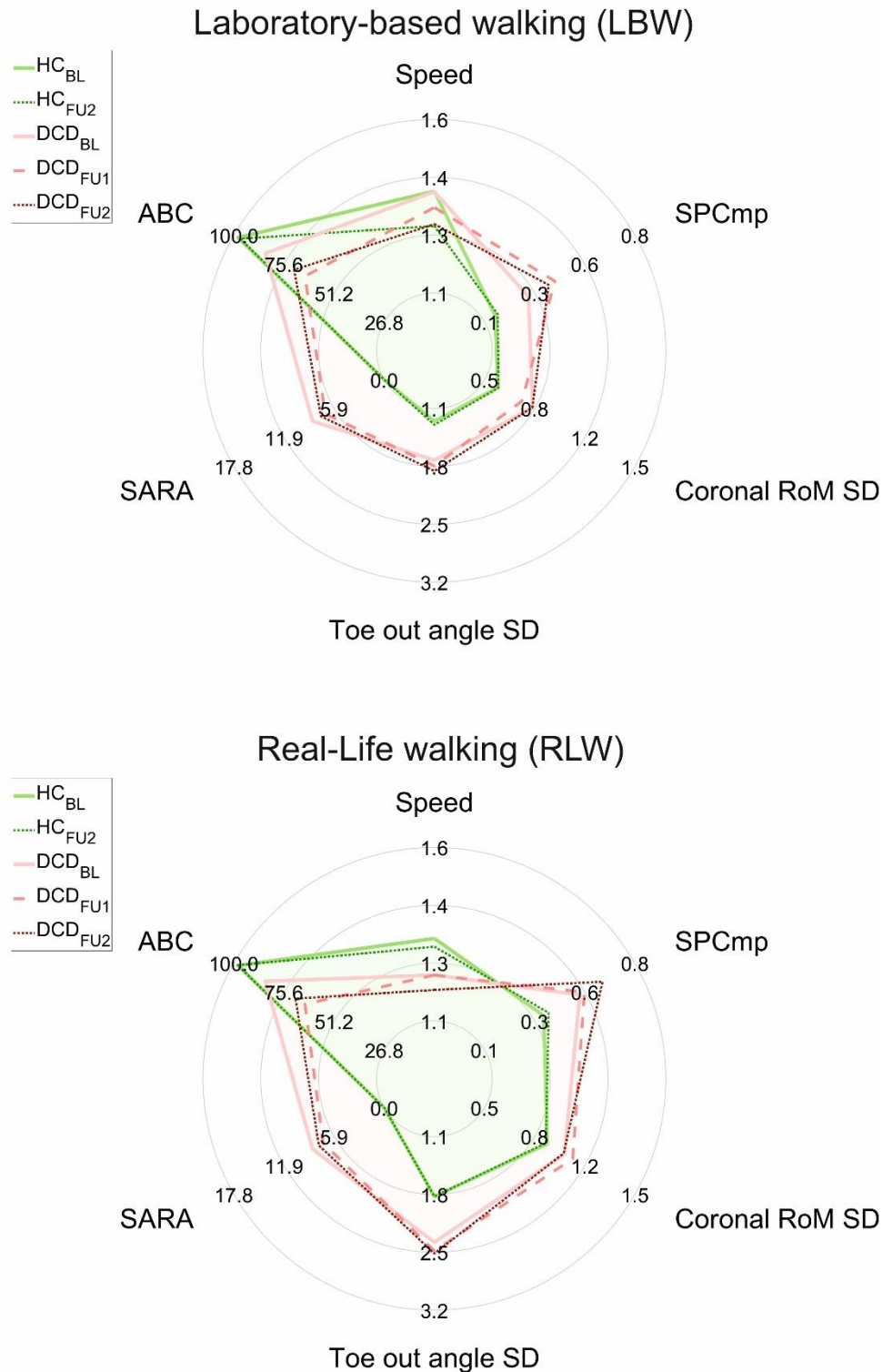


Figure 1 Radar plots illustrating cross-sectional and longitudinal differences on SARA, ABC, and four gait parameters for the gait conditions LBW (A) and RLW (B): Speed, Spatial variability compound measure (*SPCmp*), coronal RoM variability (*CorRoM<sub>SD</sub>*), and foot angle variability (*ToeOutAngle<sub>SD</sub>*). Cross-sectional differences can be seen by comparison of healthy controls (HC<sub>BL</sub>) and the DCD group (DCD<sub>BL</sub>). Given are average values for each group. Longitudinal progression can be seen comparing DCD participants at baseline (DCD<sub>BL</sub>) and 1-follow-up assessments (DCD<sub>FU1</sub>, DCD<sub>FU2</sub>) as well as for healthy controls (HC<sub>BL</sub> and HC<sub>FU2</sub>).

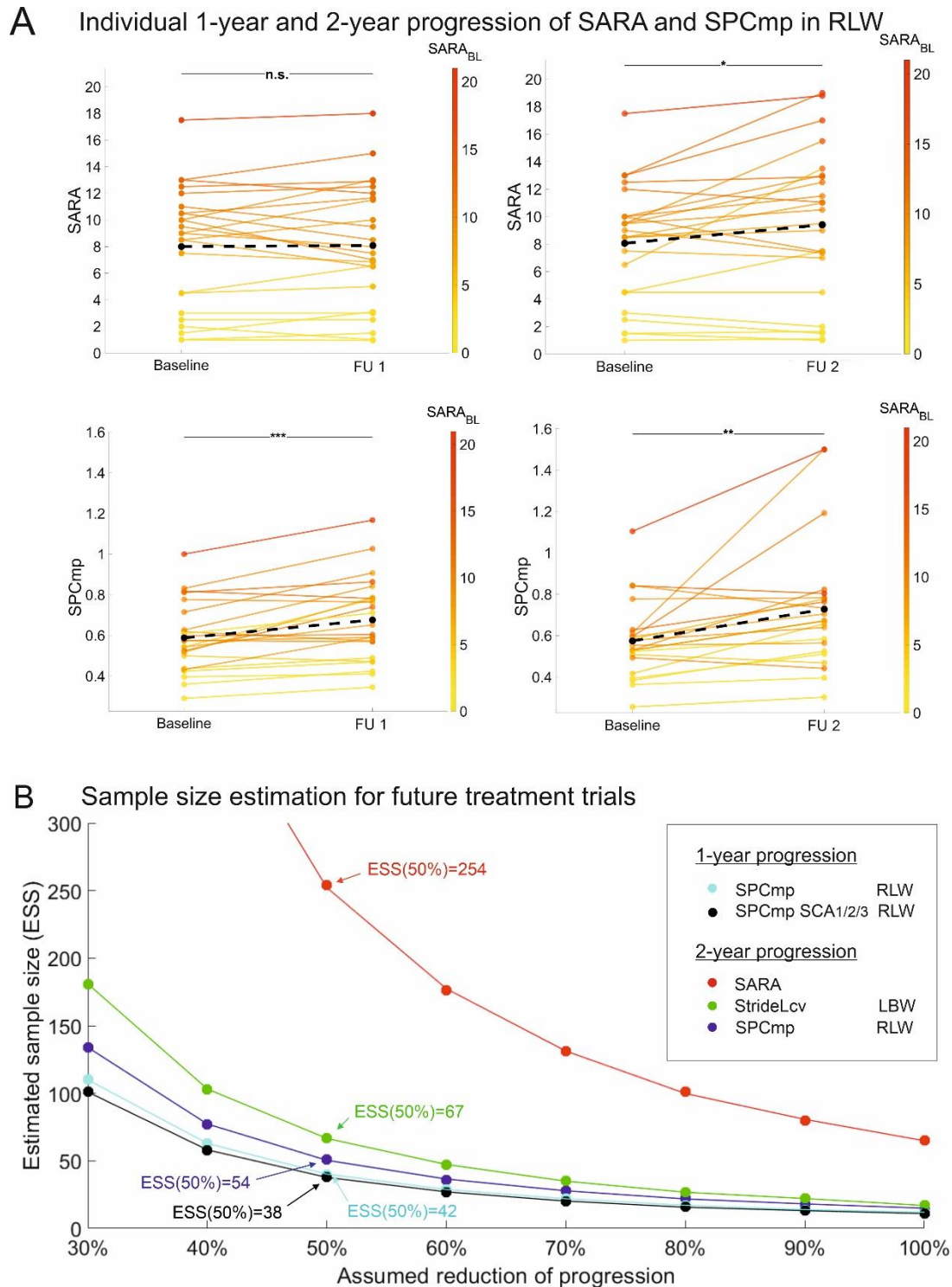


Figure 2 (A) Within-subject changes between baseline and follow-up assessments for the group of DCD participants. Shown are longitudinal differences in SARA (upper panel) and the compound measure SPCmp (lower panels). For both measures, the left panel shows the change between baseline and 1-year follow-up (FU1), and the right panel the change between baseline and 2-year follow-up (FU2). In all panels, SARA scores of individual cerebellar participants are color-coded. Black dotted line = mean change across all participants. Stars indicate significant differences between time points (\* $\equiv$   $p < 0.05$ , \*\* $\equiv$   $p < 0.0083$  Bonferroni-corrected, \*\*\* $\equiv$   $p < 0.001$ ). (B) Sample size estimates were made for future treatment trials that showed different levels of progression reduction for the various outcome measures: SARA and the gait measures SPCmp and StrideLcv in the laboratory (LBW) and real life (RLW). The estimated number of participants per study arm is plotted against the hypothesized therapeutic effect on reducing 1-year progression or 2-year progression in DCD patients or the SCA<sub>1/2/3</sub> subgroup (SCA<sub>1/2/3</sub>), respectively.

## Supplementary information

### Supplement E1- Matching of walking bouts based on bout length and number of turns

For matching, an Euclidean distance matrix  $D$  was created for each subject using bout length and #turns for each baseline bout to each follow-up bout (see Figure S1). To weight both criteria equally, the data were previously standardized per subject (centered and normalized to one standard deviation of the participants' baseline measurement). Using the distance matrix  $D$  for each subject as a cost matrix, a linear assignment problem was solved to obtain a 1:1 matching of similar bouts for baseline and follow-up. A cost for non-assignment of 0.5 was set, to obtain bouts that were as similar as possible, but at the same time to avoid losing any participants for the analysis. Thus, for each subject, only bouts with similar macroscopic gait descriptors are compared. Mismatched bouts are not included in the analysis, reducing purely contextual longitudinal differences.

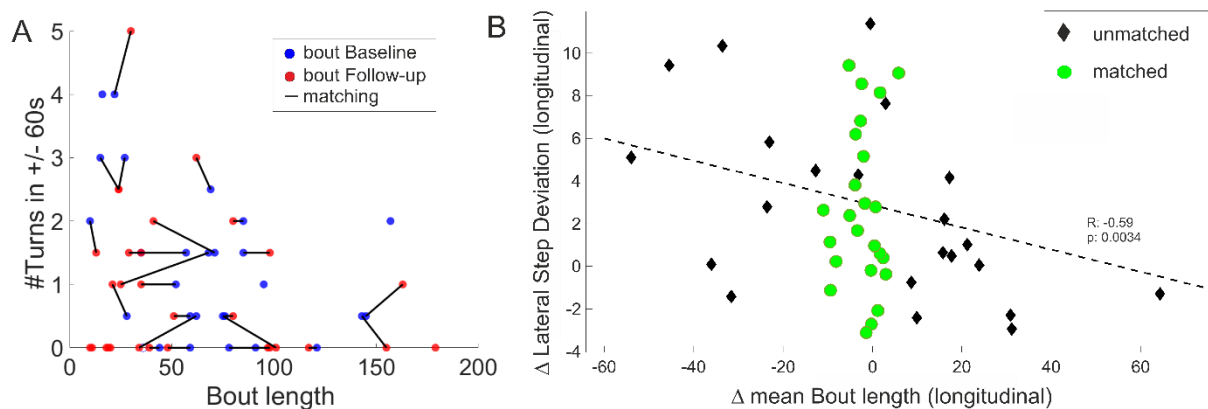


Figure S1 (A) Illustration of the Mapping procedure of corresponding real-life walking trials at baseline (blue) and the one-year follow-up (red). (B) Relationship between longitudinal differences ( $\Delta$ = one-year follow-up – baseline) differences in mean bout length and Lateral Step deviation for the DCD participants (non-matched: red; matched: blue). In the non-matched condition, there is a significant correlation between the longitudinal difference in Lateral Step Deviation and the difference in mean bout length ( $r=0.59$ ,  $p=0.0034^{**}$ ).

## Supplement E2- Details on gait measures

Lateral step deviation (*LatStepDev*): This gait measure was determined based on three consecutive walking steps, calculating the absolute amount of perpendicular deviation of the middle foot placement from the line connecting the first and the third step (Figure S2). *LatStepDev* was normalized with stride length (% of stride length), thus providing a measure independent from stride length variability, which is suggested to be increased in real-life gait.

Coronal Range of Motion (*CorRoM*): The angular range of the thoracic spine in the coronal plane (roll).<sup>12</sup>

Toe Out Angle (*ToeOutAng*): The lateral angle of the foot during the stance phase, relative to the forward motion of the gait cycle. Positive angle is outward rotation.<sup>12</sup>

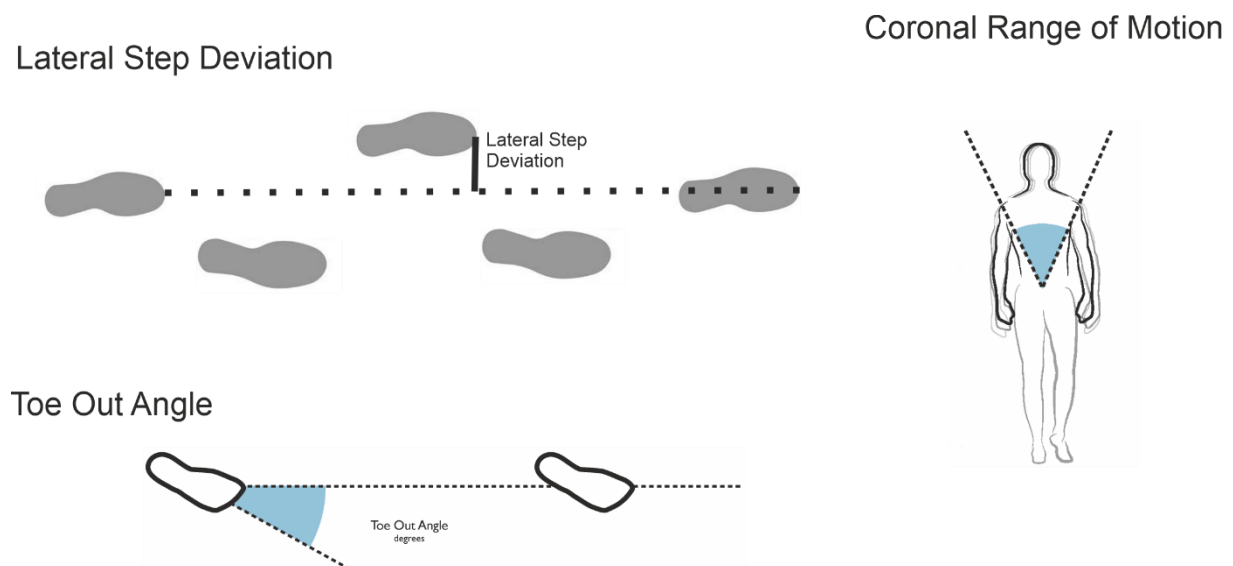
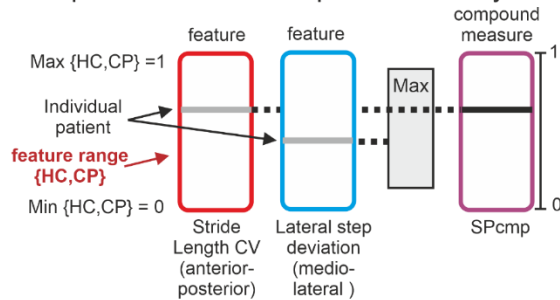


Figure S2 Illustration of the gait measures: Lateral step deviation, Toe Out Angle, and Coronal Range of Motion (Adopted from<sup>12</sup>).

## Supplement E3 – Compound measure of spatial variability

The spatial step variability compound measure  $SPCmp$  was determined in two steps: step one determines for each of the two parameters ( $StrideL_{CV}$ ) and ( $LatStepDev$ ) separately the relative value of an individual subject in comparison to the value range of all participants (resulting in values between [0-1], see Figure S3 A). In step 2, that measure out of these two measures was taken for final analysis where the individual's result showed a larger abnormality (shown by a value nearer to 1), whereas the respective other measure was not entered into the further analysis, see Figure S3-A. For the longitudinal analysis, the value range of all participants is determined based on the baseline assessment. Therefore, values in the follow-up assessment can exceed the range [0-1].

### A Compound measure of spatial variability $SPCmp$



### B Composition of the compound measure $SPCmp$

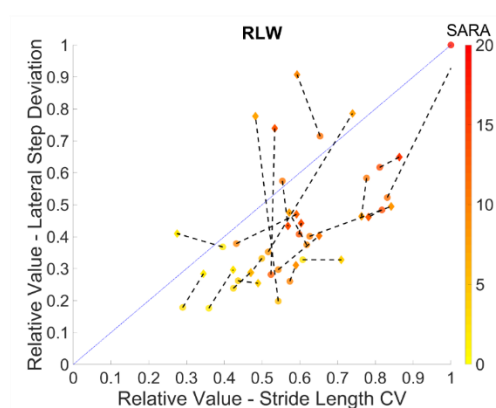


Figure S3 (A) Determination of the compound measure  $SPCmp$ . It is determined by the maximum of the relative values for the parameters  $StrideL_{CV}$  and  $LatStepDev$ . (B) Composition of the compound measure  $SPCmp$  for the walking conditions RLW in Baseline and the first follow-up assessment. Shown are the relative parameter values of each patient for the parameters  $StrideL_{CV}$  (x-axis) and  $LatStepDev$  (y-axis). The color-coding denotes the severity of gait and posture ataxia as determined by the SARA score. Dotted lines connect the baseline and the follow-up examination for the individual patients.

## Supplement E4

### Linear regression models of the matching procedure

We performed linear regression models to examine the influence of macroscopic gait characteristics on our variability measures.

#### Linear regression models:

The regression models examined the predictability of the gait measures lateral step deviation and stride length from different high-level gait descriptors like bout length, number of turns, SARA, or age of the participants. The regression was conducted on all ataxic and pre-ataxic participants in the baseline and follow-up condition, where each bout represented one data point.

Linear regression model:

$$\text{LateralStepDeviation} \sim 1 + \text{Boutlength} + \#\text{Turns} + \text{SARA} + \text{Age}$$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	0.042623	0.0011047	38.583	4.1261e-276
Boutlength	-4.0261e-05	3.2212e-06	-12.499	3.5455e-35
#Turns	0.00035816	6.5527e-05	5.4659	4.8952e-08
SARA	0.001407	5.1198e-05	27.481	4.025e-152
Age	-3.327e-05	2.3935e-05	-1.39	0.16459

Root Mean Squared Error: 0.0154

R-squared: 0.223, Adjusted R-Squared: 0.222

F-statistic vs. constant model: 280, p-value = 7.86e-212

Linear regression model:

$$\text{StrideLength}_{cv} \sim 1 + \text{Boutlength} + \#\text{Turns} + \text{SARA} + \text{Age}$$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	0.055889	0.0037609	14.861	1.245e-48
Boutlength	-7.3048e-05	1.0966e-05	-6.6611	3.0976e-11
#Turns	0.0013206	0.00022308	5.9199	3.4995e-09
SARA	0.002412	0.0001743	13.838	1.5078e-42
Age	-0.00047262	8.1485e-05	-5.8	7.1564e-09

Root Mean Squared Error: 0.0524

R-squared: 0.0733, Adjusted R-Squared: 0.0723

F-statistic vs. constant model: 77, p-value = 6.05e-63

## Supplement E5 – Comparison of statistics between matched and non-matched walking bouts

Table S5 Number of matched walking bouts and strides in the real-life walking condition (RLW). Longitudinal within-subject comparison of bout length in real life (RLW) for baseline (BL) and 1-year follow-up assessment (FU) before and after pre-processing. p-values determined by Wilcoxon signed-rank test for both follow-up assessments relative to baseline. Shown are analyses for the whole group of degenerative cerebellar disease (DCD).

RLW	Bout length (#Strides)		p Δ(FU-BL)	#Bouts		#Strides	
	BL	FU		BL	FU	BL	FU
Non-matched	47.2±32.6	61.86±48.7	0.530	1450	1404	38507	53610
Matched	58.6±63.9	57.5±57	<b>0.029</b>	447	447	26872	26306

## Supplement E6 – Longitudinal within-subject comparison for the SCA<sub>1/2/3</sub> subgroup

Table S6 Longitudinal within-subject comparison for the SCA<sub>1/2/3</sub> subgroup. Shown are results of clinical ataxia ratings (SARA score and SARA<sub>p&g</sub> posture&gait subscore<sup>41</sup>) as well as of gait measures in clinical assessment (LBW) and in real life (RLW) for baseline, 1-year and 2-year follow-up assessments. Friedman test determined within-group longitudinal differences (<sup>+</sup>, p<0.1). Post-hoc test **p**-values determined by Wilcoxon signed-rank test for both follow-up assessments relative to baseline. Stars indicate significant differences between groups (\*≡ p<0.05, \*\*≡ p<0.0083 Bonferroni-corrected, \*\*\*≡ p<0.001). Effect sizes  $r_{prb}$  determined by matched-pairs rank-biserial correlation<sup>51</sup>. . MDC<sub>90</sub> denotes the smallest reliable detectable change (90% confidence interval). The MDC column indicates whether this is smaller than the change between baseline and 1-year Follow-up (<FU1) or/and between baseline and 2-year Follow-up (<FU1+2).

GROUP		Friedman – test		Baseline	1-year Follow-Up			2-year Follow-Up			MDC <sub>90</sub>
SCA 1/2/3	Measure	$\chi^2$	p	m ± sd	m ± sd	p	$r_{prb}$	m ± sd	p	$r_{prb}$	
Clinician-reported outcomes	SARA	0.5	0.77	8.47±4.9	8.55±4.7	0.59	0.16	8.93±6.0	0.18	0.37	-
	SARA <sub>p&amp;g</sub>	1.3	0.51	3.3±2.2	3.31±2.23	0.78	0.22	3.4±2.6	0.18	0.49	-
	Patient-reported outcome	ABC	2.5	0.28	70.0±2.2	63.3±21.2	0.04*	-0.80	67.1±31.6	0.29	-0.41
Gait -	Speed	8.5	0.01*	1.44±0.14	1.37±0.13	0.01*	0.67	1.29±0.21	0.0026**	0.833	0.02 <FU1+2
LBW	StrideLCV	2.17	0.33	0.019±0.007	0.024±0.01	0.06	0.52	0.029±0.024	0.006**	0.76	0.0064 <FU2
	LatStepDev	3.2	0.20	0.033±0.01	0.034±0.013	0.3	0.29	0.034±0.014	0.25	0.35	0.004
	SPCmp	3.13	0.20	0.384±0.21	0.42±0.21	0.055	0.54	0.457±0.37	0.03*	0.63	0.05 <FU2
	CorRoM <sub>SD</sub>	6.2	0.04*	0.92±0.36	1.06±0.33	0.06	0.52	0.918±0.32	0.84	0.1	0.19
	ToeOutAng <sub>SD</sub>	0.4	0.79	1.3±0.47	1.84±0.6	0.87	0.11	1.94±0.7	0.48	0.21	0.51 <FU2
Gait -	Speed	3.9	0.14	1.24±0.16	1.18±0.12	0.09	0.47	1.16±0.23	0.01*	0.7	0.04 <FU1+2
RLW	StrideLCV	11.9	0.0026**	0.035±0.01	0.047±0.017	0.0008***	0.95	0.0497±0.03	0.04*	0.6	0.0025 <FU1+2
	LatStepDev	2.03	0.36	0.045±0.01	0.047±0.012	0.19	0.36	0.048±0.009	0.04*	0.58	0.0018 <FU1+2
	SPCmp	9.4	0.008**	0.60±0.19	0.706±0.22	0.0023**	0.86	0.71±0.30	0.0067**	0.76	0.02 <FU1+2
	CorRoM <sub>SD</sub>	7.9	0.019*	1.05±0.19	1.14±0.18	0.02*	0.64	1.09±0.26	0.45	0.23	0.2
	ToeOutAng <sub>SD</sub>	4.3	0.12	2.33±0.62	2.62±0.6	0.02*	0.63	2.67±1.02	0.56	0.18	0.22 <FU1+2



## Supplement E7 – Longitudinal within-subject comparison in the non-matched condition

Table S7 Longitudinal within-subject comparison for the group of degenerative cerebellar disease (DCD in the **non-matched condition**. Shown are gait measures in real life (RLW) for baseline, 1-year, and 2-year follow-up assessments. Friedman test determined within-group longitudinal differences (+,  $p < 0.1$ ). Post-hoc test **p**-values determined by Wilcoxon signed-rank test for both follow-up assessments relative to baseline. Stars indicate significant differences between groups (\* $\equiv p < 0.05$ , \*\* $\equiv p < 0.0083$  Bonferroni-corrected, \*\*\* $\equiv p < 0.001$ ). Effect sizes  $r_{prb}$  determined by matched-pairs rank-biserial correlation<sup>51</sup>.

non-matched	Measure	Friedman		Baseline	1-year Follow-Up			2-year Follow-Up		
		$\chi^2$	p	m ± sd	m ± sd	p	$r_{prb}$	m ± sd	p	$r_{prb}$
<b>Gait -</b>	Speed	2.1	0.33	1.22±0.14	1.19±0.13	0.33	0.23	1.17±0.21	0.02*	0.54
<b>RLW</b>	StrideLcv	8.4	0.01 <sup>+</sup>	0.036±0.013	0.04±0.015	0.01*	0.59	0.047±0.032	0.01*	0.58
	LatStepDev	10.8	0.006 <sup>+</sup>	0.044±0.01	0.046±0.010	0.07	0.42	0.047±0.009	0.10	0.39
	SPCmp	14.9	0.0005 <sup>+</sup>	0.57±0.16	0.63±0.2	0.0089**	0.62	0.71±0.40	0.01*	0.58
	CorRoMsd	6.8	0.03 <sup>+</sup>	1.01±0.19	1.08±0.19	0.01*	0.60	1.05±0.21	0.32	0.24
	ToeOutAngleSD	8.2	0.01 <sup>+</sup>	2.35±0.6	2.57±0.58	0.0081**	0.63	2.61±0.73	0.03*	0.52

## Supplement E8 – Correlations of gait measures between lab-based walking and real-life walking

Table S8 Correlations between gait measures in conditions LBW and RLW are given for the DCD group. Effect sizes of correlations are given using Spearman's  $\rho$ . (\* $\equiv$   $p < 0.05$ , \*\* $\equiv$   $p < 0.0083$  Bonferroni-corrected, \*\*\* $\equiv$   $p < 0.001$ ).

Measure	Corr LBW, RLW	
	$\rho$	<b>p</b>
Speed	0.7	0.006**
StrideLcv	0.81	<0.001***
LatStepDev	0.74	0.002**
SPCmp	0.74	0.002**
CorRoM <sub>SD</sub>	0.74	0.002**
ToeOutAng <sub>SD</sub>	0.23	0.40