

Passive Smartphone Gait Metrics in Cervical Spondylotic Myelopathy

Pranay Singh

Washington University in St. Louis School of Medicine

Faraz Arkam

Washington University in St. Louis School of Medicine

Abigail Lewis

Washington University in St. Louis School of Medicine

Erdong Chen

Washington University in St. Louis School of Medicine

Salim Yakdan

Washington University in St. Louis School of Medicine

Jing Wang

Washington University in St. Louis School of Medicine

Jang W. Yoon

University of California San Francisco School of Medicine

Shay Bess

University of Arizona College of Medicine

Thomas Mroz

Cleveland Clinic Lerner College of Medicine

Zoher Ghogawala

UMass Chan School of Medicine

Michael P. Steinmetz

Cleveland Clinic Lerner College of Medicine

Wilson Z. Ray

Washington University in St. Louis School of Medicine

Jacob K. Greenberg

jacobgreenberg@wustl.edu

Washington University in St. Louis School of Medicine

Article

Keywords: cervical spondylotic myelopathy, mobile health, digital biomarkers, passive monitoring

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10963689/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: Competing interest reported. S.B. reports royalties and patent-related interests involving VB Spine and Globus; research support from VB Spine, Medtronic, DePuy Synthes, and Globus; and educational grant support from Medtronic, Globus, VB Spine, Stryker, Orthofix, and ATEC. The remaining authors declare no financial or non-financial competing interests.

Abstract

Cervical spondylotic myelopathy (CSM) is the leading cause of spinal cord dysfunction in older adults, yet progression is monitored with episodic and subjective scales that detect decline only after symptoms emerge. We retrospectively analyzed five gait metrics (walking speed, step length, double support, asymmetry, steadiness) accrued passively up to five years by the personal iPhones of 56 patients with CSM and 42 controls. Each metric was summarized as the median and worst-5% of daily values. All five separated patients from controls (worst-5% AUC 0.74-0.85) and tracked modified Japanese Orthopaedic Association severity, with worst-5% summaries outperforming medians. On a diagnosis-anchored timeline, patient and control trajectories diverged 35-54 months before diagnosis, and a within-person detection rule flagged decline a median of 15-26 months before diagnosis. Passively collected consumer-grade gait data carry a functional signal that precedes clinical diagnosis in CSM and may enable scalable monitoring in other disorders of gradual motor decline.

Introduction

Cervical spondylotic myelopathy (CSM) is the most common cause of spinal cord dysfunction in older adults.¹ It results from chronic, age-related compression of the cervical spinal cord and affects approximately 2% of adults.^{1,2} Its prodrome, asymptomatic spinal cord compression, is more common still, present on imaging in roughly one-third of adults over 60, up to one-quarter of whom develop symptomatic myelopathy.^{3,4} CSM produces progressive neurological impairment, most commonly gait dysfunction and loss of hand dexterity, along with declines in physical function and quality of life.⁵ Surgery is the only treatment with high-level evidence for halting progression, and outcomes are more favorable when performed before irreversible neurological injury occurs.⁵ Asymptomatic spinal cord compression and some cases of mild disease are therefore typically managed by observation, with documented progression often serving as the principal trigger for surgical intervention.^{6,7} Effective management under this treatment paradigm therefore depends on detecting such disease progression early and reliably.

In routine practice, progression is monitored through patient report and clinician evaluation, typically once or twice per year.⁸ The most widely used disease-specific instrument is the modified Japanese Orthopaedic Association (mJOA) score, a clinician-administered ordinal scale that is largely based on patient report and does not incorporate objective measures of neurological ability.⁹ Therefore, by definition the mJOA generally only detects neurological progression once it has led to a symptomatic decline. Consequently, developing objective tools capable of detecting early or subclinical neurological worsening has been identified as a priority research goal by the AO Spine RECODE-DCM initiative.¹⁰ Further support for such efforts has come from a recent qualitative study of patients with CSM, spine surgeons, and therapists, which found current tools inadequate for individualized care, particularly citing an inability to capture daily function outside the clinic.¹¹

Smartphone and wearable digital biomarkers offer a means of addressing this gap. Sensor-derived measures can be collected repeatedly and remotely with minimal burden and have shown discriminative and monitoring value in Parkinson's disease, multiple sclerosis, stroke, and cognitive decline.^{12,13} Recently, we developed SynapTrack, a smartphone platform for active assessment of gait, dexterity, and vibratory sensation in CSM, and demonstrated preliminary reliability and validity against reference clinical assessments.¹⁴⁻¹⁶ Like many digital assessments in neurological conditions, SynapTrack relies on active testing, in which the patient performs a prescribed task. Active testing is controllable and clinically interpretable, but it requires sustained adherence and may not capture the full spectrum of disease fluctuations in everyday life.^{17,18} On the other hand, passive monitoring, for example, continuous

gait measurements, offers potential advantages that can complement active testing.¹⁹ Data accrue from a device the patient already owns, removing hardware or adherence barriers that constrain scalable remote monitoring.²⁰

One of the most widely available sources of such data is Apple Health, the activity-tracking framework preinstalled on every iPhone. Since iOS 14 (2020), a compatible iPhone carried during ordinary walking automatically estimates a panel of gait metrics without any dedicated task or accessory,²¹ with walking steadiness added in iOS 15 (2021) (Apple Support, *Monitor your walking steadiness in Health on iPhone*, <https://support.apple.com/guide/iphone/monitor-your-walking-steadiness-iphff1bf03ed/ios>, accessed August 2026). Because these metrics accrue by default, a patient who has carried an iPhone may hold several years of continuous gait history reaching back to the metric's release, a longitudinal window that prospective studies cannot recover. Data on their application in neurological disease, however, remain sparse.^{22,23}

We therefore sought to evaluate the putative utility of passively collected, consumer-grade Apple Health gait metrics as a digital biomarker of gait dysfunction in CSM. Specifically, we examined whether these metrics distinguish patients with CSM from healthy controls, scale with mJOA-defined severity, and follow distinct longitudinal trajectories, including whether they begin to change before diagnosis. Because gait varies appreciably from day to day and impairment may surface first on worse-performance days, each metric was summarized both as the median and as the worst 5th percentile. Because these data are generated passively at no cost or burden, this approach could extend CSM monitoring beyond the clinic and may generalize to other neurological conditions of gradual motor decline.

Methods

Study design and participants

This single-center cohort study was conducted at a tertiary academic medical center and comprised a retrospective analysis of passively collected Apple Health gait data in patients with CSM and disease-free controls. Eligible participants were English-speaking adults aged 18 years or older. Patients with CSM required a diagnosis confirmed by a fellowship-trained spine surgeon on the basis of clinical presentation and cervical imaging demonstrating spinal cord compression; controls had no history of cervical myelopathy or any other condition affecting lower-extremity function. Participants were excluded for major comorbid neurological disorders that could confound neurological assessment (e.g., Parkinson's disease, stroke, traumatic brain injury), severe cognitive impairment limiting adherence to study instructions, musculoskeletal or physical limitations preventing completion of testing, substantial visual, auditory, or motor impairments unrelated to CSM that could interfere with app use, and non-degenerative causes of myelopathy (e.g., tumor, infection, inflammatory disease). The study was approved by the Washington University in St. Louis Institutional Review Board (IRB #202401026), and written informed consent was obtained from all participants.

Data acquisition and variables

Participants downloaded SynapTrack and used it to upload their Apple Health data,¹⁵ providing continuous gait metrics that personal devices had recorded passively over up to five years prior to enrollment. The mJOA score, obtained at enrollment, served as the clinical reference standard for myelopathy severity. It is the most widely used disease-specific severity measure in CSM with validated severity strata of mild (15-17), moderate (12-14), and severe (0-11).^{24,25} Clinical diagnosis and surgery dates were obtained from medical records.

Five gait metrics were analyzed: walking speed, step length, double support, walking asymmetry, and walking steadiness. All five are produced by Apple algorithms whose implementations are not fully disclosed. Walking speed, step length, and double support are conventional spatiotemporal gait parameters with standard biomechanical definitions.²² Gait asymmetry is likewise a well-established construct, but has no universally accepted operationalization.²⁶ Apple's implementation reports the percentage of time within a walking bout during which steps are detected as asymmetric.²¹ By contrast, walking steadiness is a proprietary machine-learning composite score expressed as a percentage, generated approximately weekly rather than per walking bout, and it has no equivalent construct in conventional gait analysis. For patients, all analyses were based on preoperative data; control data were used in full.

Apple Health returns multiple observations per metric per day. These were first averaged to one value per metric per day. For each participant and metric, the distribution of daily values was then summarized as the median of daily values, and as the worst-5% of daily values (the 5th percentile for metrics in which higher values are favorable, and the 95th percentile for metrics in which higher values are unfavorable).

Statistical analysis

All hypothesis tests were two-sided with $\alpha = 0.05$. Because gait summaries were skewed and bounded, nonparametric tests were used throughout. Benjamini-Hochberg false discovery rate (FDR) adjustment was applied to cross-sectional group comparisons, mJOA correlations, ordered severity trends, time-by-group interactions, and comparisons of decline rules.²⁷ Unless otherwise stated, p values reported in the text, tables, and figures are FDR-adjusted with $p < 0.05$ considered significant.

Cross-sectional discrimination and clinical associations

We used cross-sectional comparisons to evaluate how well passively collected Apple Health metrics discriminate CSM from controls and how strongly they track myelopathy severity. Continuous demographic variables were compared between groups using two-sample t-tests, and categorical variables using the chi-square test or Fisher's exact test, as appropriate. For cross-sectional discrimination, CSM and control distributions of both the median and worst-5% summaries were compared for each metric using the Wilcoxon rank-sum test. Discriminative performance was quantified by the area under the receiver-operating-characteristic curve (AUC) with 95% DeLong confidence intervals (CIs).²⁸ Because baseline characteristics differed between groups, a sensitivity analysis was performed using linear regression for each gait summary, with group, age, and sex as covariates (Supplementary Table 1). Race and educational attainment were excluded because sparse cell counts produced unstable estimates.

Within the CSM group, associations between each gait metric and the mJOA score were assessed by Spearman rank correlation. Dose-response across controls and mJOA-defined severity strata (mild, moderate, severe) was tested by rank-based correlation between each metric and the ordinal severity level.²⁴

Longitudinal trajectory modeling

These analyses evaluated whether the rate of gait change over the monitoring period differed between patients and controls. Gait was modeled as a function of time since each participant's first available observation. To reduce autocorrelation and uneven sampling density, daily values were aggregated into per-participant monthly summaries. Weekly rather than monthly grouping was evaluated as a sensitivity analysis (Supplementary Figure 1). Within each bin, both the median and the worst-5% were computed to yield a per-participant trajectory of gait. Linear mixed-effects models with participant-level random effects were fit by maximum likelihood. Because passively collected

gait data are irregularly sampled and serially correlated, no single covariance structure was assumed. For each metric, different functional forms of the trajectory (linear, quadratic, and natural spline) and covariance structures were tested, with the Akaike Information Criterion (AIC) used to guide model selection.²⁹ Full selection landscapes are reported as supplementary figures. The time-by-group interaction reflected whether the rate of gait change differed between groups.

Timing of group-level divergence relative to clinical diagnosis

To estimate how long before a CSM diagnosis patients' gait first deviated from controls at the group level, CSM participants were aligned to a common timeline anchored at clinical diagnosis; controls were anchored to their most recent observation as a contemporaneous within-person reference. Group trajectories were estimated using the same model-selection procedure described above, applied separately to each metric within the pre-diagnosis window. To identify when trajectories diverged, the estimated patient-control difference was evaluated at every month of the disease-time grid. Testing each month separately would inflate the false-positive rate across the ~60 months examined, so divergence was instead based on a simultaneous (sup-t) confidence band, which is constructed so that the entire difference curve falls inside the band with 95% probability rather than each month individually.³⁰ The difference curve and its covariance across the monthly grid were obtained from the selected mixed model by the delta method, and the critical value for the band was obtained by simulation (20,000 draws from the estimated multivariate normal distribution of the difference curve).³¹ Divergence was defined as the earliest disease-time month at which this band excluded zero for at least two consecutive months, among months contributed to by at least 10 CSM participants. The ≥ 10 participant requirement was used only as the minimum sample to declare divergence, while models were fit using all available participants. As a complementary measure of separation over time, discrimination (patient versus control AUC) was estimated within successive pre-diagnosis intervals, with bootstrap confidence intervals.

Individual-level early detection

To assess whether deterioration was detectable in individual patients before clinical diagnosis, a within-person change-detection rule was applied in which each participant served as their own reference. Each participant's personal baseline was the mean and standard deviation of their monthly values over the first 6 months of monitoring. Because walking steadiness became available only on later operating-system versions, its baseline was taken as each participant's first 6 available monthly values. Month-to-month reliability of each metric was characterized using these same six baseline monthly values, which precede diagnosis and were verified to show no systematic drift across the baseline window (all $p \geq 0.168$). A two-way random-effects intraclass correlation coefficient for single measurements, ICC(2,1), was computed for each metric across patients, with 95% confidence intervals obtained by patient-level bootstrap. ICCs were interpreted as poor (<0.50), moderate (0.50-0.75), good (0.75-0.90), or excellent (>0.90).³²

Because there is not an absolute standard for defining a clinically meaningful gait decline in CSM, two individualized thresholds were evaluated: a fixed 1.5-standard-deviation (SD) change from baseline and an individualized 90% change threshold based on the minimal detectable change (MDC90) framework. The MDC90 was defined as the smallest within-person change that can be distinguished from measurement error with 90% confidence. It was calculated as $1.645 \times \sqrt{(1 + 1/k) \times \text{the standard error of measurement (SEM)}}$, where $k = 6$ baseline monthly observations, yielding $1.777 \times \text{SEM}$.³³ Rather than deriving the SEM from a population-level reliability coefficient ($\text{SD} \times \sqrt{(1 - \text{ICC})}$), as in a conventional test-retest MDC, we defined it as each participant's within-person SD across six baseline monthly values. This approach incorporates both measurement error and natural within-person biological

variability and is conceptually closer to an individualized reference-change threshold. It was chosen because a personalized monitoring trigger should reflect an individual's own baseline variability rather than population-level measurement error. The ICC is reported to characterize metric stability but does not enter the threshold calculation. A participant was flagged at the first worsening relative to baseline that persisted for at least two consecutive months.

Lead time, defined as the interval by which the first sustained flag preceded clinical diagnosis, was summarized per metric as the median and interquartile range (IQR). Time to first flag was estimated by Kaplan-Meier methods and is reported as the cumulative probability of a first flag, with participants whose monitoring ended before any flag treated as censored.³⁴ Because both decline rules were applied to the same patients, the rules were compared using a marginal Cox proportional-hazards model with a cluster-robust (sandwich) variance clustered on participant, which estimates the marginal hazard ratio while correcting for within-person correlation. The proportional hazards assumption was assessed by testing the scaled Schoenfeld residuals against time.³⁵

As an alternative for participants without a full 6-month baseline, the same analysis was repeated using weekly rather than monthly bins, with a 13-week baseline and a trigger of 4 abnormal weeks out of any 6 consecutive weeks. All analyses were performed in R 4.4.1.

Results

Cohort characteristics

The analysis cohort comprised 42 controls and 56 patients with CSM (Table 1). Controls and patients did not differ significantly in age (57.0 ± 11.6 vs 61.3 ± 11.4 years; mean $\pm$ SD; $p = 0.068$) or in the proportion of female participants (50% vs 57%; $p = 0.48$). Most participants were White (83% of controls, 84% of patients) and non-Hispanic (93% and 100%). A greater proportion of controls held a graduate degree (52% vs 27%), whereas college education was similar between groups (45% vs 48%). Among patients with CSM, the mean Neck Disability Index (NDI) was 29.0 ± 9.5 , and the mean mJOA score was 13.5 ± 2.6 (Table 1). Passive gait history spanned a median of 36.0 months in both controls and patients with CSM ($p = 0.50$), with a median pre-diagnosis span of 34.5 months among patients with CSM.

Table 1
Demographic and clinical characteristics by group. Percentages are based on available data for each variable; denominators vary because of missing responses.

Characteristic	Control	Myelopathy	Statistic (p)
n	42	56	
Age, years, mean (SD)	57.0 (11.6)	61.3 (11.4)	p = 0.068
Female, n (%)	21 (50)	32 (57)	p = 0.48
Ethnicity, n (%)			p = 0.091
Hispanic or Latino	3 (7)	0 (0)	
Not Hispanic or Latino	38 (93)	49 (100)	
Race, n (%)			p = 0.002
White	35 (83)	43 (84)	
African American/Black	0 (0)	7 (14)	
Asian	3 (7)	1 (2)	
Other	4 (10)	0 (0)	
Education, n (%)			p = 0.003
Less than high school degree	0 (0)	1 (2)	
High school degree	1 (2)	12 (23)	
College degree	19 (45)	25 (48)	
Graduate degree	22 (52)	14 (27)	
mJOA total, mean (SD)	–	13.5 (2.6)	
NDI total, mean (SD)	–	29.0 (9.5)	
Observation window			
Total span, months, median [IQR]	36.0 [20.6–42.5]	36.0 [31.2–55.2]	p = 0.5
Pre-diagnosis span, months, median [IQR]	–	34.5 [25.5–52.0]	

Fisher's exact test was used for ethnicity, race, and education; chi-square test was used for sex; two sample t-test was used for age. CSM = Cervical Spondylotic Myelopathy; SD = standard deviation; IQR = interquartile range; mJOA = modified Japanese Orthopaedic Association; NDI = Neck Disability Index.

Cross-sectional discrimination of CSM from controls

Both median and worst-5% summaries separated CSM patients from controls (Fig. 1). At the participant level, patients showed worse direction-corrected values than controls for every metric under both summaries, and all differed significantly after FDR correction (all adjusted $p \leq 0.002$; Fig. 1a). Discrimination, quantified as the AUC, was better for the worst-5% than for the median in four of the five metrics (Fig. 1b). Worst-5% AUCs ranged from 0.74 to 0.85 depending on the metric (Fig. 1b). The largest tail-associated gain occurred for walking asymmetry, which rose from an AUC of 0.70 for the median to 0.74 for the worst-5% (Fig. 1b). In a sensitivity analysis adjusting for age and

sex, between-group differences in every metric retained their direction and magnitude and remained significant after FDR correction (all adjusted $p \leq 0.024$; Supplementary Table 1).

Association with clinical severity

Among patients with CSM, the worst-5% summary showed stronger correlations with mJOA scores than the median summary for all five gait metrics. Correlations were strongest for walking speed (worst-5% $\rho = 0.54$, adjusted $p = 0.002$) and weakest for walking steadiness, with the remaining metrics showing intermediate associations (Fig. 2a). Before adjustment, four worst-5% correlations and two median correlations were statistically significant. After FDR adjustment, walking speed and walking asymmetry remained significant under both summaries (worst-5% adjusted $p = 0.002$ and 0.015), whereas step length and double support did not (both adjusted $p = 0.066$). Across groups stratified by mJOA severity, all five metrics showed significant ordered trends with increasing severity (adjusted $p < 0.001$; Fig. 2b).

Longitudinal trajectories

The rate of change over the monitoring period differed between groups for most gait metrics, and the time-by-group interaction was stronger for the worst-5% than the median in four of the five metrics (Fig. 3). After FDR correction under the worst-5% summary, the interaction was significant for four of the five metrics (adjusted $p \leq 0.040$) versus three under the median; step length was significant only under the worst-5% (adjusted $p = 0.040$) and walking steadiness under neither (Fig. 3). Double support showed the strongest interaction overall ($F = 18.0$ for the worst-5% vs 8.4 for the median), whereas step length showed the greatest proportional increase in the interaction statistic ($F = 5.4$ vs 1.6).

The model-selection landscapes underlying these fits are reported as supplements for the weekly-thinned, median-summarized data (Supplementary Fig. 1) and for the monthly-thinned data computed for both the median and worst-5% tail (Supplementary Fig. 2).

Timing of group-level divergence relative to clinical diagnosis

Whereas the preceding analysis modeled time from each participant's first observation and asked whether the rate of change differed between groups, this analysis realigned patients to their diagnosis date and asked when the two groups first separated. When patients were aligned to clinical diagnosis (month 0), patient and control group trajectories diverged well before diagnosis across all five metrics (Fig. 4). The earliest disease-time month at which the simultaneous (sup-t) confidence band of the patient-control difference excluded zero for at least two consecutive months was -54 months for double support, step length, and walking speed, -50 months for walking asymmetry, and -35 months for walking steadiness (Fig. 4). The AIC landscape for these disease-clock trajectory models is reported in Supplementary Fig. 3.

In a complementary analysis of discrimination across six pre-diagnosis disease-time bins (midpoints -54 , -42 , -30 , -18 , -9 , and -3 months), patient-versus-control AUC rose as diagnosis approached for all five primary metrics (Supplementary Fig. 4). By roughly 30 months before diagnosis, discrimination exceeded the chance level of 0.5 for every metric (AUC 0.63 – 0.79), and increased to 0.72 – 0.86 in the bin nearest diagnosis. In the earliest bin (54 months before diagnosis), where fewer patients contributed, double support, step length, and walking speed were already above chance (0.72 , 0.73 , and 0.68).

Individual-level early detection

Month-to-month reliability across the six baseline monthly values was excellent for double support (ICC 0.97, 95% CI 0.94–0.98), step length (0.92, 0.87–0.95), and walking speed (0.90, 0.83–0.93) and good for walking steadiness (0.89, 0.83–0.93). Walking asymmetry had moderate reliability (0.65, 0.45–0.75).

Within-person change detection flagged sustained gait decline a median of 15 to 26 months before clinical diagnosis, depending on the metric (Fig. 5a). For the fixed 1.5-SD rule, the median lead time and number of patients flagged out of those eligible were double support 26 months (24/41), walking asymmetry 25 months (27/39), step length 23 months (25/40), walking speed 22 months (26/40), and walking steadiness 15 months (20/33) (Fig. 5a).

On the cumulative-probability curves, the 1.5-SD rule reached an equal or higher cumulative probability of first flag than the individualized 90% change threshold (MDC90) for every metric, ranging from 65% (double support) to 78% (walking speed) under the 1.5-SD rule (Fig. 5b). In cluster-robust marginal Cox models, the two thresholds produced closely similar flag timing (Fig. 5b). Only step length reached significance (unadjusted $p = 0.019$), and no metric remained significant after FDR adjustment (all adjusted $p \geq 0.076$). The proportional hazards assumption showed no evidence of violation for any metric (all $p \geq 0.724$), so a single hazard ratio adequately summarizes each comparison. Walking steadiness is not shown because it is only on newer operating-system versions, so its curve is not timing-comparable on the shared axis.

The weekly alternative, defined as four of six consecutive weeks beyond the threshold, produced similar results (Supplementary Fig. 5). Under the 1.5-SD rule, median lead times ranged from 18 to 27 months, and the cumulative probability of first flag ranged from 70% (walking asymmetry) to 81% (step length).

Discussion

In this study, passively collected consumer-grade gait data showed substantial promise and multidimensional capabilities in serving as a digital biomarker of gait dysfunction in CSM. To our knowledge, this study is the first characterization of gait in a neurological condition from retrospective, passively collected consumer health-platform data. iPhone-derived gait metrics distinguished patients from controls, scaled with mJOA-defined severity, and followed group-specific longitudinal trajectories. Cross-sectional discrimination was acceptable to good, led by walking steadiness and double support, with four of five metrics at or above an AUC of 0.75. Likewise, passive gait metrics distinguished CSM profiles from those of controls at the group and individual level years before a clinical diagnosis was made. Collectively, these findings highlight the overall promise of passive gait digital monitoring in neurological disease and specifically the capabilities of leveraging readily-available consumer trackers.

A key finding was that the worst-5% of gait performance consistently outperformed the median across analyses, revealing a distributional signal in CSM that has not previously been recognized. Longitudinal gait monitoring has conventionally relied on central-tendency summaries, yet gait varies appreciably from day to day.^{13,36} Features beyond the mean or median, such as stride-to-stride variability, distributional percentiles of bout-level walking speed, and activity fragmentation, have added discriminative or predictive value for functional or neurological decline in related disorders.^{37–39} However, to our knowledge, the relative value of lower-tail versus central-tendency gait summaries has not previously been evaluated in CSM.^{40,41} In this study, the benefit was largest for double support longitudinally, where the time-by-group interaction more than doubled. This directionality aligns with a growing body of research. In Parkinson's disease, distributional measures of gait outperformed the average for distinguishing motor state, grading severity, and tracking progression.^{42,43} The clearest regulatory precedent is the stride velocity 95th centile in Duchenne muscular dystrophy, which the European Medicines Agency qualified as a primary trial endpoint in 2023, outperforming median stride velocity in sensitivity.⁴⁴ Our findings identify the lower tail of gait

performance as a previously unrecognized source of disease signal in CSM, suggesting that the most impaired episodes of everyday walking may be more informative than a patient's typical performance. Importantly, this signal can be captured passively using ubiquitous consumer hardware. Because this approach requires only a distribution of repeated measurements, it is broadly applicable to conditions in which continuous passive data are available.

The signal residing in gait quality fits the pathophysiology of CSM, which impairs the control of movement, such as coordination, proprioception, and balance. Spatiotemporal gait-quality measures have similarly been used to detect early disease progression in Parkinson's disease, cerebellar ataxia, and neuromuscular disorders.^{45–47} The metrics examined here index the spatiotemporal quality of gait, and their ordinal association with mJOA severity supports their ability to track clinical myelopathy impairment.

Arguably, the most impactful potential application of passive Apple Health data relates to earlier clinical recognition of CSM. At the group level on a common disease timeline anchored at clinical diagnosis, patient and control trajectories separated roughly 35 to 54 months before diagnosis across the five metrics with further separation of the trajectories thereafter. For walking speed, step length, and double support, the – 54 month divergence coincided with the earliest month at which the pre-specified minimum of 10 contributing patients was met. These estimates are therefore bounded by data availability, and true divergence may predate the observation window. Because clinical diagnosis is influenced by symptom recognition, healthcare-seeking behavior, and referral patterns, these findings should be interpreted as a gait signal that precedes clinical diagnosis rather than evidence of preclinical CSM. Nonetheless, the several-year interval over which gait trajectories differed exceeded the reported 1–2 year delay between symptom onset and diagnosis, suggesting that passive gait monitoring may capture functional changes before CSM comes to clinical attention.⁴⁸ If confirmed prospectively, these findings could indicate a prolonged, previously undercharacterized phase of functional change preceding CSM diagnosis. Similar temporal patterns have been reported in other neurodegenerative diseases. In initially healthy older adults monitored using a lower-back inertial sensor, gait pace deviated from that of non-converters roughly four years before a clinical diagnosis of Parkinson's disease.⁴⁹ Similarly, machine-learning models trained on passive UK Biobank wrist accelerometry distinguished individuals who later developed Parkinson's disease from healthy controls up to seven years before clinical diagnosis.⁵⁰ Prospective studies in individuals with asymptomatic spinal cord compression could determine whether passive gait trajectories characterize this transition and identify patients before clinically apparent CSM.

Distinct from this group-level signal, the within-person change-detection rule flagged sustained decline a median of 15 to 26 months before diagnosis using each patient as their own baseline. Because change is referenced to an individual's own prior gait rather than to a control distribution, a sustained deviation is meaningful for that person independent of where they fall in the population. Individual-level lead times were shorter than group-level ones because flagging sustained change within a single record requires overcoming within-person biological and analytical variability as opposed to detecting a difference between groups.^{51,52} Furthermore, the within-person declines are notable because they occurred within populations that had generally not yet presented for clinical evaluation. More rapid declines may be expected in individuals with established CSM, in whom 20–62% deteriorate neurologically within 3–6 years without surgery, a substantially higher rate than the progression to myelopathy observed among individuals with asymptomatic cord compression.⁵³ Thus, within-person change could in principle serve as a personalized trigger to seek clinical evaluation. However, these rules were applied and evaluated entirely within this cohort; consequently, these findings should be considered proof of concept and require prospective validation to determine clinical implications and prognostic importance.

Our results carry particular weight when considering the natural history of CSM, which is heterogeneous and poorly predictable at the individual level.⁵³ CSM phenotypes include a stepwise pattern of decline, slow steady deterioration, and rapidly progressive functional decline.^{54,55} Importantly, no clinical or radiographic feature reliably identifies who will progress.⁵³ The rate of decline in these phenotypes is itself poorly characterized. In our data, patients declined substantially faster than controls in four of the five metrics under the worst-5% summary. The separation was consistently larger in the lower tail, most strikingly for double support ($F = 8.4$ for the median vs 18.0 for the worst-5%), and for step length, where a rate difference was detectable only in the tail. Passive monitoring may therefore offer a means of quantifying progression rate at the resolution the phenotypic heterogeneity of CSM demands.

Our detection results also appear to reflect CSM heterogeneity. Under the more sensitive 1.5-SD rule, the cumulative probability of a first sustained flag plateaued between roughly 65% (double support) and 78% (walking speed), leaving a consistent minority who never flagged. There are several potential explanations for these unflagged patients. First, while gait is a critical feature of CSM, patients may have declined in other domains (e.g., hand dexterity) rather than gait. Second, some patients may have experienced a much more drawn out and slower decline not fully captured within the monitoring window. A less stringent progression threshold may have revealed such cases. Finally, our definitions focus exclusively on objective changes, which may not fully correlate with subjective disease impressions. Nonetheless, these results suggest that routinely collected consumer data may feasibly separate disease progression, a question that may be defined more comprehensively in larger, prospective cohorts.

A related question is whether these detection results depend on the specific analytic choices made. Although the individualized 90% change threshold is more conservative than the 1.5-SD rule, once the variance factor is corrected for a six-month averaged baseline the two are numerically close (1.78 versus 1.50 baseline SD) and produced nearly identical results. Substituting a 13-week baseline and a trigger of four abnormal weeks in any six consecutive weeks likewise produced comparable lead times and detection rates. Taken together, the detection findings appear robust to both the threshold and the temporal resolution chosen, and the approach does not require a full six-month baseline to be available.

Reliability differed by metric. Month-to-month ICCs among patients were good to excellent for double support, step length, walking speed, and walking steadiness (0.89–0.97), indicating that repeated passive measurements of these metrics are stable enough to support individual-level monitoring. Walking asymmetry was a clear exception (ICC 0.65), consistent with instrumented gait studies in which symmetry measures have shown considerably poorer test–retest reliability than pace-related measures such as walking speed and step length.^{56,57} Walking asymmetry may therefore be a less optimal metric for tracking disease severity within an individual, even though it discriminated patients from controls at the group level.

This study has several limitations. The study was retrospective and single-center, of modest size (56 patients, 42 controls), and the discrimination, severity, and longitudinal estimates were derived without external or hold-out validation. While a sensitivity analysis adjusting for age and sex did not change group separation, groups differed in race and educational attainment. By design, the cohort comprised iPhone owners, who tend to be younger and have a higher income.⁵⁸ Nonetheless, approximately 91% of US adults own a smartphone (Pew Research Center, <https://www.pewresearch.org/internet/fact-sheet/mobile/>, 2025), and iPhones account for roughly 59% of the US smartphone market (StatCounter Global Stats, <https://gs.statcounter.com/os-market-share/mobile/united-states-of-america/>, 2026). Furthermore, although these metrics have been described at a conceptual level, they were generated by proprietary algorithms, have not been specifically validated in CSM or other neurodegenerative disease populations, and may be impacted by silent algorithm updates across operating-system versions.⁵⁹ Independent peer-reviewed validation against a reference inertial sensor system exists only for walking speed, step length, and

double support, and only in healthy cohorts.²² Individual-metric findings will therefore require confirmation against a reference standard in a myelopathic cohort. Finally, these gait metrics are likely not specific to CSM and may be influenced by a range of comorbidities and individual factors that affect free-living gait. Lumbar stenosis, hip or knee osteoarthritis, peripheral neuropathy, vestibular disease, parkinsonism, multiple sclerosis, frailty, general activity level, and cardiopulmonary limitations may influence gait independent of CSM. Despite the added variance introduced by such factors, the robustness of our results lends further support to the use of consumer-grade technology. More comprehensive longitudinal validation against research-grade technology would further clarify the strengths and limitations of consumer-grade monitoring.

Recognizing the strengths and limitations in our results, prospective, external validation in a larger cohort is now indicated to more definitively characterize the performance characteristics of Apple Health monitoring. Beyond the mJOA, future work should evaluate these measures against objective neurological examination findings, quantitative gait assessments, and prospectively documented neurological progression. These metrics should also be compared to research-grade trackers for further validation. Finally, as Android-based smartphone platforms expand availability of gait quality metrics, further work will be needed to establish the reliability of such algorithms, particularly given greater variability in smartphone sensors in Android devices.⁶⁰

In summary, passively collected, consumer-grade gait data carry a clinically meaningful neurological signal in CSM that distinguished patients from controls, correlated with mJOA-defined disease severity, and identified sustained within-person gait change preceding clinical diagnosis, with worst-day summaries outperforming medians throughout. These findings support prospective, externally validated studies of passive gait monitoring in CSM and other disorders of gradual motor decline.

Declarations

Data availability

The datasets generated and analyzed during this study are not publicly available because they contain potentially identifying passive sensor data from individual participants, but are available from the corresponding author on reasonable request and with institutional review board approval.

Code availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Acknowledgements

This study was funded by the Washington University Institute of Clinical and Translational Sciences and The Foundation for Barnes-Jewish Hospital (CTFRP1916), supported by the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH; UL1TR002345). Additional funding was provided by the Washington University Here and Next Research Program (PJ000029167), the University of Missouri Spinal Cord Injury/Disease Research Program (AW00012496/GR0032927), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (1K23AR082986-01A1), and NCATS/NIH (TL1TR002344). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funders had no role in study design, data collection, analysis or interpretation, or manuscript preparation. The authors thank the patients and volunteers who contributed data to this study.

Author contributions

P.S., F.A., E.C., J.W., and J.K.G. conceived of and designed the study. P.S. developed the analysis pipeline, performed the statistical analysis, and drafted the original manuscript, with assistance from F.A., E.C., and J.W. P.S., F.A., A.L., E.C., S.Y., J.W., J.W.Y., S.B., T.M., Z.G., M.P.S., W.Z.R., and J.K.G. contributed to the analysis and interpretation of the data and revised the manuscript for important intellectual content. J.K.G. provided study supervision and administrative support. All authors read and approved the final submitted version of the manuscript.

Competing interests

S.B. reports royalties and patent-related interests involving VB Spine and Globus; research support from VB Spine, Medtronic, DePuy Synthes, and Globus; and educational grant support from Medtronic, Globus, VB Spine, Stryker, Orthofix, and ATEC. The remaining authors declare no financial or non-financial competing interests.

References

1. Badhiwala, J. H. *et al.* Degenerative cervical myelopathy - update and future directions. *Nat. Rev. Neurol.* 16, 108–124 (2020).
2. Grodzinski, B., Stubbs, D. J. & Davies, B. M. Most degenerative cervical myelopathy remains undiagnosed, particularly amongst the elderly: modelling the prevalence of degenerative cervical myelopathy in the United Kingdom. *J. Neurol.* 270, 311–319 (2023).
3. Kovalova, I. *et al.* Prevalence and Imaging Characteristics of Nonmyelopathic and Myelopathic Spondylotic Cervical Cord Compression. *Spine* 41, 1908–1916 (2016).
4. Wilson, J. R. *et al.* Frequency, timing, and predictors of neurological dysfunction in the nonmyelopathic patient with cervical spinal cord compression, canal stenosis, and/or ossification of the posterior longitudinal ligament. *Spine* 38, S37–54 (2013).
5. Theodore, N. Degenerative Cervical Spondylosis. *N. Engl. J. Med.* 383, 159–168 (2020).
6. Kane, S. F., Abadie, K. V. & Willson, A. Degenerative Cervical Myelopathy: Recognition and Management. *Am. Fam. Physician* 102, 740–750 (2020).
7. Ozaki, M. *et al.* Symptom-specific Analysis of Surgical Outcomes and Predictors of Residual Symptoms in Patients With Mild Degenerative Cervical Myelopathy: An Analysis of Cases With Severe Cord Compression or Progressive Symptoms. *Spine* 51, 556–563 (2026).
8. Brannigan, J. F. M. *et al.* Management of mild degenerative cervical myelopathy and asymptomatic spinal cord compression: an international survey. *Spinal Cord* 62, 51–58 (2024).
9. Tetreault, L., Nouri, A., Kopjar, B., Côté, P. & Fehlings, M. G. The Minimum Clinically Important Difference of the Modified Japanese Orthopaedic Association Scale in Patients with Degenerative Cervical Myelopathy. *Spine* 40, 1653–1659 (2015).
10. Davies, B. M. *et al.* Development of a core measurement set for research in degenerative cervical myelopathy: a study protocol (AO Spine RECODE-DCM CMS). *BMJ Open* 12, e060436 (2022).
11. Singh, P. *et al.* A Qualitative Study of Patient and Healthcare Provider Perspectives on Mobile Health Assessments for Cervical Spondylotic Myelopathy. 2026.03.04.26347622 Preprint at <https://doi.org/10.64898/2026.03.04.26347622> (2026).
12. Breasail, M. Ó. *et al.* Wearable GPS and Accelerometer Technologies for Monitoring Mobility and Physical Activity in Neurodegenerative Disorders: A Systematic Review. *Sensors* 21, 8261 (2021).

13. Ali, F., Padilla, H., Blazek, A. M., Barnard, L. & Kaufman, K. R. Gait Analysis in Neurologic Disorders: Methodology, Applications, and Clinical Considerations. *Neurology* 105, e214154 (2025).
14. Lewis, A. et al. Validation of Gait Tasks in SynapTrack Mobile App for Cervical Spondylotic Myelopathy. 2026.05.27.26354225 Preprint at <https://doi.org/10.64898/2026.05.27.26354225> (2026).
15. Singh, P. et al. A Neurosurgeon's Guide to Mobile Health Application Development With a Case Study for Cervical Myelopathy. *Neurosurgery* 98, 726–736 (2026).
16. Yakdan, S. et al. Preliminary Reliability and Validity of SynapTrack, a Smartphone-Based Digital Biomarker Platform for Remote Assessment of Cervical Spondylotic Myelopathy. 2026.05.29.26354454 Preprint at <https://doi.org/10.64898/2026.05.29.26354454> (2026).
17. Lipsmeier, F. et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. *Mov. Disord. Off. J. Mov. Disord. Soc.* 33, 1287–1297 (2018).
18. Bot, B. M. et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. *Sci. Data* 3, 160011 (2016).
19. Rama, E., Zuberi, S., Aly, M., Askari, A. & Iqbal, F. M. Clinical Outcomes of Passive Sensors in Remote Monitoring: A Systematic Review. *Sensors* 25, 3285 (2025).
20. Antonini, A. et al. Smartphone as a Sensor in mHealth: Narrative Overview, SWOT Analysis, and Proposal of Mobile Biomarkers. *Sensors* 25, 3655 (2025).
21. Apple Inc. Measuring Walking Quality Through iPhone Mobility Metrics. https://www.apple.com/healthcare/docs/site/Measuring_Walking_Quality_Through_iPhone_Mobility_Metrics.pdf (2021).
22. Werner, C. et al. Validity and reliability of the Apple Health app on iPhone for measuring gait parameters in children, adults, and seniors. *Sci. Rep.* 13, 5350 (2023).
23. Brodke, D. J. et al. The Future Is Mobile: Pilot Validation Study of Apple Health Metrics in Orthopaedic Trauma. *J. Bone Joint Surg. Am.* 107, 1825–1832 (2025).
24. Tetreault, L. et al. The modified Japanese Orthopaedic Association scale: establishing criteria for mild, moderate and severe impairment in patients with degenerative cervical myelopathy. *Eur. Spine J. Off. Publ. Eur. Spine Soc. Eur. Spinal Deform. Soc. Eur. Sect. Cerv. Spine Res. Soc.* 26, 78–84 (2017).
25. Kopjar, B., Tetreault, L., Kalsi-Ryan, S. & Fehlings, M. Psychometric properties of the modified Japanese Orthopaedic Association scale in patients with cervical spondylotic myelopathy. *Spine* 40, E23-28 (2015).
26. Patterson, K. K., Gage, W. H., Brooks, D., Black, S. E. & McIlroy, W. E. Evaluation of gait symmetry after stroke: a comparison of current methods and recommendations for standardization. *Gait Posture* 31, 241–246 (2010).
27. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300 (1995).
28. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44, 837–845 (1988).
29. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. in *Selected Papers of Hirotugu Akaike* (eds Parzen, E., Tanabe, K. & Kitagawa, G.) 199–213 (Springer, New York, NY, 1998). doi:10.1007/978-1-4612-1694-0_15.
30. Sachs, M. C., Brand, A. & Gabriel, E. E. Confidence bands in survival analysis. *Br. J. Cancer* 127, 1636–1641 (2022).
31. Montiel Olea, J. L. & Plagborg-Møller, M. Simultaneous confidence bands: Theory, implementation, and an application to SVARs. *J. Appl. Econom.* 34, 1–17 (2019).

32. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* 15, 155–163 (2016).
33. Alkan, A. et al. Minimal Detectable Changes of the Health Assessment Questionnaire-Disability Index, Patient-Reported Outcomes Measurement Information System-29 Profile Version 2.0 Domains, and Patient Health Questionnaire-8 in People With Systemic Sclerosis: A Scleroderma Patient-Centered Intervention Network Cohort Cross-Sectional Study. *Arthritis Care Res.* 76, 1549–1557 (2024).
34. Kaplan, E. L. & Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* 53, 457–481 (1958).
35. Grambsch, P. M. & Therneau, T. M. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 81, 515–526 (1994).
36. Shah, V. V. et al. Laboratory versus daily life gait characteristics in patients with multiple sclerosis, Parkinson's disease, and matched controls. *J. Neuroengineering Rehabil.* 17, 159 (2020).
37. Moon, Y., Sung, J., An, R., Hernandez, M. E. & Sosnoff, J. J. Gait variability in people with neurological disorders: A systematic review and meta-analysis. *Hum. Mov. Sci.* 47, 197–208 (2016).
38. Tian, Q. et al. Association of Combined Slow Gait and Low Activity Fragmentation With Later Onset of Cognitive Impairment. *JAMA Netw. Open* 4, e2135168 (2021).
39. Soltani, A. et al. Real-world gait speed estimation, frailty and handgrip strength: a cohort-based study. *Sci. Rep.* 11, 18966 (2021).
40. Mandelli, F. et al. Gait function assessed using 3D gait analysis in patients with cervical spinal myelopathy before and after surgical decompression: a systematic review and meta-analysis. *Spine J. Off. J. North Am. Spine Soc.* 24, 406–416 (2024).
41. Li, Z. et al. Wearable sensor-based gait evaluation in cervical spondylotic myelopathy: detecting asymptomatic gait dysfunction and heterogeneous gait impairment. *Spine J. Off. J. North Am. Spine Soc.* 26, 1104–1123 (2026).
42. Mirelman, A. et al. Digital Mobility Measures: A Window into Real-World Severity and Progression of Parkinson's Disease. *Mov. Disord. Off. J. Mov. Disord. Soc.* 39, 328–338 (2024).
43. Corrà, M. F. et al. Comparison of Laboratory and Daily-Life Gait Speed Assessment during ON and OFF States in Parkinson's Disease. *Sensors* 21, 3974 (2021).
44. Servais, L. et al. Stride Velocity 95th Centile: Insights into Gaining Regulatory Qualification of the First Wearable-Derived Digital Endpoint for use in Duchenne Muscular Dystrophy Trials. *J. Neuromuscul. Dis.* 9, 335–346 (2022).
45. Shah, V. V. et al. Quantity and quality of gait and turning in people with multiple sclerosis, Parkinson's disease and matched controls during daily living. *J. Neurol.* 267, 1188–1196 (2020).
46. Shah, V. V. et al. Gait Variability in Spinocerebellar Ataxia Assessed Using Wearable Inertial Sensors. *Mov. Disord. Off. J. Mov. Disord. Soc.* 36, 2922–2931 (2021).
47. Statland, J. M., Karanevich, A., Bruetsch, A. & Huisinga, J. A pilot study of the responsiveness of wireless motion analysis in facioscapulohumeral muscular dystrophy. *Muscle Nerve* 60, 590–594 (2019).
48. Pope, D. H., Mowforth, O. D., Davies, B. M. & Kotter, M. R. N. Diagnostic Delays Lead to Greater Disability in Degenerative Cervical Myelopathy and Represent a Health Inequality. *Spine* 45, 368–377 (2020).
49. Del Din, S. et al. Gait analysis with wearables predicts conversion to parkinson disease. *Ann. Neurol.* 86, 357–367 (2019).
50. Schalkamp, A.-K., Peall, K. J., Harrison, N. A. & Sandor, C. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. *Nat. Med.* 29, 2048–2056 (2023).

51. Iglesias, N., Petersen, P. H. & Ricós, C. Power function of the reference change value in relation to cut-off points, reference intervals and index of individuality. *Clin. Chem. Lab. Med.* 43, 441–448 (2005).
52. Lefante, J. J. The power to detect differences in average rates of change in longitudinal studies. *Stat. Med.* 9, 437–446 (1990).
53. Fehlings, M. G. *et al.* Symptomatic progression of cervical myelopathy and the role of nonsurgical management: a consensus statement. *Spine* 38, S19-20 (2013).
54. Matz, P. G. *et al.* The natural history of cervical spondylotic myelopathy. *J. Neurosurg. Spine* 11, 104–111 (2009).
55. Mechas, C.-A., Sarraj, M. & Helgeson, M. Diagnosis and Management of Degenerative Cervical Myelopathy. *J. Am. Acad. Orthop. Surg.* 33, e1462–e1476 (2025).
56. Kollmansperger, S., Decker, J., Jahn, K., Möhwald, K. & Wuehr, M. Reliability and minimal clinically important differences of gait characteristics in peripheral vestibular disorders. *Front. Neurol.* 17, 1818995 (2026).
57. Soulard, J., Vaillant, J., Balaguier, R. & Vuillerme, N. Spatio-temporal gait parameters obtained from foot-worn inertial sensors are reliable in healthy adults in single- and dual-task conditions. *Sci. Rep.* 11, 10229 (2021).
58. Ungerer, M. N. & Gumbinger, C. A Comparative Survey on Daily Health Habits Among iPhone and Android Smartphone Users. *Am. J. Lifestyle Med.* 15598276241268195 (2024) doi:10.1177/15598276241268195.
59. Sturrock, S. L., Moineddin, R., Gesink, D., Woodruff, S. & Fuller, D. The effect of software and hardware version on Apple Watch activity measurement: A secondary analysis of the COVFIT retrospective cohort study. *PLOS Digit. Health* 4, e0000727 (2025).
60. Halabi, R. *et al.* Comparative Assessment of Multimodal Sensor Data Quality Collected Using Android and iOS Smartphones in Real-World Settings. *Sensors* 24, 6246 (2024).

Figures

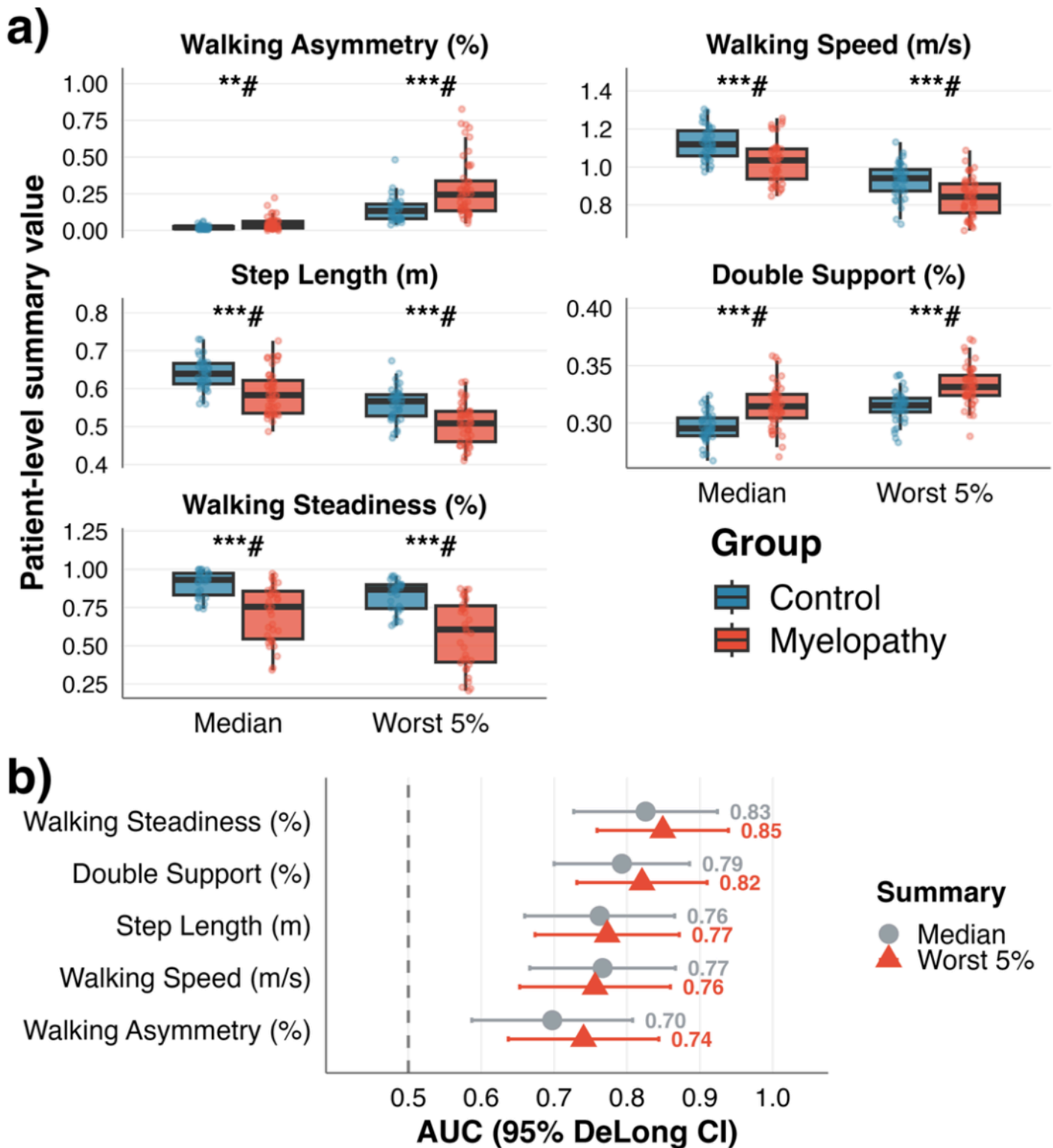


Figure 1

CSM and control gait differences are accentuated in the worst-case tail. Gait metrics compared between controls (blue) and patients (red) at the participant level. a) Boxplots of patient median versus worst-5% (direction-corrected per metric); points are individual participants with significance levels annotated. b) Control-versus-patient discrimination (AUC) per metric for the median (gray circle) and worst-5% (red triangle). The dashed vertical line marks chance (0.5). Groups were compared by Wilcoxon test with Benjamini-Hochberg FDR. Asterisks denote

unadjusted significance (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$); # denotes comparisons that remain significant after FDR adjustment.

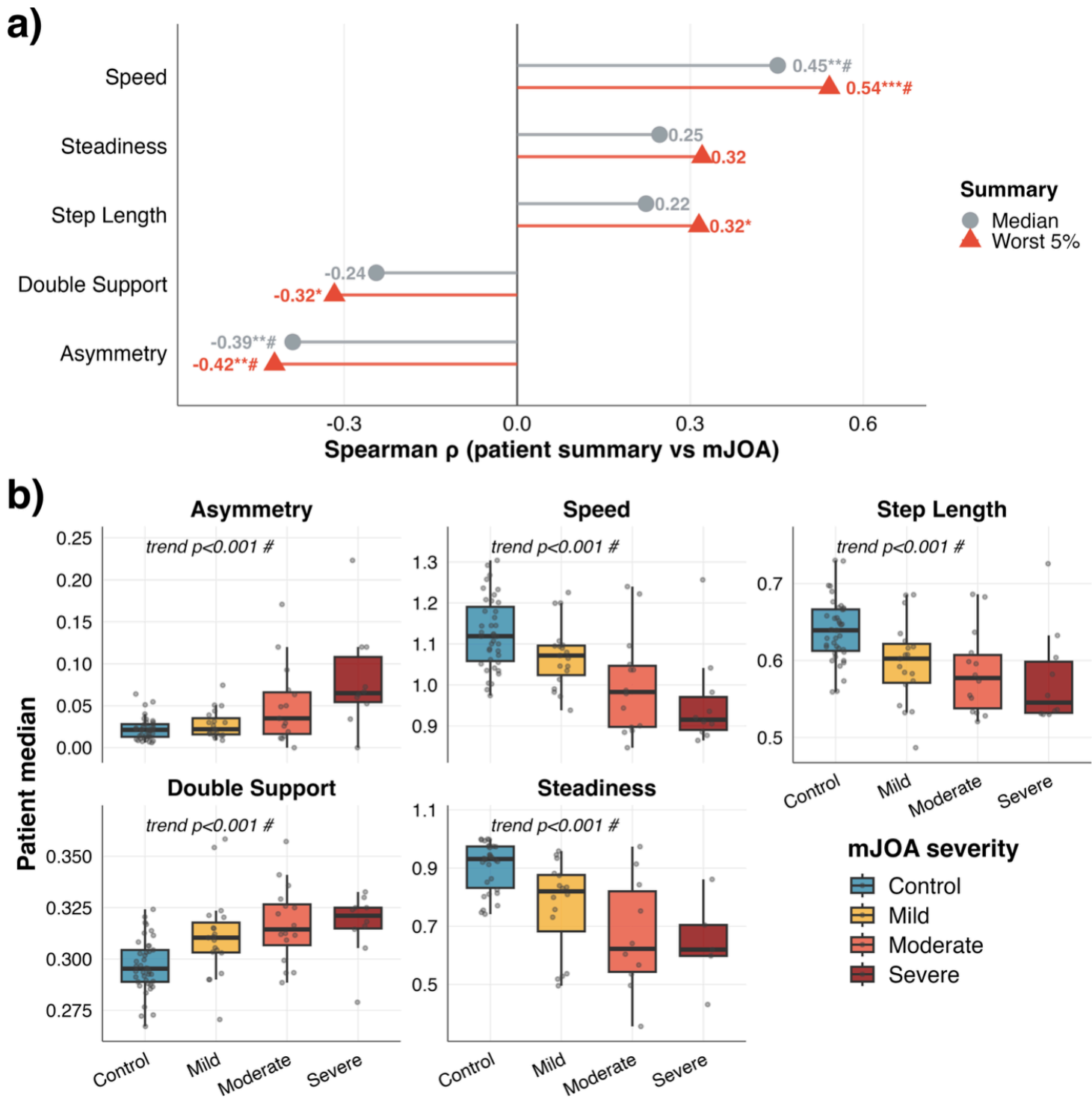


Figure 2

CSM gait metrics track clinical severity. a) Among CSM patients, Spearman correlations between each of the five primary gait metrics and mJOA. Each metric is summarized with the patient-level median (gray circle) and the direction-corrected worst-5% (red triangle). b) Participant-level median of each metric across participant groups. Controls and CSM patients stratified by mJOA (mild ≥ 15 , moderate 12-14, severe < 12) shown as boxplots with individual patients overlaid. An ordered-trend Spearman p-value (metric median vs ordinal severity rank) is annotated

per metric. Asterisks denote unadjusted significance (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$); # denotes comparisons that remain significant after FDR adjustment.

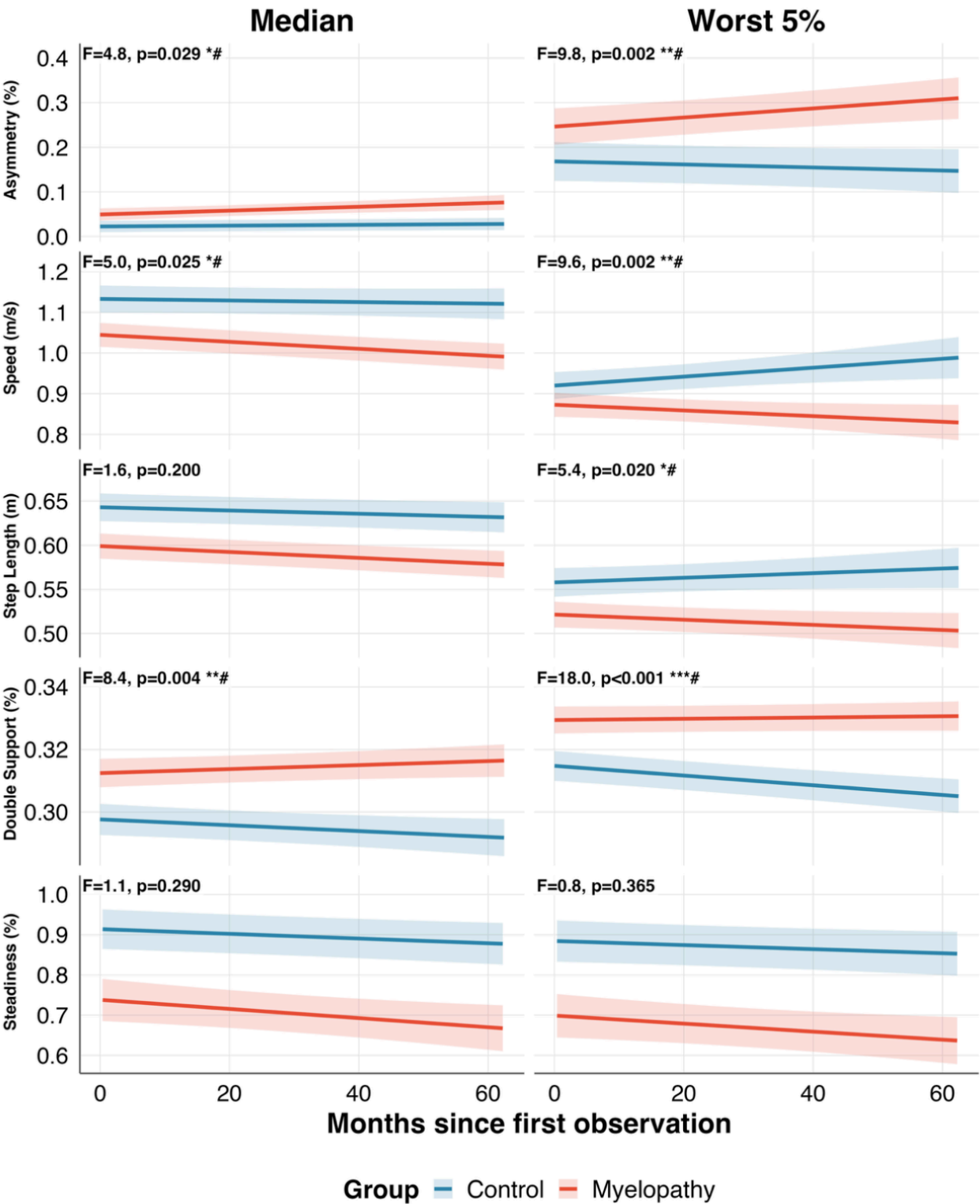


Figure 3

Longitudinal gait trajectories and model selection. Change in the five primary metrics over months since each participant's first observation. Best-fit linear mixed-model group trajectories (controls blue, patients red; ribbons represent 95% CI), with metrics in rows and the median versus worst-5% summary in columns; each panel notes the

time $\times$ group interaction (F, p). Models were fit by maximum likelihood (nlme) with AR(1) and residual-variance options. Asterisks denote unadjusted significance (*p < 0.05, **p < 0.01, ***p < 0.001); # denotes comparisons that remain significant after FDR adjustment.

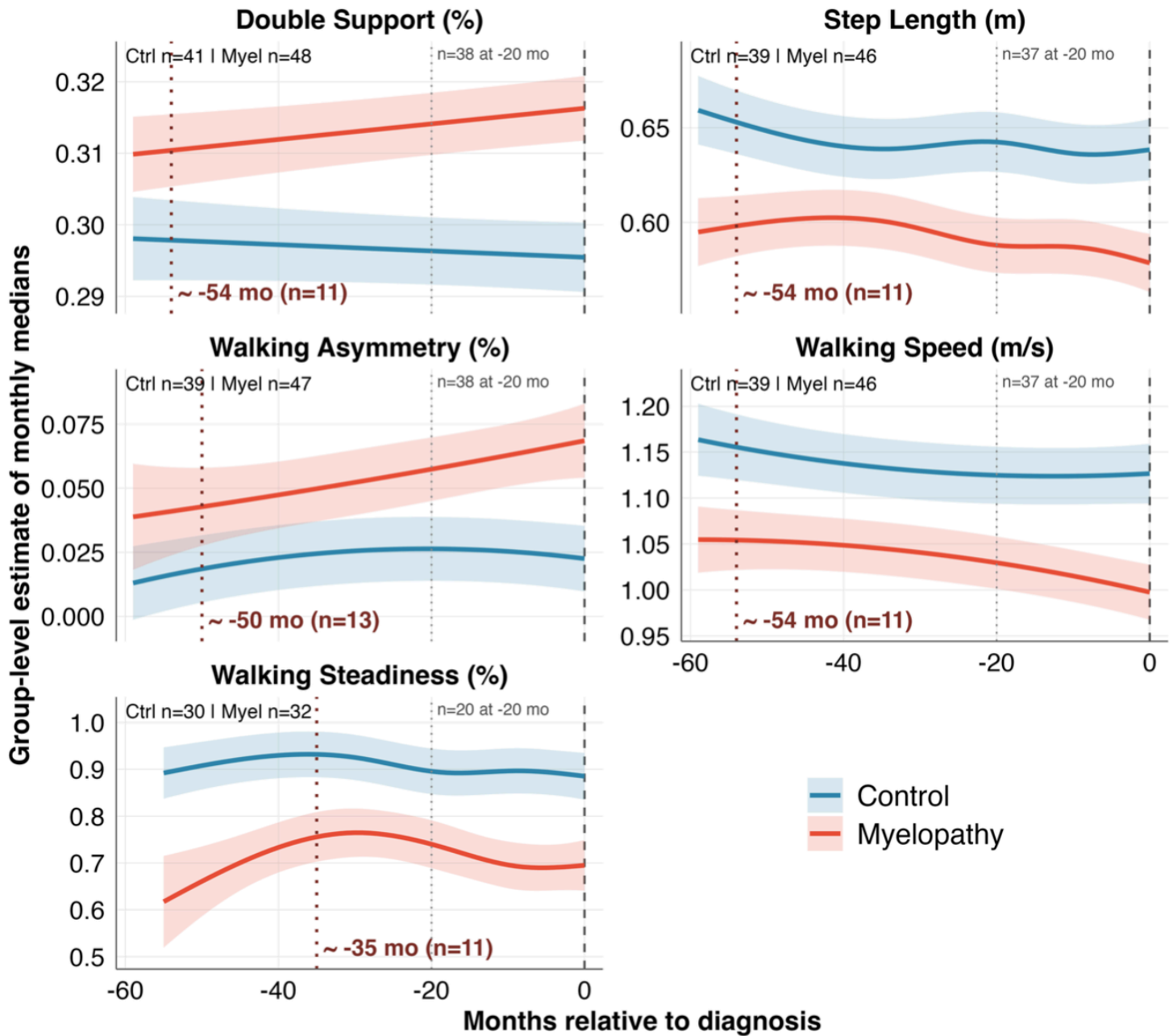


Figure 4

Diagnosis-aligned group trajectories diverge from controls before diagnosis. Patients aligned to diagnosis (month 0); controls aligned to their most recent observation. Best-fitting AIC-selected mixed-model group-level estimates of monthly medians for the five metrics (controls blue, patients red; ribbons represent pointwise 95% CI) over -60 to 0 months; dashed line = diagnosis, dotted line = marked group divergence. Divergence was the earliest disease-time month at which simultaneous (sup-t) confidence band of the patient-control difference excluded zero for ≥ 2 consecutive months with ≥ 10 patients contributing.

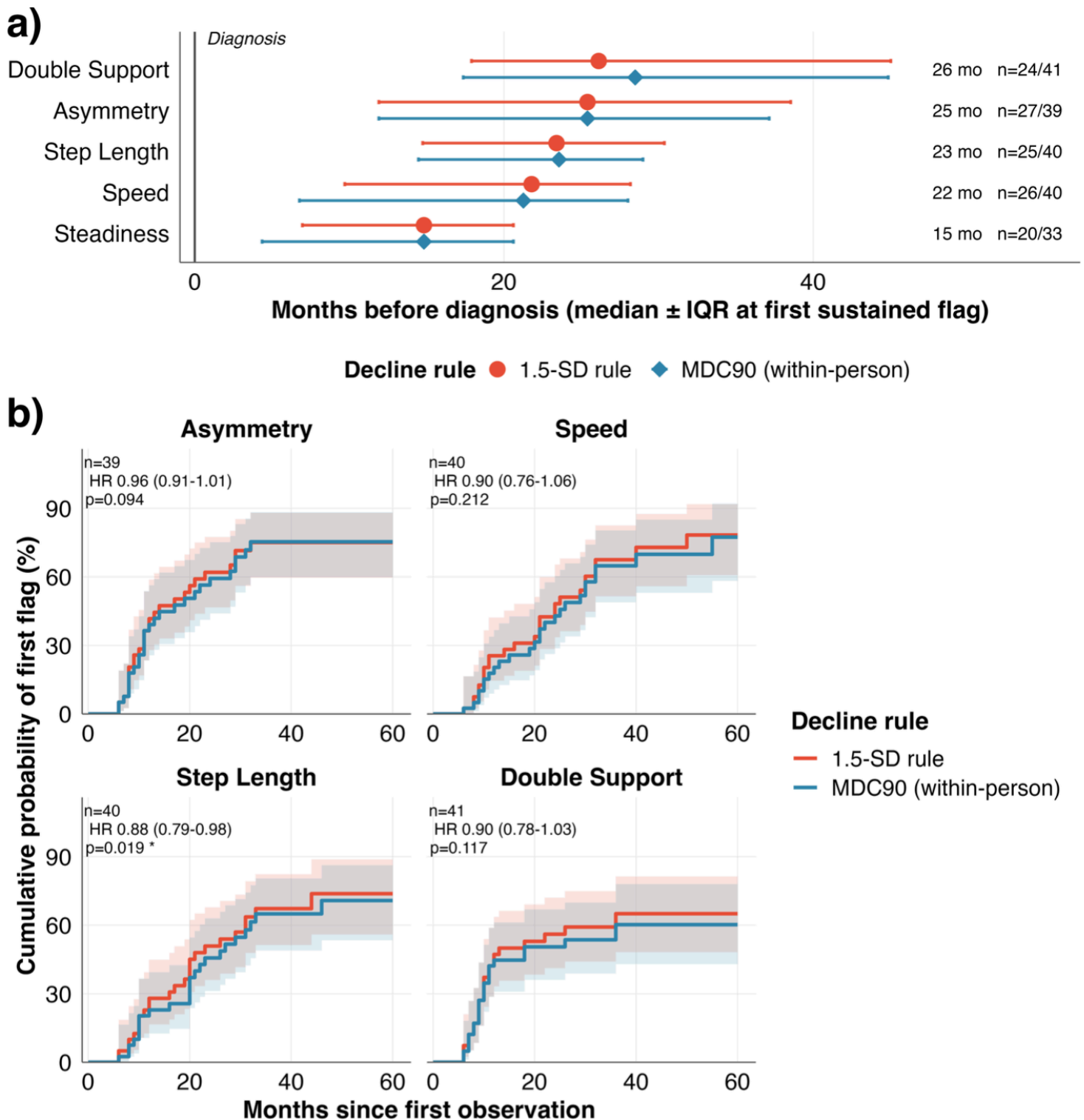


Figure 5

Within-person monitoring flags decline before clinical diagnosis. Individual early detection using each CSM patient as their own baseline. A patient's personal baseline was the mean and SD of the first 6 months of monitoring for all metrics. A month was counted as abnormal when the change from that baseline exceeded a threshold, and a patient was flagged at the first run of ≥ 2 consecutive abnormal months. Two thresholds are shown: 1.5 baseline SD, and an individualized 90% change threshold based on the minimal detectable change framework ($MDC90 = 1.645 \times \sqrt{(1 + 1/k) \times SEM}$, where k is the number of baseline monthly values). a) Lead-time forest plot. For each metric, the median

(point) and IQR (whisker) months before diagnosis at which the first sustained flag occurred, for the 1.5-SD rule (red circle) and MDC90 rule (blue diamond); the vertical line marks diagnosis (0), and the right-hand column lists the median lead time and the number of patients flagged out of those eligible. b) Time to first flag, shown as a Kaplan-Meier cumulative-probability curve ($1 - \text{survivor function}$) on the monitoring-time axis (months since each patient's first observation), with one curve per decline rule within each metric. Patients who had not yet flagged were censored at their last monitored month. Bands are 95% confidence intervals (log-log), and the per-facet n is the number of patients entering the analysis. Each panel is annotated with n , the hazard ratio for the MDC90 rule versus the 1.5-SD rule from a cluster-robust marginal Cox model with 95% CI, and the corresponding p -value. Asterisks denote unadjusted significance (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.pdf](#)