Digital Phenotypes of Instability and Fatigue Derived From Daily Standing Transitions in Persons With Multiple Sclerosis

Tyler VanDyk, Brett Meyer, *Graduate Student Member, IEEE*, Paolo DePetrillo, Nicole Donahue, Aisling O'Leary, Sam Fox, Nick Cheney, Melissa Ceruolo, Andrew J. Solomon, and Ryan S. McGinnis, *Senior Member, IEEE*

***Abstract*— Impairment in persons with multiple sclerosis (PwMS) can often be attributed to symptoms of motor instability and fatigue. Symptom monitoring and queued interventions often target these symptoms. Clinical metrics are currently limited to objective physician assessments or subjective patient reported measures. Recent research has turned to wearables for improving the objectivity and temporal resolution of assessment. Our group has previously observed wearable assessment of supervised and unsupervised standing transitions to be predictive of fall-risk in PwMS. Here we extend the application of standing transition quantification to longitudinal home monitoring of symptoms. Subjects (N=23) with varying degrees of MS impairment were recruited and monitored with accelerometry for a total of ∼6 weeks each. These data were processed using a preexisting framework, applying a deep learning activity classifier to isolate periods of standing transition from which descriptive features were extracted for analysis. Participants completed daily and biweekly assessments describing their symptoms. From these data, Canonical Correlation Analysis was used to derive digital phenotypes of MS instability and fatigue. We find these phenotypes capable of distinguishing fallers from non-fallers, and further that they demonstrate a capacity to characterize symptoms at both daily and sub-daily resolutions. These results represent promising support for future applications of wearables, which may soon augment or replace current metrics in longitudinal monitoring of PwMS.**

***Index Terms*— Wearables, digital phenotypes, falls, fatigue, multiple sclerosis.**

I. Introduction

MULTIPLE Sclerosis (MS) is an immune-mediated, demyelinating disease of the central nervous system, notable for its variable clinical presentation [1]. Persons with MS (PwMS) may experience a broad range of symptoms depending on the quantity, localization, and magnitude of demyelinating lesions. Symptoms often include a combination of cognitive (e.g., fatigue, dementia) and/or sensorimotor deficits (e.g., visual impairment, reduced mobility) [1], [2], [3].

Postural instability is a highly prevalent presentation with significant implications for patient quality of life (QoL). Ambulation impairment, *e.g.*, presents in up to 75% of PwMS and can cause significant reduction in patient mobility and independence [4], [5]. Falls are a particularly fearsome complication of postural instability and may have long-term consequences to QoL including injury, further loss of mobility, persistent fear of falling, and death [6], [7]. Unfortunately, even with close monitoring and queued interventions, more than 50% of PwMS report falls within any 3-month period [6], [7], [8]. Fatigue also presents prominently, impacts fall risk, and is often cited as a primary concern in PwMS [9], [10]. Significant effort has been invested studying metrics to predict patient symptoms and indicate timely interventions. Clinical assessments, such as the expanded disability status scale (EDSS, [11]) and performance based measures (PBMs), *e.g.*, timed-walk tests provide symptom insight [12]. However, these measures are limited to active observations or physician assessment [13].

Patient reported measures (PRMs) are perhaps the most widely used tool for accurate and time-resolved insight on patient symptoms [14]. The Activity-specific Balance Confidence Scale (ABC or ABCS, [15]) and Modified Fatigue Impact Scale (MFIS, [16]) have both been seen to correlate well with clinical assessments and fall risk (e.g., [9]). However, PRMs also have limitations; patient adherence to accurate documentation tends to trend inversely with the frequency of surveys and PRMs are fundamentally subjective and so may be influenced by unmodeled factors. Recent efforts have begun to transition toward objective biomarkers of symptomology [17], [18]. Applications of wearable devices have broad potential in this space, ranging from improved objectivity in standard clinical assessments to remote symptom monitoring.

A key challenge of remote monitoring is the need to identify meaningful data to analyze amongst high volume datasets. One common technique is to inspect data from standardized tasks, which are either assigned to the patient

Manuscript received 7 November 2022; revised 18 March 2023; accepted 25 April 2023. Date of publication 28 April 2023; date of current version 15 May 2023. This work was supported in part by the U.S. National Institutes of Health under Grant EB027852 and in part by Medidata Solutions. *(Corresponding author: Ryan S. McGinnis.)*

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Vermont Institution Review Board under Application No. STUDY00000401 (initial approval 7/7/2019).

Tyler VanDyk and Andrew J. Solomon are with the Department of Neurological Sciences, The University of Vermont, Burlington, VT 05405 USA.

Brett Meyer, Nicole Donahue, Aisling O'Leary, Sam Fox, Nick Cheney, and Ryan S. McGinnis are with the Department of Electrical and Biomedical Engineering, The University of Vermont, Burlington, VT 05405 USA (e-mail: ryan.mcginnis@uvm.edu).

Paolo DePetrillo and Melissa Ceruolo are with Medidata Solutions, Dassault Systèmes, New York, NY 12601 USA.

Digital Object Identifier 10.1109/TNSRE.2023.3271601

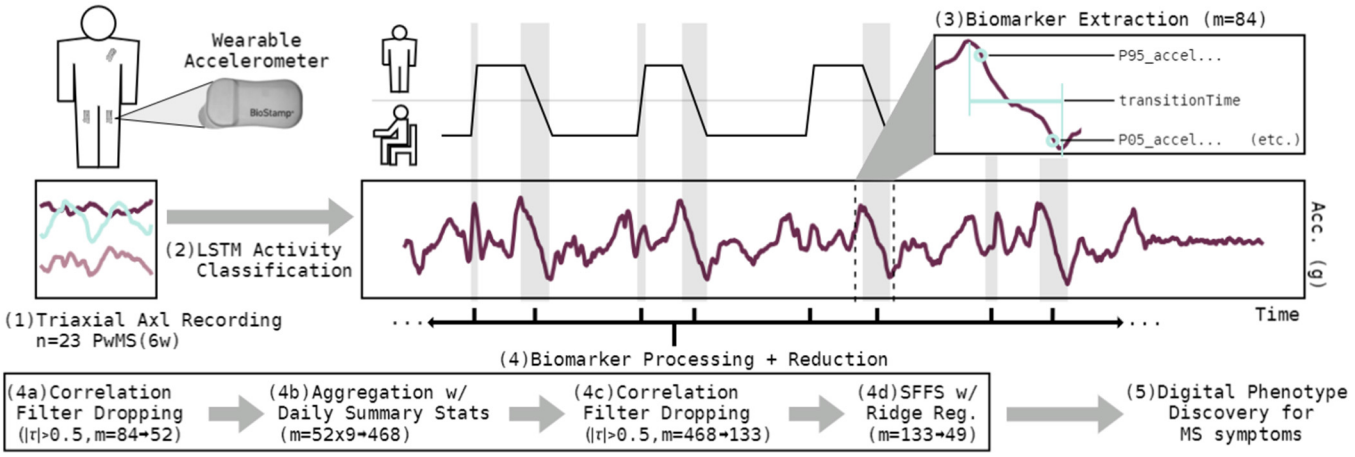

Fig. 1. Summary of approach for identifying digital phenotypes of instability and fatigue from wearables-derived measures of standing transitions.

(e.g., [19], [20], [21]) or performed naturally during daily life (e.g., [22], [23], [24], [25]). An example of the latter are daily transitions from sitting to standing (sist) or standing to sitting (stsi). These transitions have been used extensively to assess patient stability, dating back to the introduction of the "get-up and go" test in 1986 [26]. More modernly, wearable monitoring of such tasks has been applied to augmentation of PBMs and home-monitoring of symptoms [22], [23], but our understanding of these metrics remains limited [24], [27].

The purpose of this study is to investigate the ability of wearables-derived biomarkers from daily standing transitions to capture the constructs of balance confidence, fatigue, and fall risk in PwMS (Fig. 1). If successful, these biomarkers could provide objective and continual monitoring to capture a more complete picture of patient symptoms and their fluctuations. These data could be invaluable to expand our understanding of symptoms and to inform comprehensive care of PwMS.

II. Materials and Methods

A. Experimental Protocol

We considered data from 23 PwMS (5:18 Male:Female, mean ± standard deviation age 50 ± 9.7 y/o) recruited from the Multiple Sclerosis Center at University of Vermont Medical Center (inclusion: no condition affecting balance/mobility other than MS, ambulatory without aid, no known hypersensitivity to adhesives or hydrogel, not pregnant or breastfeeding).

Participants were asked to complete 12-weeks of at home monitoring with biweekly sensor wear, yielding 6 weeks of sensor data for analysis. All participants completed 2 weeks of sensor wear, 22 participants completed 5 weeks, and 20 completed 6 weeks. During the sensor-wear weeks, subjects were instrumented with three BioStamp nPoint sensors for all hours of the day (Fig. 1, left upper chest, bilateral anterior thigh; recording acceleration - 31.25 Hz, ±16G - and surface biopotentials - 250 Hz). Participants were asked to complete a daily 30-second chair stand test, a one-minute walk, and a 30-second standing balance assessment on each sensor-wear day. Participants also completed a daily falls survey each evening; survey questions covered the occurrence of "true fall(s)" (defined: unintentionally coming to rest on the ground, floor or other lower level) and/or "near fall(s)" (defined: a period of instability relative to your baseline, but you regain stability such that you ultimately do not fall) each day and documented the approximate time of each occurrence. Participants were also asked to complete the ABCS nightly and the Modified Fatigue Impact Scale (MFIS) at the end of each non-sensor wear week (4-week recall period). Participants also completed the Patient-Determined Disease Steps (PDDS) once [28]. The mean (std) survey results for our cohort were: PDDS 0.88 (1.05); ABCS 77.6 (21.9); MFIS 28.3 (16.1). Protocol was approved by the UVM Institutional Review Board (STUDY00000401).

B. Activity Identification and Feature Extraction

To identify daily activity transitions, we employed our activity classification pipeline, which leverages a deep learning model trained on 4s observations (n >100,000) of acceleration from a variety of patient populations (as previously described [23], [24]), to identify periods of walking, sitting, standing, and lying with an accuracy of 96.7% on a held-out test set. Herein, this model was used to identify subsequent periods of sitting and standing. From matched periods, each sist or stsi transition was identified using an established technique; cranial-caudal acceleration was filtered and inspected for a transition from 1g towards 0 g (stsi) or from 0 g to 1g (sist) within an 18-second window of data centered between the classified activities, automated detection was previously validated by visual inspection of ∼500 transitions [23]. After identifying transitions, the following features were calculated from the thigh and chest accelerations for both the cranial-caudal (CC) and horizontal (horz) planes: $5^{th}/50^{th}/95^{th}$ acceleration percentile (F5, F50, F95), jerk [23], $5^{th}/50^{th}/95^{th}$ percentile frequency, total power, and approximate entropy (ApEn). We also computed (SEF) for both chest and thigh [29], as well as transition time and postural sway features (Jerk, Range, and 50^{th} percentile frequency of chest acceleration) from the standing bout immediately preceding or following the transition [24], [30]. Data were computed

from custom MATLAB scripts using Medidata's Sensor Cloud Network Analytics service.

C. Defining Constructs of Instability and Fatigue

For this study we focused on symptoms of instability and fatigue in PwMS. Instability was described by the ABCS, sampled at daily resolution and well correlated to fall-risk and physical functioning [9], [15]. Fatigue was described by the MFIS, which has been validated for high and low levels of fatigue in PwMS [16], [31]. To match the daily resolution of ABCS, MFIS was super-sampled and averaged over overlapping 4-week recall periods. For normalization, we used published means in similar cohorts – ABCS 63.7 (n = 84, [32]) MFIS 44.2 (n = 268, [31]) – to center and scale survey results without data leakage in cross-validation. These scales are not completely independent, ABCS has been shown to decrease significantly in PwMS who report high fatigue on the MFIS [33]. In our analysis, we considered each scale as informative of both symptom constructs, but primarily descriptive of the designated symptom.

D. Deriving Digital Phenotypes of Instability and Fatigue

All captured transitions across subjects were inspected to derive digital phenotypes of instability and fatigue. Steps included 1) data cleaning, aggregation, and reduction, 2) principal component analysis, and 3) canonical correlation analysis. Analysis was performed in Python using *Numpy* (v1.23.1) [34], *Pandas* (v1.4.3) [35], *SciPy* (v1.7.3) [36], *Scikit-learn* (v0.24.2) [37], *Statannotations* [38], and *Pingouin* [39].

1) Data Cleaning, Aggregation, and Reduction: Extracted features ($m = 42 \times 2$ transition types $= 84$, initial features) were filtered for extreme outliers by masking feature values with a subject-specific T-Score of greater than 4.0. To reduce multicolinearity, features for each transition were filtered by supervised pair-wise correlation to each other using Kendall's Tau. Highly correlated feature pairs ($|\tau| > 0.5$, [40], [41]) were compared to PRM targets (ABCS & MFIS) with a repeated measures correlation and the least predictive feature in each pair was dropped to reduce the initial feature set ($m = 84 \rightarrow 52$, uncorrelated features). Features were aggregated to daily resolution for each subject using 9 summary statistics: mean (mu), standard deviation (sig), minimum (min), median (eta), maximum (max), and 5/25/75/95-percentiles ($m = 52 \times 9 = 468$, aggregated features). These aggregated features were again processed for multicollinearity and filtered with the pair-wise correlation protocol described above to generate a reduced feature set to be used as "biomarkers" for regression ($m = 468 \rightarrow 133$, uncorrelated aggregated features). To select a reasonable estimate for the number of necessary features we applied principal components analysis (method of Tipping and Bishop [43]) and quantified meaningful dimensions of our dataset by the method of Gavish and Donoho ($m = 133 \rightarrow 49$, selected features, noted in Fig. 3) [44]. A final dimension reduction was performed by sequential forward feature selection using ridge regression (tuned by grid search over $\alpha \in$ [10e-2,10e1]) to predict PRM targets, extracting the most informative and least collinear features [45], [46], [47]. All reduction steps listed in Fig. 1.

2) Canonical Correlation Analysis (CCA): CCA was applied to identify digital phenotypes, mapping transition biomarkers to PRM-defined instability and fatigue [42]. CCA is a supervised regression method wherein an input feature set, $X_{n\times m}$, undergoes change of basis into a new projection, $\hat{X}_{n\times k}$. Loadings are tuned to optimize correlation of $\hat{X}_{n\times k}$ – in this case a linear combination of transition biomarkers – onto similarly derived components, $\hat{Y}_{n\times k}$, of a PRM target set. Subject-stratified k-fold cross-validation was applied to the CCA model to ensure against over-fitting with the final reduced feature set. All observations were then used to produce a best-fit CCA model and paired components of $\hat{X}, \hat{Y}$ were designated as digital phenotypes of *Instability* and *Fatigue* respectively, based upon correlation to PRM surveys.

E. Statistical Analysis

Mann-Whitney U-Test was used to test for significant differences between groups. Repeated measures correlation, which compares paired distributions using analysis of covariance to correct Pearson's-r for non-independent data samples [48], was used to evaluate the association between transition features and the ABCS and MFIS. When interpreting these results, we note that repeated measures correlation is associated with higher statistical power than a traditional Pearson's correlation. A value of $r_m = 0.1$, with n = 23 subjects and an average of 30 observations per subject corresponds to a power of >80% [48]. For Pearson's-r, effect size must be much greater ($r \approx 0.55$) to achieve power of 80% with n=23 averaged datapoints, while $r = 0.1$ would correspond to a power of only ~10% [48]. Welch's one-way ANOVA with subsequent Tukey-Kramer pairwise comparison was used to test *Instability* and *Fatigue* distributions across subjects [49]. Subject distributions were first tested for normalcy using a Shapiro-Wilk test and for equal variance (within a factor of 0.33-3.0) [50], [51].

III. Results

A. Transition Biomarkers Correlate to Symptoms

We observe that features derived from dynamic transitions act as digital biomarkers and trend predictably with patient symptoms. *E.g.*, transition time for the sist transition is an easily interpretable biomarker that trends positively with ABCS, MFIS, and fall status (Fig. 2). Other less grossly appreciable biomarkers may also exhibit trends, the 95th percentile of thigh horizontal acceleration (*P95_accel_horz_Thigh*) for the stsi transition demonstrates a negative relationship with fatigue and fall status, although relates less predictably to balance confidence (Fig. 2). These patterns imply measures of standing transitions describe different aspects of symptoms in PwMS.

To further elucidate these relationships, biomarkers were aggregated with summary statistics to a daily resolution and each aggregated feature ($X_{m\times n}, m = 468$ aggregated features, $n = 695$ days) was compared to PRMs using repeated measures correlation.

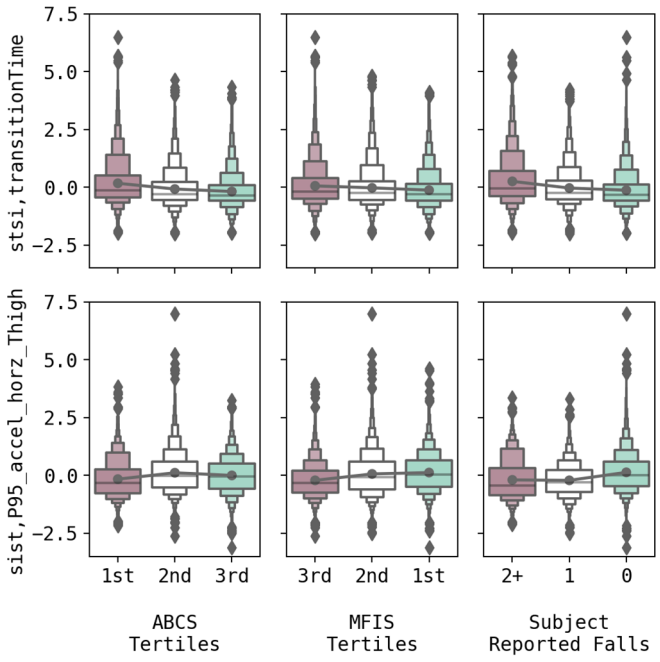

Fig. 2. Boxenplots of selected feature distributions stratified by patient symptom markers. Displayed features are selected only as representatives of different types of features present in the set. Top: Transition time for the stand-to-sit transition, Bot: 95th percentile of horizontal acceleration (recorded at thigh). "ABCS/MFIS Tertiles" calculated from all recorded scores across subjects, boxes contain all recorded transitions for all days within tertile. "Subject reported falls" boxes contain all features for subjects grouped by total reported falls. Color indicates severity of symptoms as tracked by each target. Line trace on each panel connects distribution means. All distributions shown were indicated statistically distinct by pair-wise application of the Mann-Whitney U-test ($\alpha = 0.05$).

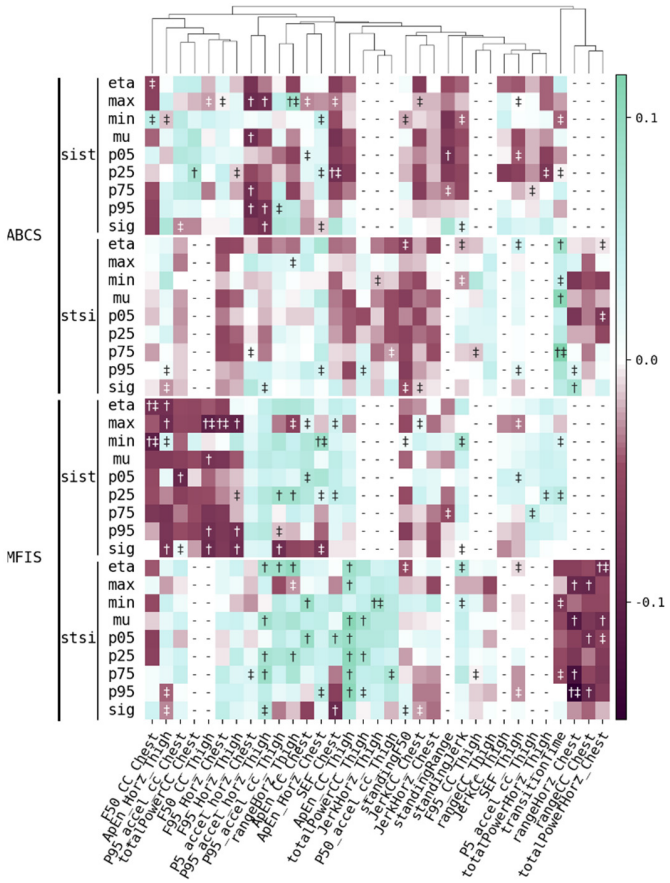

Fig. 3. Repeated measures correlations of aggregated biomarkers to target surveys. Biomarkers on the x-axis are those remaining after supervised correlation filtering. On the y-axis the PRM target, transition type and aggregation statistic for each correlation are displayed. Color indicates direction/magnitude of correlation. Dendrogram displays hierarchical clustering of transition features by similarities in correlation. We note several groups of features share similarities in the contained information (e.g. MFIS compared to transitionTime, rangeHorz_Chest, rangeCC_Chest, and totalPowerHorz_Chest for the stsi transition). Annotations: biomarkers with statistically significant correlation to target (†, $\alpha = 0.05$), selected for CCA (‡), and dropped by filtering (-).

Fifteen aggregated biomarkers were found to be correlated at a statistically significant level ($\alpha = 0.05$) with patient ABCS and 48 with MFIS, without overlap between the two constructs (Fig. 3). These significant correlations ranged from $|r_m| = 0.076-0.148$. Despite low magnitudes (small effect-size), these may be interpreted to represent significant statistical power for repeated measures correlation as discussed in Methods (II-E).

Hierarchal clustering grouped biomarkers by similarities in correlation (Fig. 3). *E.g.*, F50_CC and F95_Horz derived at either the chest or thigh during sist transitions correlate with increased fatigue (lower MFIS). Interestingly these same biomarkers are essentially uncorrelated with PRMs when derived for the stsi transition. In isolation, these features are therefore likely to be more important in the sist transition. Conversely, transition time appears to correlate far better to PRMs for the stsi transition than the sist transition. Transition time of the stsi transition also appears to be among the more important predictors of stability with three aggregation statistics (average, median, and 75th percentile) being significantly correlated to ABCS.

B. CCA Identifies Phenotypes of Instability and Fatigue

The results of Figs. 2 and 3 suggest that individual features contain information about patient symptoms and may describe different aspects of those symptoms. To investigate whether combinations of biomarkers might improve predictive potential, we applied a more deliberate regression-based approach. CCA with 49 appropriately selected features demonstrated successful prediction of test sets on cross-validation (k= 23; Multiple-R^2 = 0.*42*[0.092], mean[SD]), and the model was passed all available observations to generate a single best-fit Multiple-R^2 = 0.48 (Fig. 4).

This regression demonstrates correlation between biomarkers and components of the ABCS and MFIS, and seems to treat the two PRMs as largely orthogonal. The primary direction, $\hat{X}_1$, correlates to $\hat{Y}_1$ (r^2 = 0.62), which exhibits a high magnitude negative correlation to ABCS and a lesser positive correlation to MFIS (Fig. 4). As such it might be considered a digital phenotype of *Instability* with a higher $\hat{X}_1$ value predicting a day with lower balance confidence. In contrast, the secondary direction, $\hat{X}_2$ (r^2 = 0.33) is aligned with a greater positive MFIS correlation, and a negligible positive correlation to ABCS (Fig. 4). As such this may be

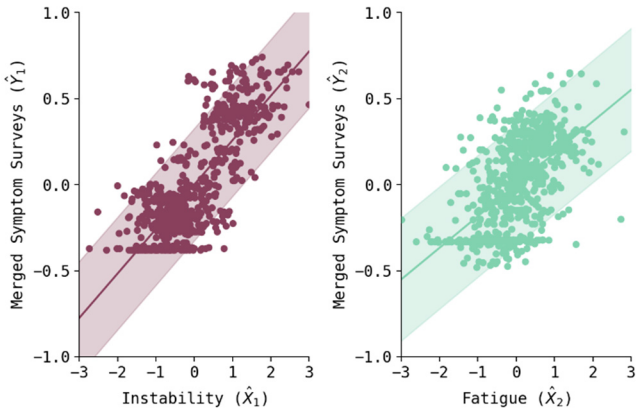

Fig. 4. Resulting CCA projections $\hat{X}$ against $\hat{Y}$. Scattered data for all recorded dates across all subjects. Traces display best-fit linear regression with shaded prediction intervals for future residuals (α = 0.05). Left: phenotype of Instability (survey correlations: r [ABCS, MFIS] = -0.98, 0.67), best-fit $R^2 = 0.62$. Right: phenotype of Fatigue (r [ABCS, MFIS] = 0.17, 0.74), best-fit $R^2 = 0.33$. Uncorrelated pairings ($\hat{X}_1$ v. $\hat{Y}_2$ and $\hat{X}_2$v. $\hat{Y}_1$) not visualized, but R^2 values were checked to validate orthonormalcy ($R^2 < 0.01$).

thought of as a phenotype of *Fatigue* with $\hat{X}_2$ proportional to the degree of subject fatigue on a measured day.

C. Phenotypes of Instability and Fatigue Stratify Fall Risk

Treated as phenotypes of *Instability* and *Fatigue*, the distribution of daily observations should stratify patients by symptoms. Phenotype distributions were isolated for each subject and compared using Welch's one-way ANOVA along each phenotype axis ([*Instability*, *Fatigue*]: F(22, 155) = [80.47, 27.11]; p = [<0.01, <0.01], Fig. 5) with Tukey-Kramer (T-K) groupings.

From this we were able to reject the null hypothesis that these phenotypes have equivalent means across subjects. Furthermore, T-K groups allowed us to identify subjects with similar phenotype distributions (sharing at least one T-K letter group, Fig. 5), and subject groups that might be considered statistically distinct (share no letter groups, Fig. 5). Distributions of *Instability* stratified subjects such that most subjects that reported true falls exhibited distributions near the upper end of the instability spectrum. In fact, all subjects who reported multiple falls during the study were sorted into a single Tukey-Kramer grouping (T-K Group 'B', Fig. 5A). In contrast, $\hat{X}_{2^\circ}$ seems to contain information less associated with fall risk. However, $\hat{X}_{2^\circ}$ may enhance the ability of $\hat{X}_{1^\circ}$ to discriminate fall risk when used jointly (Fig. 6). *E.g.*, we noted two subjects – S0026 and S0017 – who each reported a single fall during the study despite instability distributions toward the stable end of the spectrum. Interestingly, these subjects share a *Fatigue* grouping (T-K group 'A', Fig. 5B) with many of the frequent-fallers, potentially suggesting fatigue independently contributes to fall-risk.

The joint distribution of *Instability*vs. *Fatigue* helps to visualize the differences between subjects who fall more or less frequently (Fig. 6). When subjects are grouped by the number of falls reported, never-fallers (reported true falls = 0) and frequent-fallers (true falls > 1) exhibit distinct peak densities.

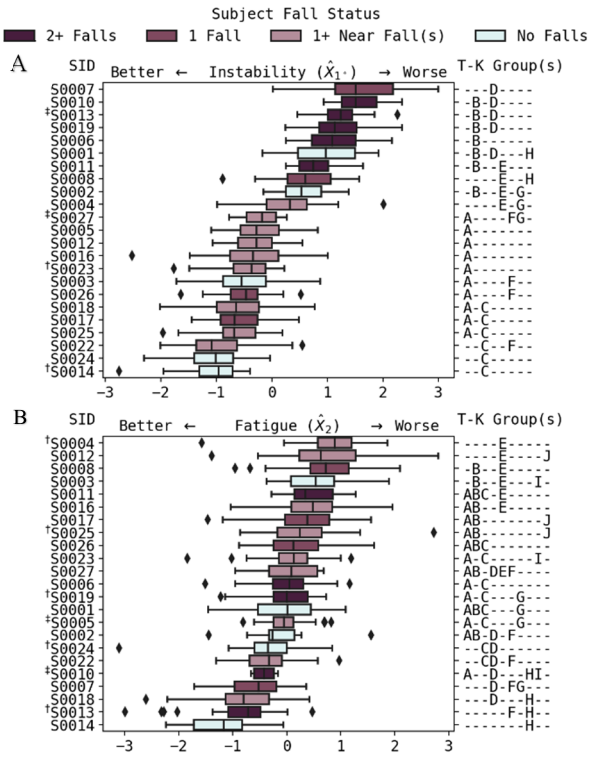

Fig. 5. Left: Boxplots of daily biomarker values for each subject. Color indicates the total number of reported falls for each subject during the study as denoted by the legend. Right: Tukey-Kramer groupings for each subject (a shared letter group between two subjects indicates insufficient evidence to conclude the means are distinct, α =0.05). Subject SID annotations denote results of pre-ANOVA tests for normalcy and equal variance (†= Failed to reject null of non-normalcy, ‡= Variance < 33% of max).

These groups can be further subdivided to separate days with/without reported fall events (Fig. 6B). By these metrics, patients at risk for frequent falls seem equally likely – on any given day – to experience a true fall, no falls, or a near fall where some intervention (*e.g.* catching themselves against a wall) prevents them from falling to the ground. This favors the hypothesis that context plays a large role in the manifestation of true falls. Similarly, never-fallers demonstrate consistently low *Instability* scores, even on days with "near fall" events.

Unlike frequent-fallers and never-fallers, single-fallers (true falls = 1) exhibit a bimodal distribution (Fig. 6). At daily resolution (Fig. 6B), days with near falls in this subset have a peak density more aligned with that of frequent-fallers, while days without reported fall events are better aligned with never-fallers. It may be that PwMS who demonstrate borderline or transient instability are more likely to fluctuate daily than late-stage or asymptomatic patients. In this subset of PwMS, continuous monitoring with a wearable device may be feasibly applied to detect dangerous ranges of instability and suggest just-in-time temporary interventions.

D. Transition Feature Phenotypes Exhibit Daily Trends

Phenotypes would ideally exhibit predictive capacity at sub-daily resolution. To explore this, standing transitions were

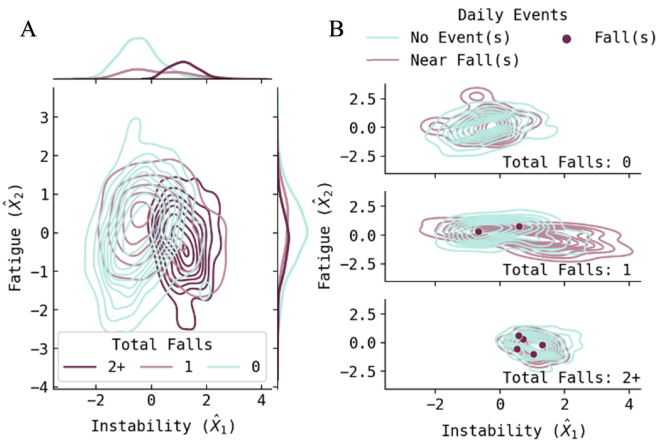

Fig. 6. (A) Joint distributions displayed as kernelized density estimates (KDE), colors indicate subject subsets grouped by the total number of falls reported by each subject during the study. (B) Joint distributions grouped by the worst reported event each day for population subsets; scatter points for days with a reported true fall, KDE for days with near falls or no fall events. Top→Bottom: subjects with 0 reported true fall events (n=15), 1 reported true fall (n=4, note 2 subjects reported a fall on an unmonitored day), 2+ reported true falls (n=5, note only 5 of 17 true falls occurred on monitored dates).

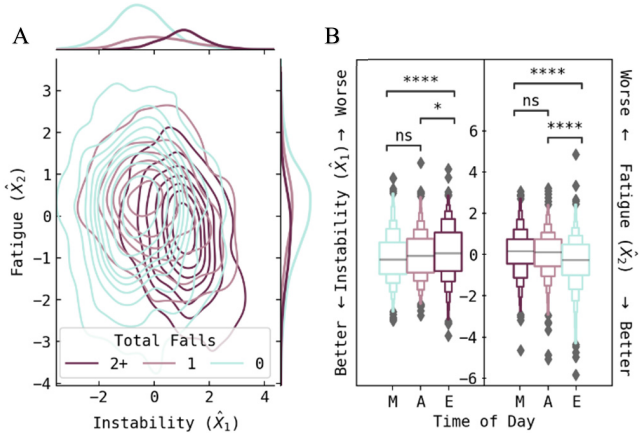

Fig. 7. (A) $\hat{X}$ joint distributions displayed as kernelized density estimates, color indicates subject grouping by total number of reported falls. (B) Distributions of $\hat{X}$ across subjects stratified by time of day, color indicates symptom severity. Statistically distinct distributions indicated as determined by Mann-Whitney U-test ($\alpha = 0.05$).

re-binned and summarized using 7-hour windows (Morning: 05:00-12:00, Afternoon: 12:00-19:00, Evening: 19:00-02:00).

Identical feature selections (m= 49, n= 1830 recorded windows) and CCA loadings as determined at daily resolution were used to recalculate *Instability* and *Fatigue* for each bin. Sub-daily aggregation reduced predictive power; peak densities are less discrete than those observed at daily resolution (Fig. 6A vs. Fig. 7A). This may be a ramification of feature selection, which favored summary statistics describing the extremities of feature values (*e.g.*, min, max, P05, P95), which are less appreciable when aggregated over shorter periods. More intriguing are the apparent trends in both biomarkers over the course of a day (Fig. 7B). *Instability* appears to drift upwards from morning to evening ($\hat{X}_1 = -0.14[1.17]/-0.07[1.11]/0.15[1.29]$, Morning / Afternoon / Evening, mean[SD]), supporting a decline in balance confidence throughout the day in PwMS. In contrast, *Fatigue* drifts downwards ($\hat{X}_2 = 0.20[1.00]/0.12[0.92]/-0.17[1.14]$, M/A/E, mean[SD]), which implies fatigue is worst in the morning and improves as the day progresses, perhaps in association with activity.

IV. Discussion

Here we demonstrate the potential efficacy of wearable devices to augment symptom monitoring in PwMS. We propose phenotypes of *Instability* and *Fatigue* formulated using only features of patients' daily standing transitions, an easily distinguished and frequently repeated task [22], [23].

We observe that these metrics demonstrate correlation with traditional clinical measures (Figs. 4 and 5). Unlike PRMs or clinical surveys, these biomarkers require minimal conscious input from the patient, thereby reducing the likelihood of response bias and patient non-adherence. Importantly, the instability index was able to stratify patients by fall status (Figs. 5 and 6) and highlights single-fallers as having a broad range of daily stability (Fig. 6B). This population, with transient motor symptoms and a borderline fall risk, may benefit greatly from continuous monitoring with subsequent forewarning and queued intervention of particularly unstable periods.

Furthermore, both biomarkers demonstrate promise to provide sub-daily resolution of patient symptoms. Though sub-daily aggregation proved worse than daily aggregation at discretion of fall status (Fig. 6 vs. Fig. 7), it's worth noting that the CCA regression model treats each aggregation period as a discrete and independent data point. More robust regression, such as a model with sequential memory, may enhance the temporal resolution and predictive power, but would likely require more training data. Future work may also probe these biomarkers for predictive extrapolation, rather than the reported stability over the window captured by accelerometry. With sub-daily resolution we hypothesize that digital biomarkers could be used to predict stability within the near future, particularly valuable for just-in-time warning systems in PwMS.

Without further data collection and retraining, these phenotypes may still provide previously inaccessible insight into symptoms in PwMS. Much evidence exists to suggest that symptoms in PwMS and particularly in those with Relapsing Remitting MS, are often transient and may arise over the course of hours to days [52], [53], [54]. However, there is a paucity of information describing fluctuations of symptoms over the course of a day. Previous studies, often limited by sample size, study duration, and active assessments, have shown little in the way of observable within-day changes in gait, joint angle variability, or walking capacity [55], [56], [57].

In contrast, we observed a distinct declining trend in patient stability over the course of the day (Fig. 7B). Interestingly, meta-analysis has shown that falls mostly occur in the morning/afternoon (36/39%) rather than in the evening (21%) which our findings would suggest is the time of day at greatest risk [8]. This may further support the importance of context

in considering fall risk. Patients settling down in their home environment fall less often, despite nightly worsening of their pathologic instability.

Daily trends in fatigue have also been previously studied. Our findings predict improvement in fatigue over the course of the day (Fig. 7) in agreement with some studies [57], but contrasts with findings of others that posit fatigue tends to instead worsen as the day wears on [55], [58]. These discrepancies in findings, particularly for a symptom with such high incidence and impact on QoL, suggest underlying effect modifiers. One such factor may be patient activity levels, daily exercise has been shown to improve both fatigue and postural stability in PwMS [59]. Further longitudinal monitoring of these symptoms with wearable devices may help elucidate true patterns in symptoms and may indicate potential therapeutic targets.

These data represent promising support for the application of wearable monitoring for PwMS. We find that digital biomarkers of instability and fatigue are able to match PRMs in repeatable, meaningful assessment of symptom fluctuations in PwMS at daily resolution. Further, with evidence of attainable sub-daily resolution there is potential to extend this technology, perhaps using these biomarkers as a touchstone for calibration and interpretation of less standardized data such as continuous gait analysis. We see these results to be encouraging for the application of wearable technologies in clinical medicine to enable QoL-modifying interventions for PwMS.

REFERENCES

[1] P. A. Calabresi, "Diagnosis and management of multiple sclerosis," *Amer. Family Physician*, vol. 70, no. 10, pp. 1935–1944, 2004.

[2] M. L. B. Ferreira, "Cognitive deficits in multiple sclerosis: A systematic review," *Arquivos de Neuro-Psiquiatria*, vol. 68, no. 4, pp. 632–641, Aug. 2010.

[3] H. L. Zwibel, "Contribution of impaired mobility and general symptoms to the burden of multiple sclerosis," *Adv. Therapy*, vol. 26, no. 12, pp. 1043–1057, Dec. 2009, doi: 10.1007/S12325-009-0082-X.

[4] J. Hobart, "The multiple sclerosis impact scale (MSIS-29): A new patient-based outcome measure," *Brain*, vol. 124, no. 5, pp. 962–973, May 2001, doi: 10.1093/brain/124.5.962.

[5] N. G. LaRocca, "Impact of walking impairment in multiple sclerosis," *Patient, Patient-Centered Outcomes Res.*, vol. 4, no. 3, pp. 189–201, Sep. 2011, doi: 10.2165/11591150-000000000-00000.

[6] Y. Nilsagård, C. Lundholm, E. Denison, and L.-G. Gunnarsson, "Predicting accidental falls in people with multiple sclerosis—A longitudinal study," *Clin. Rehabil.*, vol. 23, no. 3, pp. 259–269, Mar. 2009, doi: 10.1177/0269215508095087.

[7] P. N. Matsuda, A. Shumway-Cook, A. M. Bamer, S. L. Johnson, D. Amtmann, and G. H. Kraft, "Falls in multiple sclerosis," *PM&R*, vol. 3, no. 7, pp. 624–632, Jul. 2011, doi: 10.1016/j.pmrj.2011.04.015.

[8] Y. Nilsagård et al., "Falls in people with MS—An individual data meta-analysis from studies from australia, sweden, united kingdom and the United States," *Multiple Sclerosis J.*, vol. 21, no. 1, pp. 92–100, Jan. 2015, doi: 10.1177/1352458514538884.

[9] S. Tajali et al., "Predicting falls among patients with multiple sclerosis: Comparison of patient-reported outcomes and performance-based measures of lower extremity functions," *Multiple Sclerosis Rel. Disorders*, vol. 17, pp. 69–74, Oct. 2017, doi: 10.1016/j.msard.2017.06.014.

[10] T. J. Braley and R. D. Chervin, "Fatigue in multiple sclerosis: Mechanisms, evaluation, and treatment," *Sleep*, vol. 33, no. 8, pp. 1061–1067, Aug. 2010.

[11] J. F. Kurtzke, "Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS)," *Neurology*, vol. 33, no. 11, pp. 1444–1452, Nov. 1983, doi: 10.1212/wnl.33.11.1444.

[12] M. D. Goldman, R. W. Motl, J. Scagnelli, J. H. Pula, J. J. Sosnoff, and D. Cadavid, "Clinically meaningful performance benchmarks in MS: Timed 25-foot walk and the real world," *Neurology*, vol. 81, no. 21, pp. 1856–1863, Nov. 2013, doi: 10.1212/01.wnl.0000436065.97642.d2.

[13] H. Inojosa, D. Schriefer, and T. Ziemssen, "Clinical outcome measures in multiple sclerosis: A review," *Autoimmunity Rev.*, vol. 19, no. 5, May 2020, Art. no. 102512, doi: 10.1016/j.autrev.2020.102512.

[14] T. Ziemssen, R. Kern, I. Voigt, and R. Haase, "Data collection in multiple sclerosis: The MSDS approach," *Frontiers Neurol.*, vol. 11, p. 445, Jun. 2020.

[15] L. E. Powell and A. M. Myers, "The activities-specific balance confidence (ABC) scale," *J. Gerontol. Ser. A, Biol. Sci. Med. Sci.*, vol. 50A, no. 1, pp. M28–M34, Jan. 1995, doi: 10.1093/gerona/50a.1.m28.

[16] R. D. Larson, "Psychometric properties of the modified fatigue impact scale," *Int. J. MS Care*, vol. 15, no. 1, pp. 15–20, Apr. 2013, doi: 10.7224/1537-2073.2012-019.

[17] M. Sparaco, L. Lavorgna, R. Conforti, G. Tedeschi, and S. Bonavita, "The role of wearable devices in multiple sclerosis," *Multiple Sclerosis Int.*, vol. 2018, pp. 1–7, Oct. 2018, doi: 10.1155/2018/7627643.

[18] J. Dunn, R. Runge, and M. Snyder, "Wearables and the medical revolution," *Personalized Med.*, vol. 15, no. 5, pp. 429–448, Sep. 2018, doi: 10.2217/pme-2018-0044.

[19] L. J. Tulipani, B. Meyer, D. Allen, A. J. Solomon, and R. S. McGinnis, "Evaluation of unsupervised 30-second chair stand test performance assessed by wearable sensors to predict fall status in multiple sclerosis," *Gait Posture*, vol. 94, pp. 19–25, May 2022, doi: 10.1016/j.gaitpost.2022.02.016.

[20] L. J. Tulipani, B. Meyer, D. Larie, A. J. Solomon, and R. S. McGinnis, "Metrics extracted from a single wearable sensor during sit-stand transitions relate to mobility impairment and fall risk in people with multiple sclerosis," *Gait Posture*, vol. 80, pp. 361–366, Jul. 2020, doi: 10.1016/j.gaitpost.2020.06.014.

[21] B. M. Meyer et al., "Wearables and deep learning classify fall risk from gait in multiple sclerosis," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1824–1831, May 2021, doi: 10.1109/JBHI.2020.3025049.

[22] L. Adamowicz et al., "Assessment of sit-to-stand transfers during daily life using an accelerometer on the lower back," *Sensors*, vol. 20, no. 22, p. 6618, Nov. 2020, doi: 10.3390/s20226618.

[23] L. J. Tulipani, B. Meyer, S. Fox, A. J. Solomon, and R. S. Mcginnis, "The sit-to-stand transition as a biomarker for impairment: Comparison of instrumented 30-second chair stand test and daily life transitions in multiple sclerosis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1213–1222, 2022, doi: 10.1109/TNSRE.2022.3169962.

[24] B. M. Meyer et al., "How much data is enough? A reliable methodology to examine long-term wearable data acquisition in gait and postural sway," *Sensors*, vol. 22, no. 18, p. 6982, Sep. 2022, doi: 10.3390/s22186982.

[25] B. M. Meyer et al., "Open-source dataset reveals relationship between walking bout duration and fall risk classification performance in persons with multiple sclerosis," *PLOS Digit. Health*, vol. 1, no. 10, Oct. 2022, Art. no. e0000120, doi: 10.1371/journal.pdig.0000120.

[26] S. Mathias, U. S. Nayak, and B. Isaacs, "Balance in elderly patients: The 'get-up and go' test," *Arch. Phys. Med. Rehabil.*, vol. 67, no. 6, pp. 387–389, Jun. 1986.

[27] R. S. McGinnis and E. W. McGinnis, "Advancing digital medicine with Wearables in the wild," *Sensors*, vol. 22, no. 12, p. 4576, Jun. 2022, doi: 10.3390/s22124576.

[28] Y. C. Learmonth, R. W. Motl, B. M. Sandroff, J. H. Pula, and D. Cadavid, "Validation of patient determined disease steps (PDDS) scale scores in persons with multiple sclerosis," *BMC Neurol.*, vol. 13, no. 1, p. 37, Apr. 2013, doi: 10.1186/1471-2377-13-37.

[29] N. Millor, P. Lecumberri, M. Gomez, A. Martinez-Ramirez, and M. Izquierdo, "Kinematic parameters to evaluate functional performance of sit-to-stand and stand-to-sit transitions using motion sensor devices: A systematic review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 5, pp. 926–936, Sep. 2014, doi: 10.1109/TNSRE.2014.2331895.

[30] M. Mancini, P. Carlson-Kuhta, C. Zampieri, J. G. Nutt, L. Chiari, and F. B. Horak, "Postural sway as a marker of progression in Parkinson's disease: A pilot longitudinal study," *Gait Posture*, vol. 36, no. 3, pp. 471–476, Jul. 2012, doi: 10.1016/j.gaitpost.2012.04.010.

[31] D. Amtmann, A. M. Bamer, V. Noonan, N. Lang, J. Kim, and K. F. Cook, "Comparison of the psychometric properties of two fatigue scales in multiple sclerosis," *Rehabil. Psychol.*, vol. 57, no. 2, pp. 159–166, 2012, doi: 10.1037/a0027890.

[32] Y. Nilsagård, A. Carling, and A. Forsberg, "Activities-specific balance confidence in people with multiple sclerosis," *Multiple Sclerosis Int.*, vol. 2012, pp. 1–8, Oct. 2012, doi: 10.1155/2012/613925.

[33] H. Garg, S. Bush, and E. Gappmaier, "Associations between fatigue and disability, functional mobility, depression, and quality of life in people with multiple sclerosis," *Int. J. MS Care*, vol. 18, no. 2, pp. 71–77, Mar. 2016, doi: 10.7224/1537-2073.2015-013.

[34] C. R. Harris et al., "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, doi: 10.1038/s41586-020-2649-2.

[35] W. McKinney, "Data structures for statistical computing in Python," in *Proc. Python Sci. Conf.*, 2010, pp. 51–56.

[36] P. Virtanen et al., "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.

[37] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 10, pp. 2825–2830, Jul. 2017.

[38] F. Charlier and M. Weber. (2020). *Statannotations*. [Online]. Available: https://github.com/trevismd/statannotations

[39] R. Vallat, "Pingouin: Statistics in Python," *J. Open Source Softw.*, vol. 3, no. 31, p. 1026, Nov. 2018, doi: 10.21105/JOSS.01026.

[40] D. A. Walker, "JMASM9: Converting Kendall's tau for correlational or meta-analytic analyses," *J. Modern Appl. Stat. Methods*, vol. 2, no. 2, pp. 525–530, Nov. 2003, doi: 10.22237/jmasm/1067646360.

[41] C. F. Dormann et al., "Collinearity: A review of methods to deal with it and a simulation study evaluating their performance," *Ecography*, vol. 36, no. 1, pp. 27–46, Jan. 2013, doi: 10.1111/j.1600-0587.2012.07348.x.

[42] J. Wegelin, *A Survey of Partial Least Squares (PLS) Methods, With Emphasis on the Two-Block Case*. Seattle, WA, USA: Univ. Washington, 2000.

[43] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Comput.*, vol. 11, no. 2, pp. 443–482, Feb. 1999, doi: 10.1162/089976699300016728.

[44] M. Gavish and D. L. Donoho, "The optimal hard threshold for singular values is 4/sqrt(3)," 2013, *arXiv:1305.5870*.

[45] S. Jomthanachai, W. P. Wong, and K. W. Khaw, "An application of machine learning regression to feature selection: A study of logistics performance and economic attribute," *Neural Comput. Appl.*, vol. 34, no. 18, pp. 15781–15805, Sep. 2022, doi: 10.1007/s00521-022-07266-6.

[46] J. Fan and R. Li, "Comment: Feature screening and variable selection via iterative ridge regression," *Technometrics*, vol. 62, no. 4, pp. 434–437, Oct. 2020, doi: 10.1080/00401706.2020.1801256.

[47] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its Oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001, doi: 10.1198/016214501753382273.

[48] J. Z. Bakdash and L. R. Marusich, "Repeated measures correlation," *Frontiers Psychol.*, vol. 8, p. 456, Apr. 2017, doi: 10.3389/FPSYG.2017.00456/BIBTEX.

[49] H. Liu, *Comparing Welch's ANOVA, a Kruskal–Wallis Test and Traditional ANOVA in Case of Heterogeneity of Variance*. Richmond, VA, USA: Virginia Commonwealth University, 2015.

[50] A. Dean and D. Voss, Eds., *Design and Analysis of Experiments* (Springer Texts in Statistics). New York, NY, USA: Springer, 2017, doi: 10.1007/0-387-22634-6_5.

[51] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, nos. 3–4, pp. 591–611, Dec. 1965, doi: 10.1093/biomet/52.3-4.591.

[52] F. Andermann, J. B. Cosgrove, D. Lloyd-Smith, and A. M. Walters, "Paroxysmal dysarthria and ataxia in multiple sclerosis; a report of 2 unusual cases," *Neurology*, vol. 9, no. 4, pp. 211–215, Apr. 1959, doi: 10.1212/wnl.9.4.211.

[53] Y. Zhang, S. Fan, F. Han, and Y. Xu, "Paroxysmal symptoms as the first manifestation of multiple sclerosis mimicking a transient ischemic attack: A report of two cases," *Frontiers Neurol.*, vol. 8, p. 585, Nov. 2017, doi: 10.3389/fneur.2017.00585.

[54] J. Freiha, N. Riachi, M. A. Chalah, R. Zoghaib, S. S. Ayache, and R. Ahdab, "Paroxysmal symptoms in multiple sclerosis—A review of the literature," *J. Clin. Med.*, vol. 9, no. 10, p. 3100, Sep. 2020, doi: 10.3390/jcm9103100.

[55] M. E. Morris, "Changes in gait and fatigue from morning to afternoon in people with multiple sclerosis," *J. Neurol., Neurosurgery Psychiatry*, vol. 72, no. 3, pp. 361–365, Mar. 2002, doi: 10.1136/jnnp.72.3.361.

[56] S. J. Crenshaw, T. D. Royer, J. G. Richards, and D. J. Hudson, "Gait variability in people with multiple sclerosis," *Multiple Sclerosis J.*, vol. 12, no. 5, pp. 613–619, Sep. 2006, doi: 10.1177/1352458505070609.

[57] P. Feys et al., "Effect of time of day on walking capacity and self-reported fatigue in persons with multiple sclerosis: A multi-center trial," *Multiple Sclerosis J.*, vol. 18, no. 3, pp. 351–357, Mar. 2012, doi: 10.1177/1352458511419881.

[58] D. J. H. Powell, C. Liossi, W. Schlotz, and R. Moss-Morris, "Tracking daily fatigue fluctuations in multiple sclerosis: Ecological momentary assessment provides unique insights," *J. Behav. Med.*, vol. 40, no. 5, pp. 772–783, Oct. 2017, doi: 10.1007/s10865-017-9840-4.

[59] J. Redlicka, E. Zielińska-Nowak, A. Lipert, and E. Miller, "Impact of moderate individually tailored physical activity in multiple sclerosis patients with fatigue on functional, cognitive, emotional state, and postural stability," *Brain Sci.*, vol. 11, no. 9, p. 1214, Sep. 2021, doi: 10.3390/brainsci11091214.