Check for updates

CPT: Pharmacometrics & Systems Pharmacology

ARTICLE OPEN ACCESS

Integrated Modeling of Digital-Motor and Clinician-Reported Outcomes Using Item Response Theory: Towards Powerful Trials for Rare Neurological Diseases

Alzahra Hamdan[1] | Andreas Traschütz[2,3] | Lukas Beichert[2,3] | Xiaomei Chen[1] | Cynthia Gagnon[4,5,6] | Bart P. van de Warrenburg[7] | Filippo M. Santorelli[8] | Nazlı Başak[9] | Giulia Coarelli[10] | Rita Horvath[11] | Stephan Klebe[12,13] | PROSPAX consortium | EVIDENCE-RND consortium | Rebecca Schüle[2,14] | Andrew C. Hooker[1] | Matthis Synofzik[2,3] | Mats O. Karlsson[1]

[1]Pharmacometrics Research Group, Department of Pharmacy, Uppsala University, Uppsala, Sweden | [2]Division Translational Genomics of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research and Center of Neurology, University of Tübingen, Tübingen, Germany | [3]German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany | [4]Groupe de recherche interdisciplinaire sur les maladies neuromusculaires (GRIMN), Centre intégré universitaire de santé et de services sociaux du Saguenay–Lac-Saint-Jean, Québec, Canada | [5]Centre de recherche du Centre intégré universitaire de santé et de services sociaux du Saguenay–Lac-Saint-Jean, Québec, Canada | [6]Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Québec, Canada | [7]Department of Neurology, Radboud University Medical Center, Nijmegen, the Netherlands | [8]IRCCS Stella Maris Foundation, Pisa, Italy | [9]School of Medicine, KUTTAM-NDAL, Koç University, İstanbul, Turkey | [10]Sorbonne Université, Institut du Cerveau – Paris Brain Institute-ICM, Inserm, CNRS, APHP, Hôpital de la Pitié Salpêtrière, Paris, France | [11]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK | [12]Department of Neurology, University Hospital Essen, Essen, Germany | [13]Department of Neurology, Klinikum Vest, Recklinghausen, Germany | [14]Division of Neurodegenerative Diseases, Department of Neurology, Heidelberg University Hospital, Heidelberg, Germany

Correspondence: Mats O. Karlsson (mats.karlsson@uu.se)

Received: 11 February 2025 | **Revised:** 19 June 2025 | **Accepted:** 7 July 2025

Funding: This work was funded by the European Joint Programme on Rare Diseases (EJP RD) Joint Transnational Call 2019 for the EJP RD WP20 Innovation Statistics consortium "EVIDENCE-RND" and under the EJP RD Grant Agreement (n°825575) (to M.O.K, R.S., and M.S.), and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) No 441409627, as part of the PROSPAX consortium, both under the EJP RD COFUND-EJP N° 825575 (to R.S. and M.S). In addition, it was supported by the European Union, project European Rare Disease Research Alliance (ERDERA, # 101156595) (to M.S., R.S, and M.O.K). Moreover, work in this project was supported by the Clinician Scientist program "PRECISE.net"funded by Else Kröner-Fresenius-Stiftung (to M.S., R.S., A.T, and L.B.). This work was also supported by ZonMw grant number 463002002 (to B.P.W.). The Italian Ministry of Health Ricerca Corrente 2024 supported work in FMS clinic.

Keywords: clinical outcome assessments | digital-motor outcomes | genetic ataxias | Item Response Theory | rare neurological diseases

ABSTRACT

Robust and highly sensitive outcomes are crucial for small trials in rare diseases. Combining different outcome types might improve sensitivity to identify disease severity and progression, yet innovative methodologies are scarce. Here we develop an Item Response Theory framework that allows integrated modeling of both continuous and categorical outcomes (ccIRT). With degenerative ataxias, a group of rare neurological coordination diseases, as a showcase, we developed a ccIRT model integrating two ataxia outcome types: a clinician-reported outcome (Scale for the Assessment and Rating of Ataxia; SARA; categorical data) and digital-motor outcomes for gait and limb coordination (continuous data). The ccIRT model leveraged data from 331 assessments from a natural history study for spastic ataxias. The model describes SARA items and digital-motor outcomes data as functions of a common underlying ataxia severity construct, evaluating 9 *gait* and 17 *limb coordination* digital-motor measures for their ability to add to SARA in estimating individual ataxia severity levels. Based on our proposed workflow for assessing digital-motor

Matthis Synofzik and Mats O. Karlsson shared last authorship.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2025 The Author(s). *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

outcomes in ccIRT models, the final model selected three *digital gait* and three *limb coordination* measures, reducing average uncertainty in ataxia severity estimates by 49% (10% SD) compared to the SARA-only IRT model. Trial simulations showed a 49% and 61% reduction in sample sizes needed to detect disease-modifying effects in two genotypes. Overall, our ccIRT framework for combining multiple outcome domains, even of different variable types, facilitates a more precise estimation of disease severity and a higher power, which is particularly relevant for rare diseases with inherently small and short trials.
Trial Registration: ClinicalTrials.gov: NCT04297891

1 | Introduction

Although more than 7000 distinct rare diseases are known nowadays, only 5% of those have approved treatments [1]. A major bottleneck of therapeutic progress in rare diseases is the lack of innovative methods tailored to provide substantial evidence supporting trial designs and analyses in small cohorts. A critical aspect is the need for robust and sensitive outcome assessments to measure disease progression and treatment effects in limited populations. Disease outcome assessments from several modalities are often used, including clinician-reported outcomes, patient-reported outcomes, digital-motor outcomes, molecular biomarkers, and medical imaging. However, these outcomes are currently analyzed separately when used in observational or randomized treatment trials. Innovative analysis methods that efficiently integrate different outcome modalities with different data types (e.g., continuous and categorical) might enhance power and thus enable more feasible trial designs for rare diseases.

Item Response Theory (IRT) is a statistical tool used to analyze composite data on the item level, thus capturing more information from data compared to simply analyzing total scores [2]. Several studies showed that the IRT method, coupled with pharmacometric methods, increased the statistical power compared to other traditional analyses (e.g., analyses based on total composite score) [3–7]. IRT use is so far limited to clinical outcome assessments with categorical data (e.g., clinician-reported and patient-reported outcomes). In this work, we propose a novel, extended pharmacometric IRT methodology that integrates multiple outcomes that have different data types (e.g., both continuous and categorical data) for further improvement in power.

Degenerative ataxias are a heterogeneous group of rare neurological diseases with promising disease-modifying treatments on the horizon [8]. The disease is characterized by progressive damage to the cerebellum and its associated tracts, affecting patients' gait, balance, and limb coordination ability [9]. Ataxia severity and progression are assessed using various outcome types, including in particular clinician-reported outcomes and digital-motor outcomes [10]. The Scale for the Assessment and Rating of Ataxia (SARA) is the most commonly used clinician-reported outcome in ataxias and is currently often used as the primary outcome in many treatment trials [8, 11]. SARA consists of eight items assessing gait, stance, sitting, speech, and limb movements (finger chase, nose-finger test, alternating hand movements, and heel-shin slide) [11]. IRT analysis of SARA showed that SARA is a well-designed clinical outcome assessment for ataxia [12]. However, given the slow-progressive nature of genetic ataxias, SARA's ability to detect longitudinal changes was insufficient to support treatment trials with small sample sizes and short duration, as is inherently a requirement for rare diseases [7, 13]. Meanwhile, digital-motor outcomes offer objective and sensitive ataxia assessments even in early and pre-ataxic stages [14–17]. Digital-motor outcomes are sensor-based performance outcomes that provide quantitative measures (i.e., continuous data) of (i) gait patterns, for example, using body-worn sensors [18], and (ii) limb coordination, particularly upper limb coordination, for example, using stationary sensors [19, 20]. Digital-motor outcomes are increasingly used in natural history and observational trials for ataxia [10, 16] and in treatment trials for other rare neurological diseases such as Huntington's disease and Duchenne muscular dystrophy [21–24].

With the urgent need for feasible trial designs to advance therapeutic progress in rare neurological diseases, our proposed methodology for integrated modeling of multiple disease outcome modalities can enhance the robustness of disease severity estimation and, hence, enhance the sensitivity to detect longitudinal changes and treatment effects. Using degenerative ataxia as an example, we present here (i) an IRT framework for integrated modeling of multiple outcome domains with both continuous and categorical data, and (ii) an evaluation of the ability of digital-motor outcomes to add to SARA in the assessment of the disease severity of ataxia patients.

2 | Methods

2.1 | Data

Data were obtained from the prospective multicenter natural history study for spastic ataxias PROSPAX [25]. The PROSPAX study aims to improve the current understanding of the disease progression in spastic ataxias, that is, degenerative ataxias with cerebellar and pyramidal features [9]. The study collects longitudinal data from several ataxia outcome domains, including SARA and digital-motor outcomes. Institutional review board approvals and informed consents were obtained, complying with the regulatory standards and Declaration of Helsinki (Tübingen IRB 824/2019BO2).

The dataset used in this analysis consists of 187 spastic ataxia patients of two different genetic subpopulations (74 patients with Autosomal Recessive Spastic Ataxia Charlevoix Saguenay [ARSACS], and 113 patients with Hereditary Spastic Paraplegia Type 7 [SPG7]) with a total of 331 visits. In this analysis cohort, ataxia severity was measured using (i) SARA assessment with individual item scores reported, (ii) *digital gait outcomes* assessed by wearable sensors (Opal APDM Wearable Technologies) [26, 27], and (iii) *digital limb coordination outcomes* assessed by stationary sensors (Q-Motor sensors) [20].

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Summary

- What is the current knowledge on the topic?
  - Robust and sensitive disease outcomes are needed to facilitate the therapeutic progress in rare neurodegenerative diseases. Various outcomes are often used separately to evaluate the disease severity and progression.
- What question did this study address?
  - Can we integrate data across multiple outcome modalities (even if of different data types) using Item Response Theory (IRT) modeling to improve the precision of disease severity estimates and the statistical power of treatment trials?
- What does this study add to our knowledge?
  - A joint IRT model of multiple outcome modalities with continuous and categorical data was introduced. In a case study of rare ataxias, a joint IRT model of digital-motor and clinician-reported outcomes decreased the uncertainty of individual estimates and increased sample sizes by ~50% (compared to clinician-reported outcome IRT analyses).
- How might this change drug discovery, development, and/or therapeutics?
  - The proposed methodology enhances the use of information from multiple outcome modalities, leading to more precise individual severity estimates and decreased sample sizes. This will benefit disease-modifying trials in rare diseases that require small and short studies.

Nine *digital gait* and 17 *digital limb coordination* measures were selected for this study based on previous research and clinical hypotheses. Although SARA and *digital limb coordination* assessments were available for all PROSPAX patients (with missingness in some individual items or measures), *digital gait* measures were available for only 97 patients and 157 visits since they can only be sensibly collected from ambulatory patients. Data description of the digital-motor outcomes is shown in Table 1.

Prior to the analysis, digital-motor outcomes data were standardized (i.e., by subtracting the mean from each data point and dividing by the standard deviation [SD] as shown in Equation 1) and monotonized (i.e., unifying the severity direction of the data where a higher value means more severe ataxia which was done by multiplying data by −1 if the measure has opposite direction). This preprocessing transforms data into a consistent standard format, which facilitates data and results interpretation across all digital measures. A visual description of the raw and processed data is illustrated in Appendix S1.

$$DMO_{im} = \frac{DMO_{im,raw} - mean\left(DMO_{im,raw}\right)}{SD\left(DMO_{im,raw}\right)} \quad (1)$$

The two genotypes ARSACS and SPG7 are also classified as Autosomal Recessive Cerebellar Ataxias (ARCA). Natural history data for a wide range of ARCA genotypes (> 100) were collected in the large prospective multicenter ARCA registry [28, 29]. In previous work, we studied the performance of SARA across ARCA genotypes, including ARSACS and SPG7, using large data from the ARCA registry (990 patients, 1932 visits) which we will refer to as the "ARCA reference population" in this paper [7, 12].

2.2 | Digital-Motor Outcomes

Gait was captured using wearable motion sensors attached to both feet and the posterior trunk at the level of L5 while the patients walked up and down a constrained 10-m stretch for 1 min at self-selected speed. For each detected gait stride, a variety of metrics are quantified, which can be categorized into five gait domains: pace (1 measure), temporal variability (2 measures), spatial variability (2 measures), gait smoothness (2 measures), and foot angle variability (2 measures). Typically, 20–40 strides were detected and the measured data were aggregated using numerical descriptive statistics (e.g., mean, median, coefficient of variation). For further details on the *digital gait* outcomes and their validation for degenerative ataxia, see [26, 27, 30, 31]. Limb coordination was captured using stationary sensors while performing six motor tasks: foot tapping (3 measures), finger tapping (2 measures), alternating hand movement (3 measures), grip-lift (2 measures), spiral drawing (2 measures), and target reaching (5 measures). On a higher level, these measures capture seven features of coordinative limb movement: speed (4 measures), variability (4 measures), efficiency (2 measures), smoothness (2 measures), endpoint precision (1 measure), stability (1 measure), and force control (3 measures). For further details on the *limb coordination* outcomes and their validation for degenerative ataxia, see [20]. A detailed description of the *digital gait and digital limb coordination* measures analyzed in this work is shown in Table 1.

2.3 | Item Response Theory Model of SARA

In previous work, we developed an IRT model for SARA that could describe the observed item-level data using one common latent variable [7, 12]. The model was built using data from the ARCA registry as mentioned above in the data section [28]. In this model, the relationship between the individual latent variable (i.e., ataxia severity of patients) and the probability of the individual to score a specific item response is described using logistic item characteristic functions (ICFs), as shown in Equation (2).

$$P\left(Y_{ij} \geq k\right) = \frac{e^{a_j\left(\psi_i - b_{j,k}\right)}}{1 + e^{\left(a_j\left(\psi_i - b_{j,k}\right)\right)}} \quad (2)$$

where Y_{ij} is the observed score for the ith individual and the jth item, k is the response level for a specific item, a_j is a parameter that describes the discrimination ability of item j, $b_{j,k}$ is the difficulty parameter for the kth level of item j, and ψ_i is the individual latent variable for disease severity. The estimated values of the item parameters (a_j, $b_{j,k}$) were reported previously [12], in which the latent variable scale is defined relative to the ARCA reference population with a standard normal distribution (the mean of 0 and the standard deviation of 1). In other words, the latent

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Description of digital-motor measures in the PROSPAX dataset.

Domain/Task	Measure	No. of patients	Total no. of occasions	Mean value (SD)
Digital gait measures				
Pace	Gait speed (median)	97	157	0.91 (0.24)
Temporal variability	Gait cycle duration (CV)	97	157	0.04 (0.02)
	Swing (CV)	97	157	0.03 (0.02)
Spatial variability	Stride length (CV)	97	157	0.05 (0.03)
	Lateral step deviation (norm)	97	157	0.04 (0.01)
Gait smoothness	Harmonic ratio accelerometer vertical (median)	97	157	2.07 (0.63)
	Harmonic ratio accelerometer mediolateral (median)	97	157	1.70 (0.51)
Foot angle variability	Pitch at toe-off (median)	97	157	30.0 (5.5)
	Pitch at toe-off (MADN)	97	157	1.51 (0.63)
Digital limb coordination measures				
Foot tapping	Speed: tap duration[a] (mean)	160	267	0.31 (0.20)
	Variability: tap duration[a] (SD)	160	267	0.07 (0.06)
	Force control: AUC time-force[a] (CV)	160	267	0.51 (0.26)
Finger tapping	Speed: inter-onset interval[a] (mean)	186	325	0.35 (0.15)
	Variability: tap duration[a] (SD)	186	325	0.03 (0.03)
Alternating hand movement	Speed: inter-peak interval[a] (mean)	183	322	0.56 (0.16)
	Variability: upward hand movement[a] (SD)	183	322	0.06 (0.03)
	Force control: AUC time-force[a] (SD)	183	322	0.45 (0.50)
Grip-lift	Stability: position index[a] (mean)	184	321	1.14 (1.08)
	Force control: grip force[a] (CV)	184	321	0.07 (0.05)
Spiral drawing	Efficiency: distance from template (median)	174	292	0.07 (0.05)
	Smoothness: power in the 1–4 Hz frequency spectrum of speed	183	315	5.68 (2.49)
Target reaching	Speed: inter-target interval (median)	183	321	1140.7 (418.5)
	Variability: dtw[b] of time-speed profile	183	321	5.33 (0.44)
	Efficiency: 2D path length (median)	183	321	32.7 (2.6)
	Smoothness: SPARC[c]	183	321	−1.49 (0.08)
	Endpoint precision: dysmetria	182	313	0.77 (0.78)

Abbreviations: CV, coefficient of variation; MADN, normalized mean absolute deviation; SD, standard deviation.
[a]Nondominant side of the body.
[b]Dynamic time warp.
[c]Spectral arc length.

variable scale is a continuous unbounded scale where 95% of the reference population lies in the range of −2 to 2. Given the estimated ICFs, ψ_i can be estimated for any new population based on their SARA item scores.

In the current study, we used the published model that was built on a larger dataset with a wider range of ataxia [12] than the PROSPAX dataset with additional validation using the new, independent set of data. The validation was done by assessing the

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

agreement in estimated ψ_i values between the published model and a reestimated model based on the PROSPAX dataset. In the first model, the SARA item parameters were fixed to the estimates from the published IRT model [12], and the model was applied to PROSPAX data. Here, the latent variable parameters (i.e., the mean and variance of the population) were estimated. In the second model, the model parameters of the latent variable were fixed to the estimates from the first model to keep the ψ_i values from the two models on the same scale, and the item parameters of the IRT model were reestimated by fitting to PROSPAX dataset.

2.4 | Integration of SARA and Digital-Motor Outcomes

2.4.1 | Joint IRT Model for Continuous and Categorical Data (ccIRT)

IRT models are traditionally used to fit categorical data. Here, we propose a joint IRT framework for modeling both continuous and categorical data, capturing information from multiple outcome domains (digital-motor outcomes [continuous data] and SARA [categorical data] in this study). The IRT model for SARA was extended to incorporate the digital-motor outcomes by adding linear (Equation 3) or nonlinear function (Equation 4) between the latent variable (ψ_i) and each digital-motor measure.

$$DMO_{im} = \theta_{1m} + \theta_{2m} \cdot \psi_i + \varepsilon_{im} \quad (3)$$

$$DMO_{im} = \theta_{1m} + \theta_{2m} \cdot \left((\psi_i + a)^{\theta_{3m}} - a \right) + \varepsilon_{im} \quad (4)$$

DMO_{im} is the value of the mth digital-motor measure for the ith individual, θ_{1m} and θ_{2m} are two parameters of each model for the mth measure, and ε_{im} is the residual error following a normal distribution with mean 0 and variance σ^2_m. Equation (4) describes a nonlinear model where θ_{3m} is a power term, and a is a shift value used to ensure the base term (i.e., $\psi_i + a$) is positive upon fitting the nonlinear model. A value of three was determined for the a parameter based on the range of ψ_i estimates in the studied population. A sensitivity analysis indicated a comparable model fit when the shift parameter value was varied between 3 and 10 (Appendix S2).

In this joint ccIRT model, the ICFs for SARA (Equation 2) were fixed based on the previously published model, and only parameters related to digital-motor measures, which are θ_{1m}, θ_{2m}, and θ_{3m}, were estimated. Selection between linear and nonlinear models was based on the Bayesian information criterion for mixed-effect models [32].

2.4.2 | Assessment Framework of Digital-Motor Outcomes

The joint ccIRT model was developed in multiple model-building and analysis steps. The ultimate goal was to (i) incorporate the digital-motor outcomes in the SARA IRT model, (ii) select

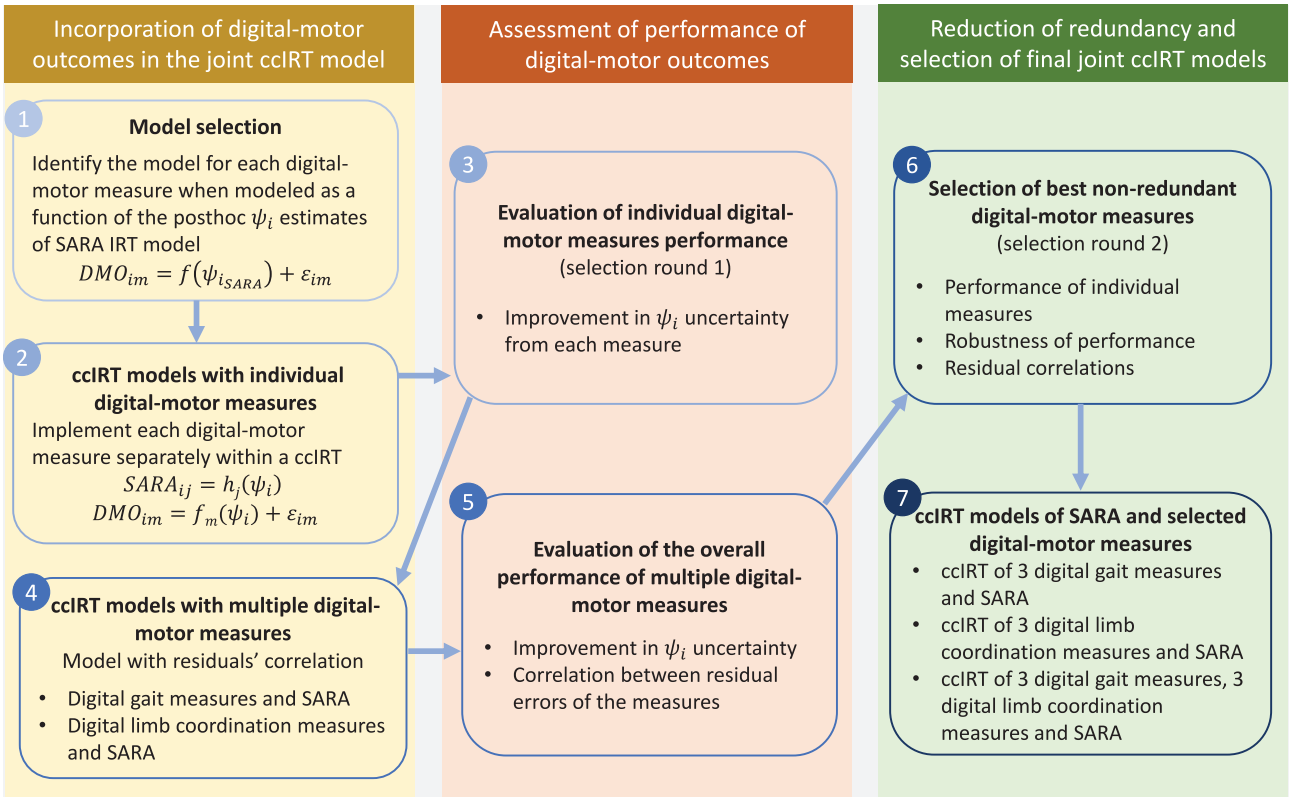


FIGURE 1 | The implementation and assessment workflow of the digital-motor outcomes within a joint Item Response Theory framework for Continuous and Categorical data (ccIRT) along with SARA clinician-reported outcome. The workflow consists of intertwined steps of model building and performance assessment steps [1 to 6] before reaching the final joint ccIRT models (step 7). DMO_{im}, digital-motor outcome value for individual i and measure m; ε_{im}, residual error term; ψ_i, individual latent variable.

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

digital-motor measures based on their ability of supporting SARA in the estimation of individual ataxia severity levels (ψ_i), and (iii) screen for redundancy among the selected measures. The general workflow is illustrated in Figure 1 and summarized in the following steps:

1. Each digital-motor measure was modeled separately as a function of the individual latent variable estimated from the SARA IRT model (ψ_{iSARA}) to select among linear and nonlinear models. The relationship between ψ_{iSARA} values and the digital-motor measures is illustrated in Appendix S1.
2. The selected models from Step 1 were used to create separate ccIRT models for each digital-motor measure. This means that both SARA data and the digital-motor measure contribute to the estimation of the individual latent variable values.
3. The performances of the digital-motor measures in supporting SARA were ranked based on separate ccIRT models and predefined evaluation criteria (the next section) for *digital gait* measures and *digital limb coordination* measures, respectively. The best-performing measures were selected for the next step (selection round 1).
4. The selected *digital gait* measures from Step 3 were implemented in one joint ccIRT model. The correlations between model residuals of the digital-motor measures were estimated. The same procedure was applied to the selected *digital limb coordination* measures.
5. The overall performances of the ccIRT models from Step 4 were evaluated.
6. The joint ccIRT models were reduced based on the results from Steps 3 and 5 (selection round 2).
7. As a final step, a joint ccIRT model of SARA, selected digital gait measures, and selected digital limb coordination measures was implemented (See Appendix S3 for the model code).

2.4.3 | Performance Evaluation and Selection Criteria

The main evaluation criterion of the joint ccIRT model was the ability of the added digital-motor measures to improve the precision of the latent variable estimates compared to SARA-only IRT model. This was measured as the percent decrease in the variance of ψ_i for each patient's visit (Equation 5) and then summarized as mean and standard deviation or median and interquartile range. Furthermore, the robustness of the performance of the digital-motor measure was taken into consideration and was assessed as the frequency across individuals of achieving a higher precision upon the addition of the digital-motor measure to SARA IRT model. Specifically, the number of data occasions showing a particular change (i.e., decrease, increase, or no change) in the uncertainty compared to the SARA-only IRT model was counted. Finally, the magnitude of residual correlations between pairs of digital-motor measures was considered, and high correlations indicated a possibility of reducing the redundancy in the digital-motor measures.

$$\%\ decrease\ in\ uncertainty = \frac{Var(\psi_{iSARA}) - Var(\psi_{iSARA+DMO})}{Var(\psi_{iSARA})} * 100 \quad (5)$$

Moreover, the clinical significance and the captured ataxia aspect (gait metric in digital gait outcomes and movement feature in digital limb coordination outcomes) were considered upon selection among redundant outcomes. Finally, the agreement in the ψ_i estimates between the SARA-only IRT model and the joint ccIRT model was evaluated by calculating the percentage of data where $\psi_{iSARA+DMO}$ estimate falls within the 95% confidence interval of ψ_{iSARA} estimate.

2.5 | Clinical Trial Simulations and Sample Size Calculations

The stochastic simulation and estimation (SSE) algorithm was used to simulate and analyze ataxia trials for sample size calculations ($N_{\text{simulations}} = 250$) and type I error checks ($N_{\text{simulations}} = 1000$). Two-year trials with placebo-controlled balanced parallel design were simulated for the two PROSPAX genetic subpopulations (ARSACS and SPG7, separately) with repeated measurements every 6 months (total of 5 samples per subject). The final developed ccIRT model was used to simulate SARA item scores and digital-motor outcomes data. The model was extended to introduce longitudinal changes of the latent variable given the published longitudinal IRT estimates of the natural history progression of ataxia in ARSACS and SPG7 populations [7]. Hypothetical disease-modifying drug effects (i.e., inhibition of progression rate) were set at 0% and 100% magnitudes for type I error and sample size calculations, respectively. Patients were simulated with latent variables at baseline ranging from −2 to 0, corresponding to 95% probability of *Gait* + *Stance* SARA scores ≤ 7.

Simulated data were analyzed using four different analysis models: (i) IRT model of SARA only, (ii) ccIRT model of SARA and *digital gait* measures, (iii) ccIRT model of SARA and *digital limb coordination* measures, and (iv) ccIRT model of SARA, and both *digital gait* and *digital limb coordination* measures. Type I error was calculated as the percent of trials with false-positive drug effect. Sample sizes were calculated for 80% power with a significance level of 0.05 using the Parametric Power Estimation (PPE) algorithm [33]. Further details on the simulation framework and sample size calculations are available in Appendix S4.

2.6 | Software

Models were implemented using the nonlinear mixed effects modeling software NONMEM [34] version 7.5.0 and 7.5.1, with parameter estimation conducted using a Laplace approximation

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the likelihood. The SSE tool was used from Perl-speaks-NONMEM (PsN) [35] version 5.3.1.

3 | Results

3.1 | Validation of SARA IRT Model

The ψ_i estimates of PROSPAX patients from the published and the updated IRT model were in large agreement with a correlation coefficient of 0.99. All points closely align with the identity line when plotted in a scatterplot. See Appendix S5 for the detailed results, including a comparison between model parameters from both models.

3.2 | Joint ccIRT Model of SARA and Digital-Motor Outcomes

3.2.1 | Modeling and Selection of Digital-Motor Outcomes

The IRT framework of SARA was extended to incorporate digital-motor outcomes following the workflow described in Figure 1. The selected best model for each digital-motor measure and model parameter estimates are shown in Appendix S6. Based on the identified best models (Step 1), we built the joint ccIRT models (Step 2 and Step 4), assessed the performance of individual digital-motor measures (Step 3 and Step 5), and reduced the redundancy among measures (Step 6).

Adding the different *digital gait* measures to the SARA IRT model improved the precision of ψ_i estimates to varying extents, with a median decrease in ψ_i variance ranging 2.1%–49.4%. The five top-ranked measures were propagated to the next step, where they were modeled together in a joint ccIRT model with SARA. Based on (i) the estimated residual correlation matrix and (ii) the robustness of the measure performance (i.e., the number of data occasions where the ψ_i uncertainty has improved upon adding the *digital gait* measure to the SARA IRT model), we selected three measures in the final step. Detailed results of the assessment workflow and residual correlations are shown in Figure 2. Even though *Swing (CV)* ranked the top in uncertainty improvement (median: 49.9%, IQR: 7.6%–84.9%), it also showed a worsening in the uncertainty of ψ_i in 15 out of 157 assessments. *Swing* had high residual correlations with *CV of stride length* (4th rank) and *CV of gait cycle duration* (2nd rank) with $r = 0.68$ and 0.62, respectively. As a result, Swing (CV) was not selected. *MADN of pitch at toe-off* (3rd rank) had minimal correlations with the other top four digital gait measures. Hence, *CV of gait cycle duration*, *MADN of pitch at toe-off*, and *CV of stride length* were selected as the best (and unique) *digital gait* measures supporting SARA assessment of ataxia severity.

Following the same workflow, we assessed the performance of *digital limb coordination* measures. Each of the *digital limb coordiantion* measures improved the (median) ψ_i uncertainty by 0.6%–22.9%, and three measures were selected in the final step: the *speed of target pointing* (inter-target interval), the *variability of target reaching* (dynamic time warp of time-speed curve), and the *smoothness of spiral drawing* (cumulative power in the 1–4 Hz frequency spectrum; Figure 3). These three measures showed low residual correlations between each other (max correlation is 0.27). In the intermediate primary selection step, 12 *digital limb coordination* measures were selected (compared to 5 in digital gait measures) to ensure the inclusion of measures from different tasks (e.g., target reaching) and feature groups (i.e., speed, variability, efficiency, smoothness, stability, and force control).

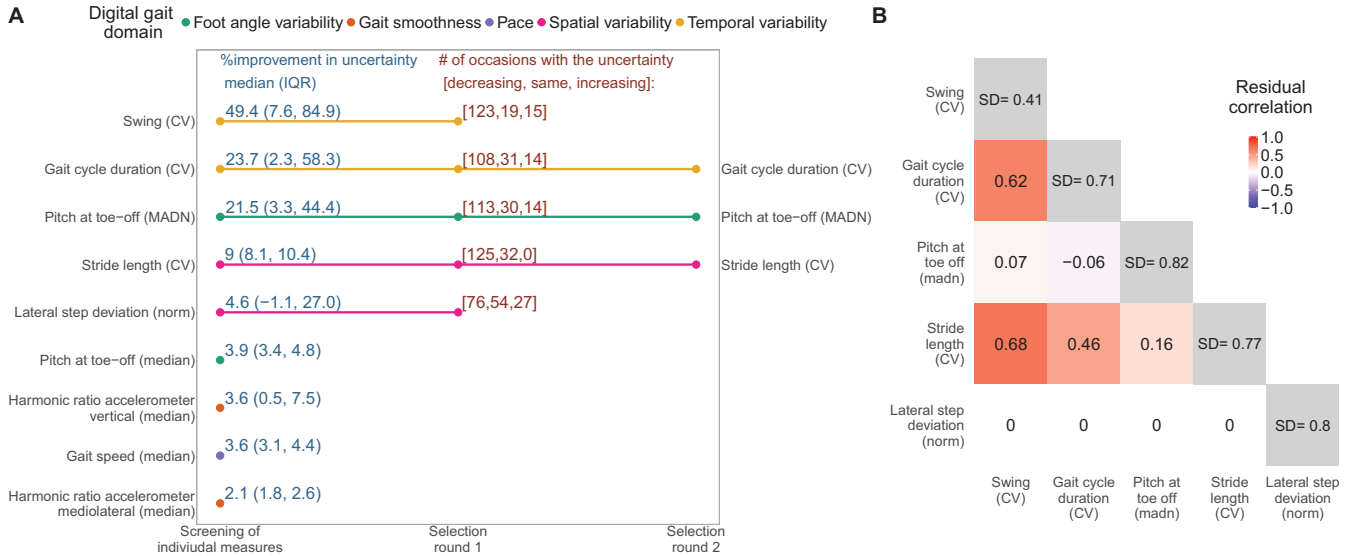


FIGURE 2 | Assessment results of digital gait measures in joint ccIRT models with SARA. Panel A: Performance and selection results of the digital gait measures starting with a screening step of each of the nine measures in separate joint models with SARA, then the first selection round of best measures (5 measures selected), and the final selection of measures (3 measures). Panel B: Correlation matrix of residual error terms in the big joint ccIRT model of selection round 1 measures and SARA. The standard deviation of the residual variability of each measure is shown on the diagonal elements. CV, coefficient of variation; IQR, interquartile range; MADN, normalized mean absolute deviation; SD, standard deviation.

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



FIGURE 3 | Assessment results of digital limb coordination measures in joint ccIRT models with SARA. The measures are named following the format "Task: Feature" (for full names see Table 1). Panel A: Performance and selection results of digital measures starting with a screening step of each of the 17 measures in separate joint models with SARA, then the first selection round of best measures (12 measures), and the final selection of measures (3 measures). Panel B: Correlation matrix of residual error terms in the big joint ccIRT model of selection round 1 measures and SARA. The standard deviation of the residual variability of each measure is shown on the diagonal elements.

3.2.2 | Performance of Joint ccIRT Models

Based on the final selection of digital-motor measures (Figures 2A and 3A), ccIRT models of SARA and either one or both of the digital-motor domains were built to estimate the ataxia severity of PROSPAX patients.

ccIRT of SARA and *digital gait* outcomes: the model resulted in a decrease of ψ_i uncertainty in 87% of the 157 occasions compared to the SARA-only IRT model. The mean of the standard error (SE) of ψ_i estimates decreased from 0.30 in the SARA IRT model to 0.24 in the ccIRT model with a mean decrease in ψ_i variances of 33%. The ψ_i estimates from both models were highly correlated with a Pearson's correlation coefficient (r) of 0.91 indicating good agreement between the models.

ccIRT of SARA and *digital limb coordination* outcomes: the model showed a decrease in ψ_i uncertainty in 98% of the occasions (total 326). The mean of ψ_i SE decreased from 0.30 to 0.24 with a mean decrease in ψ_i variances of 36%. The ψ_i estimates from both models were highly correlated ($r = 0.95$).

ccIRT of SARA, *digital gait* outcomes, and *digital limb coordination* outcomes: the model was built based on data occasions with existing assessments of all three outcomes (152 visits). Compared to SARA IRT model, the uncertainty of ψ_i estimates decreased in all data occasions with a mean of ψ_i SE decreasing from 0.30 to 0.21. In other words, the variance of the ψ_i estimates improved by 49% on average. The ψ_i estimates were close to the estimates from the SARA IRT model ($r = 0.91$). The contribution of individual model components (SARA, digital gait outcomes, and digital limb coordination outcomes) to the informativeness of the final ccIRT model is shown in Figure 4.

Figure 4A shows that the ataxia severity estimates varied depending on the availability of model components with the most prominent difference in the ψ_i distribution given *digital gait* data only. Figure 4B shows that SARA had the highest contribution to the informativeness of the developed ccIRT model (mean ψ_i SE of 0.30), followed by *digital limb coordination* outcomes (mean ψ_i SE of 0.35), and then *digital gait* outcomes (mean ψ_i SE of 0.41). Figure 4B indicates a potential differential contribution of the digital-motor outcomes relative to the ataxia severity level (over the SARA range of 4–21) with a higher reduction in uncertainty and, hence, higher informativeness in more severe ataxia, particularly given *digital gait* outcomes.

3.3 | Clinical Trial Simulations and Sample Size Calculations

Simulated data were analyzed using SARA as the sole outcome (SARA IRT model) or both SARA and digital-motor outcomes in joint ccIRT models. Sample sizes and Type I error rates are shown in Figure 5. Sample size calculations for additional trial scenarios are available in Appendix S4. Type I error was controlled under all analysis models for both ARSACS and SPG7 trials, with false-positive rates (4.2–6.0) falling in the 95% confidence interval of 5% rate [3.8–6.5]. When analyzed using the SARA IRT model, sample sizes of 356 and 1780 patients were needed to detect 100% disease-modifying drug effects in ARSACS and SPG7 trials, respectively, with 80% power. The sample sizes decreased by 20% (ARSACS) and 37% (SPG7) when analyzed using the ccIRT model of SARA and *digital gait* outcomes and by 34% (ARSACS) and 49% (SPG7) when analyzed using the ccIRT model of SARA and *digital limb coordination*. Finally, when assessed using the three outcome

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

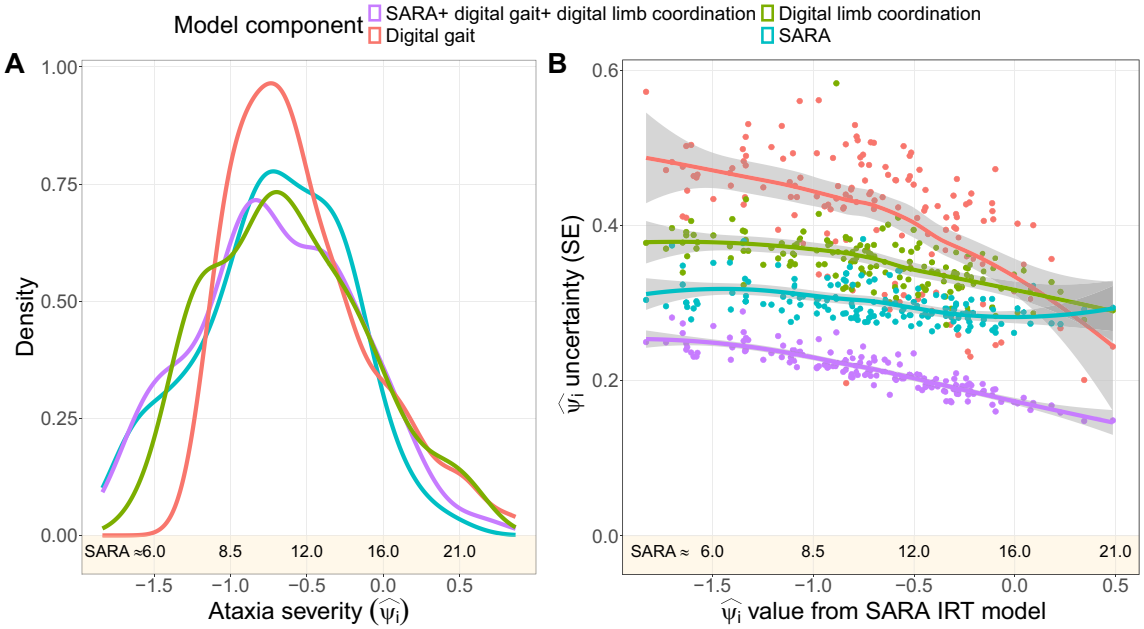


FIGURE 4 | The performance of the final ccIRT model of SARA, digital gait, and digital limb coordination outcomes and the differential contribution of the model components, in assessing the individual latent variable ($\widehat{\psi_i}$) underlying ataxia outcomes (i.e., ataxia severity). (A) The distribution of the $\widehat{\psi_i}$ estimates given data of the different ccIRT model components. (B) The relationship between the uncertainty of $\widehat{\psi_i}$ expressed as the standard error (SE) of the estimate and the value of $\widehat{\psi_i}$ (estimated using SARA IRT model). Only PROSPAX visits with reported assessments from all three outcome modalities (SARA, digital gait, and digital limb coordination) were included in the comparison ($n = 152$). Total SARA scores corresponding to different ψ_i values are reported in the shaded area above the x-axes.

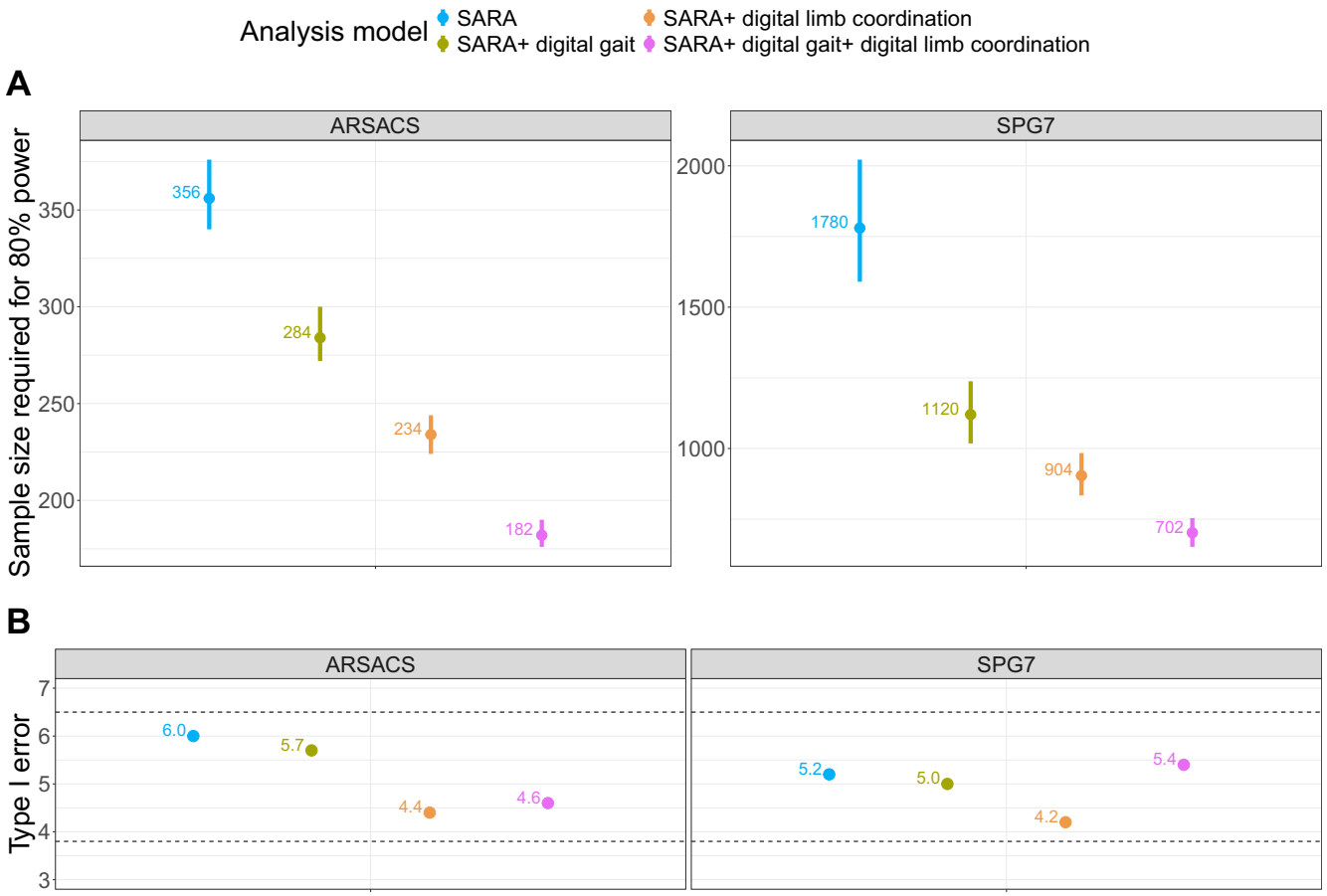


FIGURE 5 | Sample size required to detect disease-modifying drug effect with 80% power (A) and Type I error (B) under multiple analysis models with different components of the ccIRT model. The vertical lines in panel A correspond to the 95% confidence interval of estimated sample sizes. Randomized (1:1), parallel, two-year trials were simulated for ARSACS and SPG7 ataxia populations with a 100% disease-modifying drug effect.

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

modalities (ccIRT of SARA and both types of digital-motor outcomes), the sample sizes decreased by 49% and 61% compared to the SARA-only IRT analysis in ARSACS and SPG7 trials, respectively.

4 | Discussion

Sensitive outcomes are crucial for estimating disease progression and treatment effects in small, short-duration trials for rare degenerative diseases. In this work, we showed how integrating multiple outcome domains within an IRT framework, regardless of variable types, enables a more precise assessment of disease severity.

Specifically, we demonstrated the ability of digital-motor outcomes to support and improve the most widely used clinician-reported outcome assessment in degenerative ataxias: SARA assessment. Leveraging data from the trial-like PROSPAX natural history study, we successfully extended the published IRT model of SARA [12] to incorporate digital-motor outcomes. To the extent of our knowledge, this represents the first case of pharmacometric IRT modeling of both continuous and categorical data. This framework allows for estimating patients' ataxia severity levels underlying observed data of SARA and digital-motor measures with less uncertainty.

Among nine digital gait measures and 17 digital limb coordination measures, we proposed three measures from each digital-motor type that are sufficient to improve SARA estimation of individual ataxia severity levels. This is reflected by the improvement in the precision of individual latent variable estimates, particularly upon adding digital-motor measures of both domains (33%, 36%, and 49% for SARA and *digital gait* measures, SARA and *digital limb coordination* measures, and SARA and both types of *digital-motor outcomes*, respectively). These magnitudes are deemed substantial relative to SARA items' contribution, where the mean improvement ranged between 6% and 23%. Details are shown in Appendix S7. Within the studied population (mild to moderate ataxia; SARA 4–21), the digital-motor outcomes, in particular *digital gait* outcomes, showed increased informativeness in more severe ataxia levels and hence more precise estimates compared to less severe ataxia levels (Figure 4).

Clinical trial simulations with hypothetical 100% disease-modifying drug effects in ARSACS and SPG7 populations showed that the improvement in the precision of ataxia severity estimates gained upon adding the digital-motor outcomes to the SARA IRT model leads to a corresponding decrease in the required trial sample size compared to SARA-only IRT analysis (Figure 5). The inclusion of digital-motor measures from the two ataxia domains (digital gait and digital limb coordination) in the analysis IRT model on top of SARA results in a larger decrease in sample size compared to the addition of measures from one domain only (e.g., a decrease by 49% when ARSACS trial was analyzed using SARA and both *digital gait* and *digital limb measures* vs. 20% when analyzed using SARA and *digital gait* only). The larger sample sizes needed for SPG7 trials compared to ARSACS trials (e.g., 702 vs. 182 patients under ccIRT of the three outcome domains) can be explained by the slower natural history progression in SPG7 [7].

The selected digital gait measures represent three out of four group categories: *CV of gait cycle duration* as a measure of gait smoothness, *MADN of pitch at toe-off* as a measure of the variability of foot angle as it leaves the floor, and *CV of stride length* as a variability measure of the spatial distance traveled by a foot during a gait cycle. These measures have previously been validated in various ataxias across different genotypes [30, 31], and, more specifically, in the two spastic ataxias studied here: ARSACS [26] and SPG7 [27]. This validation demonstrates discriminative validity of the measures against healthy controls and convergent validity with clinical and patient-relevance scales. The findings of the current study further support the potential value of these measures as progression and treatment outcomes in ARSACS and SPG7, and possibly other ataxias. The selected digital limb coordination measures capture limb coordination across multiple movement features, including *speed* of target reaching, *variability* of target reaching, and *smoothness* of spiral drawing. Interestingly, this data-driven selection of optimal digital limb coordination measures aligns with the clinical understanding of ataxia, which has led to the implementation of both spiral drawing and target reaching as digital limb coordination tasks to capture kinematic impairments in upper limb ataxia [20].

It is important to keep in mind that the evaluation of the digital-motor outcomes in this work was limited to the assessment of their ability to *add* to the SARA in estimating individual ataxia severity levels. Hence, the digital-motor outcomes were selected to reflect the same underlying concept as in SARA. In future investigations, digital-motor outcomes can be modeled separately from SARA, where the relative performance of individual measures and their description of the underlying concept can differ from the joint IRT models. In addition, the relative informativeness of digital-motor outcomes at different ataxia stages can be studied more thoroughly and at a wider range of ataxia severity.

Previous studies with simulated degenerative ataxia trials showed that larger sample sizes are needed if simpler models, such as linear mixed-effect models, or standard hypothesis testing (e.g., *t*-test) were used for the analysis of SARA data [7, 36]. Despite the significant decrease in the sample sizes achieved by the joint IRT analysis of SARA and digital-motor outcomes compared to SARA-only IRT analysis, the required sample sizes are still not feasible for short trials in rare and ultra-rare ataxias, even for best-case treatment effect scenarios with a total stop of the disease progression. Integrating additional ataxia outcomes (e.g., fluid biomarkers and imaging markers) may lead to further reduction in ataxia trial sizes. This is a potential topic for future research. Trial analyses based on individual disease progression predictions are likely to benefit directly from the improvement in the precision of individual trajectories, which may lead to feasible N-of-1 trials [37]. Prediction of individual trajectories for ataxia patients has been recently introduced based on SARA scores and can benefit from joint modeling of multiple outcomes [13]. The impact of adding the digital-motor outcomes to the confidence interval of PROSPAX patients' severity estimates is

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

shown in Appendix S8. Furthermore, our proposed integrated framework of SARA and additional outcome types (digital-motor outcomes in this work) can serve as a bridge with the currently available natural history studies of ataxia, which weigh heavily on SARA assessments. Such bridging is possible through the common latent variable underlying SARA and other ataxia outcomes. This will facilitate the use of natural history data to inform trial designs in rare genetic ataxias and possibly allow for external natural history controls. Finally, this methodology could also be applied to other diseases, including rare degenerative conditions, or to other types of disease outcomes.

5 | Conclusion

The proposed methodology of joint IRT modeling of multiple outcome modalities, even of different variable types (continuous data, categorical data), facilitates a more precise estimation of the disease severity of patients as well as a more powerful assessment of longitudinal changes and treatment effects, particularly in rare neurological diseases where small, powerful trials are inherently needed.

Author Contributions

A.H., A.T., L.B., X.C., A.C.H., M.S., and M.O.K. wrote the manuscript; A.H., A.T., L.B., X.C., R.S., A.C.H., M.S., and M.O.K. designed the research; A.H. performed the research and analyzed the data; C.G., B.P.W., F.M.S., N.B., G.C., R.H., and S.K. contributed to data collection.

Acknowledgments

Members of the Evidence-RND consortium include Emmanuelle Comets, France Mentré, Nicole Maria Heussen, Niels Hendrickx, Oleksandr Sverdlov, Ralf-Dieter Hilgers, Thomas Klockgether, and Yevgen Ryeznik. We are grateful to Ralf Reilmann at the George Huntington Institute (Münster, Germany) for providing hardware and algorithms for digital-motor analysis of limb coordination.

Conflicts of Interest

M.S. has received consultancy honoraria from Ionis, UCB, Prevail, Orphazyme, Servier, Reata, Biogen, GenOrph, AviadoBio, Biohaven, Zevra, Solaxa, and Lilly, all unrelated to the present manuscript. A.C.H. and M.O.K. have received consultancy fees from and own stock in Pharmetheus, all unrelated to this manuscript. As an Associate Editor for *CPT: Pharmacometrics and Systems Pharmacology*, A.C.H. was not involved in the review or decision process for this paper. B.P.W. is supported by research grants from ZonMw, Hersenstichting, Dutch Research Council, and Christina foundation; receives research support from Servier and UCB; and has served on scientific advisory boards and/or provided paid consultancy for Biogen, Vico Therapeutics, and Biohaven Pharmaceuticals. All other authors declared no competing interests in this work.

Data Availability Statement

Data supporting the study findings are available on request.

References

1. S. Nguengang Wakap, D. M. Lambert, A. Olry, et al., "Estimating Cumulative Point Prevalence of Rare Diseases: Analysis of the Orphanet Database," *European Journal of Human Genetics* 28, no. 2 (2020): 165–173.

2. S. Ueckert, "Modeling Composite Assessment Data Using Item Response Theory," *CPT: Pharmacometrics & Systems Pharmacology* 7, no. 4 (2018): 205–218.

3. S. Balsis, A. A. Unger, J. F. Benge, L. Geraci, and R. S. Doody, "Gaining Precision on the Alzheimer's Disease Assessment Scale-Cognitive: A Comparison of Item Response Theory-Based Scores and Total Scores," *Alzheimer's & Dementia* 8, no. 4 (2012): 288–294.

4. S. Ueckert, E. L. Plan, K. Ito, M. O. Karlsson, B. Corrigan, and A. C. Hooker, "Improved Utilization of ADAS-Cog Assessment Data Through Item Response Theory Based Pharmacometric Modeling," *Pharmaceutical Research* 31, no. 8 (2014): 2152–2165.

5. Y. K. Lyauk, D. M. Jonker, T. M. Lund, A. C. Hooker, and M. O. Karlsson, "Item Response Theory Modeling of the International Prostate Symptom Score in Patients With Lower Urinary Tract Symptoms Associated With Benign Prostatic Hyperplasia," *AAPS Journal* 22, no. 5 (2020): 115.

6. S. Buatois, S. Retout, N. Frey, and S. Ueckert, "Item Response Theory as an Efficient Tool to Describe a Heterogeneous Clinical Rating Scale in De Novo Idiopathic Parkinson's Disease Patients," *Pharmaceutical Research* 34, no. 10 (2017): 2109–2118.

7. A. Hamdan, N. Hendrickx, A. C. Hooker, et al., "Longitudinal Analysis of Natural History Progression of Rare and Ultra-Rare Cerebellar Ataxias Using Item Response Theory," *Clinical Pharmacology & Therapeutics* 116, no. 6 (2024): 1593–1605.

8. M. Synofzik, H. Puccio, F. Mochel, and L. Schöls, "Autosomal Recessive Cerebellar Ataxias: Paving the Way Toward Targeted Molecular Therapies," *Neuron* 101, no. 4 (2019): 560–583.

9. M. Synofzik and R. Schüle, "Overcoming the Divide Between Ataxias and Spastic Paraplegias: Shared Phenotypes, Genes, and Pathways," *Movement Disorders* 32, no. 3 (2017): 332–345.

10. T. Klockgether, T. Ashizawa, B. Brais, et al., "Paving the Way Toward Meaningful Trials in Ataxias: An Ataxia Global Initiative Perspective," *Movement Disorders* 37, no. 6 (2022): 1125–1130.

11. T. Schmitz-Hübsch, S. T. du Montcel, L. Baliko, et al., "Scale for the Assessment and Rating of Ataxia: Development of a New Clinical Scale," *Neurology* 66, no. 11 (2006): 1717–1720.

12. A. Hamdan, A. C. Hooker, X. Chen, et al., "Item Performance of the Scale for the Assessment and Rating of Ataxia in Rare and Ultra-Rare Genetic Ataxias," *CPT: Pharmacometrics & Systems Pharmacology* 13, no. 8 (2024): 1327–1340.

13. N. Hendrickx, F. Mentré, A. Traschütz, et al., "Prediction of Individual Disease Progression Including Parameter Uncertainty in Rare Neurodegenerative Diseases: The Example of Autosomal-Recessive Spastic Ataxia Charlevoix Saguenay (ARSACS)," *AAPS Journal* 26, no. 3 (2024): 57.

14. W. Ilg, S. Milne, T. Schmitz-Hübsch, et al., "Quantitative Gait and Balance Outcomes for Ataxia Trials: Consensus Recommendations by the Ataxia Global Initiative Working Group on Digital-Motor Biomarkers," *Cerebellum* 23, no. 4 (2024): 1566–1592.

15. W. Ilg, Z. Fleszar, C. Schatton, et al., "Individual Changes in Preclinical Spinocerebellar Ataxia Identified via Increased Motor Complexity," *Movement Disorders* 31, no. 12 (2016): 1891–1900.

16. W. Ilg, B. Müller, J. Faber, et al., "Digital Gait Biomarkers Allow to Capture 1-Year Longitudinal Change in Spinocerebellar Ataxia Type 3," *Movement Disorders* 37, no. 11 (2022): 2295–2301.

17. L. Rochester, B. Galna, S. Lord, D. Mhiripiri, G. Eglon, and P. F. Chinnery, "Gait Impairment Precedes Clinical Symptoms in Spinocerebellar Ataxia Type 6," *Movement Disorders* 29, no. 2 (2014): 252–255.

18. "Comprehensive Gait and Balance Analysis – APDM Wearable Technologies [Internet]," (2020), APDM [cited 2024 Mar 13], https://apdm.com/mobility/.

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19. R. Reilmann and R. Schubert, "Chapter 18 – Motor Outcome Measures in Huntington Disease Clinical Trials," in *Handbook of Clinical Neurology [Internet]*, ed. A. S. Feigin and K. E. Anderson (Elsevier, 2017), 209–225, [cited 2024 Mar 14] (Huntington Disease; vol. 144), https://www.sciencedirect.com/science/article/pii/B9780128018934000183.

20. D. Hermle, R. Schubert, P. Barallon, et al., "Multifeature Quantitative Motor Assessment of Upper Limb Ataxia Including Drawing and Reaching," *Annals of Clinical and Translational Neurology* 11, no. 5 (2024): 1097–1109.

21. R. Reilmann and R. Schubert, "Motor Outcome Measures in Huntington Disease Clinical Trials," *Handbook of Clinical Neurology* 144 (2017): 209–225.

22. R. Reilmann, A. McGarry, I. D. Grachev, et al., "Safety and Efficacy of Pridopidine in Patients With Huntington's Disease (PRIDE-HD): A Phase 2, Randomised, Placebo-Controlled, Multicentre, Dose-Ranging Study," *Lancet Neurology* 18, no. 2 (2019): 165–176.

23. L. Servais, K. Yen, M. Guridi, J. Lukawy, D. Vissière, and P. Strijbos, "Stride Velocity 95th Centile: Insights Into Gaining Regulatory Qualification of the First Wearable-Derived Digital Endpoint for Use in Duchenne Muscular Dystrophy Trials," *Journal of Neuromuscular Disease* 9, no. 2 (2022): 335–346.

24. L. Servais, D. Eggenspieler, M. Poleur, et al., "First Regulatory Qualification of a Digital Primary Endpoint to Measure Treatment Efficacy in DMD," *Nature Medicine* 29, no. 10 (2023): 2391–2392.

25. PROSPAX [Internet], "PROSPAX – An Integrated Multimodal Progression Chart in Spastic Ataxias," [cited 2024 Mar 13], https://www.prospax.net/.

26. L. Beichert, W. Ilg, C. Kessler, et al., "Digital Gait Outcomes for Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay (ARSACS): Discriminative, Convergent, and Ecological Validity in a Multicenter Study (PROSPAX)," *Movement Disorders* 39, no. 9 (2024): 1544–1555.

27. L. Beichert, J. Seemann, C. Kessler, et al., "Patient-Relevant Digital-Motor Outcomes for Clinical Trials in Hereditary Spastic Paraplegia Type 7," *Neurology* 103, no. 12 (2024): e209887.

28. A. Traschütz, S. Reich, A. D. Adarmes, et al., "The ARCA Registry: A Collaborative Global Platform for Advancing Trial Readiness in Autosomal Recessive Cerebellar Ataxias," *Frontiers in Neurology* 12 (2021): 677551.

29. A. Traschütz, A. D. Adarmes-Gómez, M. Anheim, et al., "Responsiveness of the Scale for the Assessment and Rating of Ataxia and Natural History in 884 Recessive and Early Onset Ataxia Patients," *Annals of Neurology* 94, no. 3 (2023): 470–485.

30. V. V. Shah, R. Rodriguez-Labrada, F. B. Horak, et al., "Gait Variability in Spinocerebellar Ataxia Assessed Using Wearable Inertial Sensors," *Movement Disorders* 36, no. 12 (2021): 2922–2931.

31. W. Ilg, J. Seemann, M. Giese, et al., "Real-Life Gait Assessment in Degenerative Cerebellar Ataxia," *Neurology* 95, no. 9 (2020): e1199–e1210.

32. M. Delattre, M. Lavielle, and M. A. Poursat, "A Note on BIC in Mixed-Effects Models," *Electronic Journal of Statistics* 8, no. 1 (2014): 456–475.

33. S. Ueckert, M. O. Karlsson, and A. C. Hooker, "Accelerating Monte Carlo Power Studies Through Parametric Power Estimation," *Journal of Pharmacokinetics and Pharmacodynamics* 43, no. 2 (2016): 223–234.

34. S. L. Beal, L. B. Sheiner, A. Boeckmann, and R. J. Bauer, *NONMEM User's Guides (1989–2009)* (Icon Development Solutions, 2009).

35. M. O. Karlsson and R. Nordgren, "Perl-Speaks-NONMEM: User Documentation [Internet]," [cited 2025 Feb 4], https://uupharmacometrics.github.io/PsN/docs.html.

36. N. Hendrickx, F. Mentré, A. Hamdan, et al., "Comparing Randomized Trial Designs to Estimate Treatment Effect in Rare Diseases With Longitudinal Models: A Simulation Study Showcased by Autosomal Recessive Cerebellar Ataxias Using the SARA Score [Internet]," *medRxiv* (2025) [cited 2025 Jun 13], p. 2025.01.30.25321311, https://www.medrxiv.org/content/10.1101/2025.01.30.25321311v1.

37. M. Synofzik, W. M. C. van Roon-Mom, G. Marckmann, et al., "Preparing n-of-1 Antisense Oligonucleotide Treatments for Rare Neurological Diseases in Europe: Genetic, Regulatory, and Ethical Perspectives," *Nucleic Acid Therapeutics* 32, no. 2 (2022): 83–94.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

21638306, 2025, 11, Downloaded from https://ascpt.onlinelibrary.wiley.com/doi/10.1002/psp4.70081 by Koc University, Wiley Online Library on [11/12/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License