Assessing Free-Living Postural Sway in Persons with Multiple Sclerosis

Brett Meyer [1], Jenna Cohen [1], Paolo DePetrillo [1], Melissa Ceruolo [1], David Jangraw [1], Nick Cheney [1], Andrew Solomon [1], and Ryan McGinnis [2]

[1]Affiliation not available
[2]University of Vermont

October 31, 2023

Abstract

Assessments of postural sway are associated with disease status and fall risk in Persons with Multiple Sclerosis (PwMS). However, these assessments, which leverage force platforms or wearable accelerometers, are most often conducted in laboratory environments and are thus not broadly accessible. Remote measures of postural sway captured during daily life may provide a more accessible alterative, but their ability to capture disease status and fall risk has not yet been established. We explore the utility of remote measures of postural sway in a sample of 33 PwMS. Remote measures of sway differed significantly from lab-based measures, but still demonstrated moderately strong associations with patient reported measures of balance and mobility impairment. Machine learning models for predicting fall risk trained on lab data provided an AUC of 0.79, while remote data only achieved an AUC of 0.51. Remote model performance improved to an AUC of 0.74 after a new, subject-specific k-means clustering approach was applied for identifying the remote data most appropriate for modelling. This cluster-based approach for analysing remote data also strengthened associations with patient-reported measures, increasing their strength above those observed in the lab. This work introduces a new framework for analysing data from remote patient monitoring technologies and demonstrates the promise of remote postural sway for assessing fall risk and characterizing balance impairment in PwMS.

Assessing Free-Living Postural Sway in Persons with Multiple Sclerosis

Brett M. Meyer, Student Member, IEEE, Jenna G. Cohen, Paolo DePetrillo, Melissa Ceruolo, David Jangraw, Nick Cheney, Andrew J. Solomon, and Ryan S. McGinnis, Senior Member, IEEE

***Abstract*— Assessments of postural sway are associated with disease status and fall risk in Persons with Multiple Sclerosis (PwMS). However, these assessments, which leverage force platforms or wearable accelerometers, are most often conducted in laboratory environments and are thus not broadly accessible. Remote measures of postural sway captured during daily life may provide a more accessible alterative, but their ability to capture disease status and fall risk has not yet been established. We explore the utility of remote measures of postural sway in a sample of 33 PwMS. Remote measures of sway differed significantly from lab-based measures, but still demonstrated moderately strong associations with patient reported measures of balance and mobility impairment. Machine learning models for predicting fall risk trained on lab data provided an AUC of 0.79, while remote data only achieved an AUC of 0.51. Remote model performance improved to an AUC of 0.74 after a new, subject-specific k-means clustering approach was applied for identifying the remote data most appropriate for modelling. This cluster-based approach for analysing remote data also strengthened associations with patient-reported measures, increasing their strength above those observed in the lab. This work introduces a new framework for analysing data from remote patient monitoring technologies and demonstrates the promise of remote postural sway for assessing fall risk and characterizing balance impairment in PwMS.**

***Index Terms*— Postural sway, wearables, digital biomarkers, clinical validation, remote monitoring.**

I. INTRODUCTION

MULTIPLE sclerosis (MS) is an immune mediated disorder leading to demyelination of central nervous system axons that affects an estimated 2.8 million people worldwide [1]. In MS, nerve signals are altered or delayed leading to sensory impairment, motor impairment, fatigue, and postural instability. As a result, an estimated 50-80% of persons with multiple sclerosis (PwMS) have balance and gait dysfunction and over 50% experience a fall in any given 3-month period [2], [3]. This incidence of falls is similar to 80 year-old adults, however, symptoms of MS typically manifest around 30 years-old creating a long-term quality of life and health care burden [4], [5].

Postural instability and balance impairment are typically assessed with subjective patient-reported measures (PRM) [6], non-instrumented balance assessments [7], [8] and/or balance assessments using force platforms [9], [10]. Force platforms are the gold standard for postural sway analysis, which considers objective movement features captured during a period of standing for characterizing balance impairment [9]. Studies utilizing force platforms have been able to distinguish impaired individuals from controls [9] and classify the fall risk of older adults [11] and PwMS [12]. However, force platforms are expensive and limit accessibility to specialized clinics or research laboratories. To address these challenges, studies have shown that postural sway can be assessed using data from just a sacral or chest accelerometer [9], [13]–[15]. Sensor-derived postural sway measures have been used to classify fall risk in PwMS [13], distinguish between disease states [9], [12], [15]–[17] and to augment current assessment techniques [18], thereby achieving similar clinical utility to the force platform.

These promising balance assessments, however, are all performed in clinical or laboratory settings, which limits their accessibility. Recent studies of chest accelerometer-based postural sway have found stronger relationships to PRMs from remotely collected measures compared to in-clinic clinic assessment [15], [19]. These differing relationships to PRMs may be explained by differences between remote and in-clinic measures. Studies comparing remote and in-clinic gait have found that the remote parameters are significantly different and have higher variability compared to those from an in-clinic assessment [20], [21]. As a result, separate models are needed to examine in-clinic and remote gait, but it is not yet clear if these same discrepancies in data exist for postural sway.

Another challenging aspect of remote monitoring is the inherent increase in variability, compared to laboratory measures. This additional variability creates challenges for interpretability and requires additional care to be taken during analysis such that simple averaging of parameters across days or weeks may not be appropriate. One approach is to ask participants to perform repeated prescribed activities throughout the monitoring period

Manuscript received August 11, 2023. Work supported in part by the US NIH under Grant R21EB027852 and by Medidata Systems.

B. M. Meyer, J. G. Cohen, D. Jangraw, A. J. Solomon, N. Cheney, and R. S. McGinnis are with the University of Vermont, Burlington, VT 05405 USA. (e-mails: brett.meyer@uvm.edu, jenna.g.cohen@uvm.edu, ncheney@uvm.edu, david.jangraw@uvm.edu andrew.solomon@uvm.edu, ryan.mcginnis@uvm.edu).

P. DePetrillo and M. Ceruolo are with Medidata Systems, A Dassault Systèmes Company, New York, NY 10014, USA (e-mails: paolo.depetrillo@3ds.com, melissa.ceruolo@3ds.com). R. S. McGinnis is the corresponding author for this work.

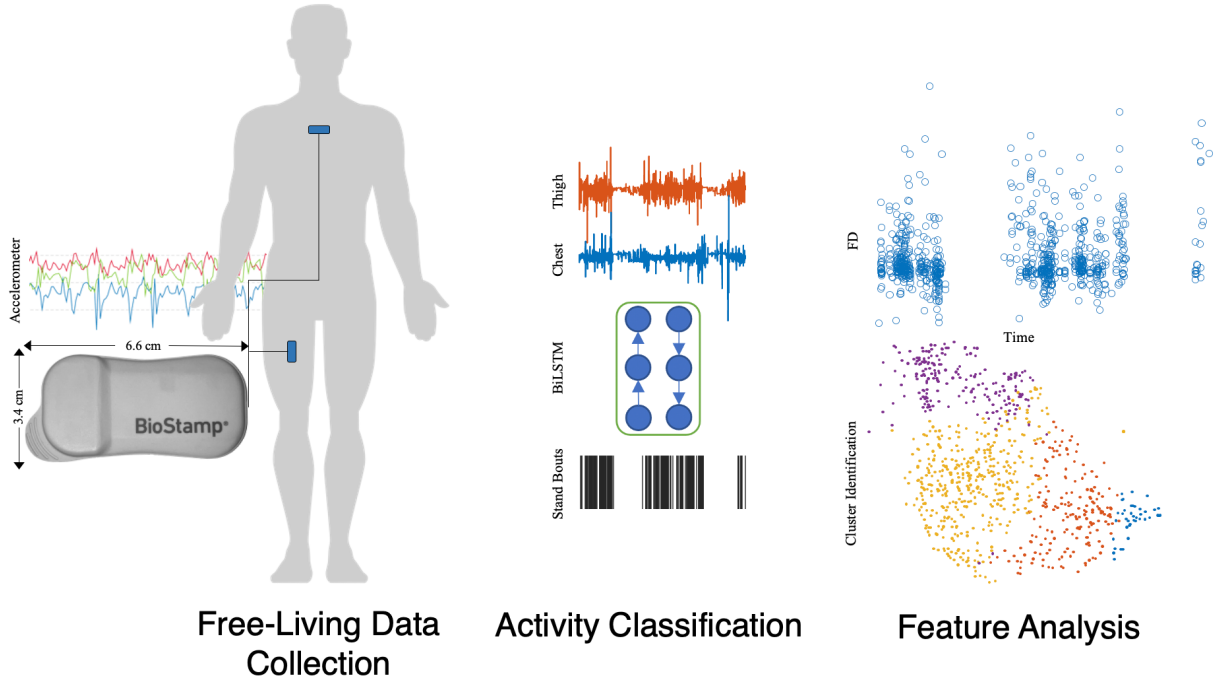

Fig 1. Data processing overview. Free-living data collected from thigh and chest accerleromerter and then classified using a deep learning classifier. Features of postural sway were computed for each standing bout. Feature values vary through the day, clustering techniques were used to find similar data.

to provide consistent context for analysis. For example, this approach has been applied to 30-second chair stand tests [22] and ten-meter walk tests [23], where participants were asked to complete multiple trials remotely. Another approach is to use GPS data to capture measurements in consistent physical locations, again providing context for analysis [24]. While these approaches help control variability, they also reduce the data available for analysis, potentially losing important information in favor of simplified analyses. There may instead be a benefit to pursuing new methods that allow us to select which data to analyze in a fully unsupervised manner, but these approaches have not yet been developed.

II. Methods

To address our goal of evaluating remote postural sway as a biomarker for balance impairment, we consider free-living data from a sample of PwMS (left, Fig. 1). A fully automated data processing pipeline enables detection of standing periods via deep learning-based activity classification (middle, Fig. 1), computation of postural sway parameters that characterize balance performance, and determination of which data are suitable for analysis via unsupervised clustering (right, Fig. 1). Remote postural sway parameters are compared to lab-based measures and PRMs of balance and mobility impairment, and their suitability for classifying fall risk is assessed.

A. Participants and Protocol

To address these objectives, we utilize a dataset of 33 PwMS (16:17 fallers:non-fallers; 10:23 Male:Female, mean ± standard deviation age 50 ± 13 y/o), recruited from the Multiple Sclerosis Center at University of Vermont Medical Center (exclusion: no major health conditions other than MS, no acute exacerbations within the previous three-months, ambulatory without the use of assistive devices). PwMS who self-reported to have fallen within the previous six-months were characterized as fallers based on the criteria "consider a fall as an event where you unintentionally came to rest on the ground or a lower level." Our analysis required a subset of the larger publicly available dataset that has been described in detail in our previous work [21]. Participants were asked to complete several PRMs, several activities of daily living, and a neurologist-administered Expanded Disability Status Scale (EDSS) assessment [25] during a laboratory visit. Participants were then asked to complete a 48-hour daily life monitoring period immediately following the laboratory visit. The PRMs utilized in this analysis were Activities-Specific Balance Confidence (ABC) [6], Multiple Sclerosis Walking Scale (MSWS) [26], and Modified Fatigue Impact Scale (MFIS) [27]. The in-laboratory assessment used in this analysis was a 2-minute standing balance assessment where participants were instructed to stand with their feet shoulder-width apart. The lab and remote assessment periods were instrumented with Biostamp nPoint® (Medidata) sensors (62 Hz ± 16G) located on the chest and thigh. The chest sensor was secured to the sternum, just below the sternoclavicular joint. The thigh sensor was on the anterior aspect of the right thigh, ~25% from the knee to hip.

B. Automated Data Processing

1) **Remote Activity Identification**

Data recorded from both the laboratory and remote sessions were first reoriented to align the cranial-caudal axis with gravity based on the first ten seconds of the lab standing trial. Following calibration, remote data were classified using a previously

described classification framework that identifies bouts of walking, standing, sitting, and lying [19], [28]. Briefly, this model uses a Bidirectional Long-Short-Term Memory Network (BiLSTM) to perform classifications on raw acceleration data from a chest and thigh sensor. This model was trained on a mix of persons with MS, Parkinson's, and healthy adults and provides a 97% accuracy on a held-out test set. This model was used to identify all remote standing bouts that were 30 seconds or longer. 30 seconds was chosen as the minimum because this is the length of the typical in-lab balance assessment [9], [15]. The first minute of the laboratory standing balance assessment was used for in-lab analysis. Data were processed using the individualized distributions [15] approach where a 30-second window is slid 5 samples over the trial to create a distribution of each sway parameter.

2) **Postural Sway Parameter Extraction**

Following the identification of standing periods in both lab and remote data, the acceleration data were down-sampled to 31.25 Hz and a 4th order, zero-phase Butterworth low-pass filter with a cutoff frequency of 3.5 Hz was applied before computing the magnitude of the acceleration in the horizontal plane. Fifteen features were computed for each 30-second lab epoch and/or valid remote standing period. These features included thirteen features from Mancini et al [9]: Jerk, Distance (Dist), Root-Mean-Square (RMS), Path, Range, Mean Velocity (MV), Mean Frequency (MF), Area, Power (Pwr), median power frequency (F50), 95% power frequency (F95), Centroidal Frequency (CF), and Frequency Dispersion (FD). We also considered two features that capture signal complexity: Approximate Entropy (ApEn) [29], and Lyapunov Exponent (LyExp) [16], [30].

3) **Data Clustering Methodology**

Unsupervised clustering (k-means, [31], [32]) was used to discover underlying structure in the remote data that may arise from participants performing other activities while standing such as washing dishes, standing in line, etc. Similar methods have been used to cluster symptoms in PwMS to increase predictability of physical activity [33]. The optimal number of clusters for each participant was chosen using MATLAB's evalclusters function with DaviesBouldin [34] criterion and Euclidean distance for 1 to 5 clusters yielding a mean of 4 clusters across participants. The z-scores of the reduced feature set were then used to identify four clusters for each participant. The clusters were labeled based on the sorted centroid of the FD feature, because it is strongly related to impairment. Cluster 1 has the highest centroid of FD, and cluster 4 has the lowest.

C. Statistical Analysis

Ranksum difference tests were used to identify differences between lab and remote postural sway features. Effect sizes were characterized with Cohen's D. Median and inter-quartile range (IQR) are also reported for each feature. Spearman correlations were used to identify significant associations between the postural sway features and PRMs. The multiple observations of features per participant were aggregated using the 5th percentile (P5), 25th percentile (P25), median (Med), 75th percentile (P75), 95th percentile (P95), and standard deviation (STD). The results of the strongest significant aggregation were reported. Using the results of the correlations to PRMs and previously published [15] cross-correlations between the features, we then selected a reduced feature-set for remote analysis containing features that demonstrate correlations to PRMs and that are not highly correlated to each other. When choosing between highly related features, the feature with the strongest remote PRM correlation was chosen. Details of feature correlations are provided in the results. The reduced remote feature set contained RMS, Range, Area, CF, FD, and LyExp.

To examine clinical significance, we computed Spearman correlations to both ABC and EDSS for the sway features from each cluster and compared the remote data features to lab data features with a Rank-sum test. Due to the reduced amount of data per participant, correlations were performed using the raw feature values instead of summary statistics. Additionally, we computed the Spearman correlation of the lab features with clustered features for the median and 95th percentile of each feature. For comparison, these same methods have been applied to the non-clustered data.

D. Fall Risk Classification

Six-month fall history was used to inform classification models for discriminating fallers from non-fallers. Logistic regression (LR) and support vector machine (SVM) models were trained, optimized, and tested separately on the lab features, all remote features, and remote features from each cluster. Leave-One-Subject-Out cross-validation (LOSOCV) was performed to ensure data from participants was not in both the test and training set. Performance was assessed using area under the receiver operating characteristic curve (AUC), accuracy (acc), sensitivity (sens), specificity (spec), and F1 score. Model performance was computed using both the outputs from each individual input and by aggregating the median decision score from each observation of an individual participant, resulting in one prediction per participant. Model hyperparameters were tuned. Lasso regularization was used with OptimizeLearnRate to train the LR models and the SVM model was found to perform best with a linear kernel and a SMO solver. A permutation analysis was conducted to compare the model AUC against random chance, using 100 run average of classification results compared to 1000 replicates of permuted labels.

III. Results

All fifteen sway features computed were significantly different between the lab and all remote data (Table I). Very high effect sizes were also observed for Pwr, Path, and RMS. When correlating the remote data to PRMs (Table II), we found the strongest relationships across all PRMs with FD (frequency dispersion). The PRMs ABC and MFIS demonstrated the most significant correlations to remote sway, however, the strongest relationship observed, r = -0.62, was the 75th percentile of FD with MSWS.

TABLE I
DIFFERENCE BETWEEN REMOTE AND LAB

Feature	Lab Median	Lab IQR	Remote Median	Remote IQR	P-value	Cohen's D
Jerk	0.043	0.041	0.103	0.132	<0.001	0.838
Dist	0.001	0.002	0.006	0.011	<0.001	0.904
RMS	0.045	0.034	0.709	0.118	<0.001	8.50
Path	1.39	1.071	22.2	3.749	<0.001	8.37
Range	0.066	0.054	0.392	0.353	<0.001	1.45
MV	1.24	1.127	21.6	4.015	<0.001	7.51
MF	195	548.7	541	2995	<0.001	0.63
Area	0.018	0.029	4.84	3.213	<0.001	2.20
Pwr	0.002	0.033	0.503	0.160	<0.001	4.64
F50	0.081	0.033	0.080	0.0002	<0.001	1.48
F95	0.225	0.026	0.217	0.003	<0.001	1.35
CF	0.701	0.057	0.715	0.011	<0.001	0.711
FD	1.85	0.251	1.788	0.035	<0.001	1.00
ApEn	0.515	0.178	0.343	0.282	<0.001	1.05
LyExp	-0.457	14.74	-2.54	5.464	<0.001	1.13

Median and interquartile range (IQR) testing of lab and remote sway metrics with rank-sum difference testing (uncorrected) and Cohen's D effect size (α = 0.05)

Measures of in-lab postural sway were found to provide strong fall risk classification results. Using a LR classifier, we observed an AUC of 0.74 when classifying fall risk based on single observations, which increased to 0.79 when we used the median of each participant's decision scores as a summary predictor ($p < 0.001$ and $p = 0.003$, respectively). The weights of the model, which we use as a proxy for a feature importance, are depicted in Fig 2. To establish a baseline remote data performance, we first fit models using all the collected data (i.e., without clustering). The best performing model was a logistic regression model, with an AUC of 0.52 before aggregation and 0.44 after, suggesting that the model is unable to perform any better than guessing likely due to noisy data ($p = 0.034$ and $p = 0.417$, respectively). More model performance details can be found in Table IV.

TABLE II
ASSOCIATION WITH PATIENT REPORTED MEASURES

Feature	EDSS		ABC		MFIS		MSWS	
Jerk	-	-	-	-	-	-	-	-
Dist	-	-	-	-	-	-	-	-
RMS	-	-	-	-	P75	0.34	-	-
Path	-	-	-	-	P75	0.32	-	-
Range	-	-	STD	0.36	-	-	-	-
MV	-	-	-	-	Med	0.33	-	-
MF	-	-	-	-	-	-	-	-
Area	-	-	P75	0.39	*STD*	*0.45*	STD	0.30
Pwr	-	-	-	-	P75	0.34	-	-
F50	-	-	-	-	-	-	-	-
F95	-	-	-	-	-	-	-	-
CF	P75	0.38	Med	-0.51	Med	0.38	Med	0.49
FD	*P75*	*-0.59*	*P75*	*0.56*	P75	-0.40	*P75*	*-0.62*
ApEn	-	-	P5	0.31	P5	-0.33	P5	-0.32
LyExp	-	-	STD	0.38	-	-	-	-

Metric showing the strongest correlations, and the strength of that correlation, between sway features derived from all remote data and patient-reported measures. P5: 5th Percentile; Med: Median; P75: 75th Percentile; STD: Standard Deviation; Strongest correlation for each measure in italics. Comparisons where no metrics met a weak significance criterion (α = 0.10, uncorrected) are omitted and replaced with a dash.

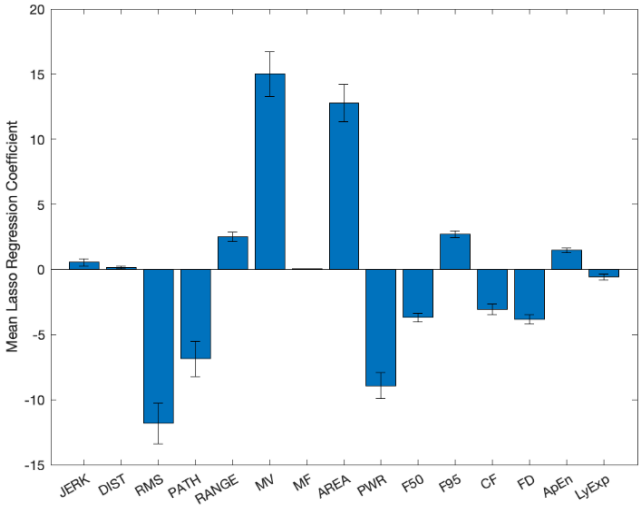

Fig 2. Feature importance of logistic regression model for in-lab fall risk classification.

Clustering methods were then applied to investigate whether selecting subsets of data would enhance performance. An average of four clusters was found to be optimal across participants (see Fig. 3 for optimal clusters by participant). Each cluster was found to have unique relationships to PRMs. As seen in Table III, the strongest correlation to ABC was observed with FD from cluster 2, however, the strongest correlation to EDSS was observed with RMS from cluster 3. Overall, features from clusters 1-3 all establish meaningful correlations to PRMs while cluster 4 does not.

When compared to lab data, all features were different between the clusters and lab data except ApEn for cluster 1 and ApEn, CF, and FD for cluster 2. Only clusters 2 and 3 have significant correlations between lab-derived and clustered remote features. The lab-derived RMS feature has a correlation of 0.36 and 0.43 with the median RMS of clusters 2 and 3, respectively. The lab-derived CF feature was also significantly anticorrelated with the CF of cluster 2 ($r = -0.46$). Interestingly, while not strongly correlated, the features Range and FD showed negative correlations between the lab data and all remote data, and between the lab data and clusters 1-3. All other features, including those from all home data, were not significantly correlated to lab-derived features.

Fig. 4 demonstrates the z-score differences between the clusters and all of the home data. Here we find Range and FD are both higher and CF is lower in clusters 1 and 2 compared to all home data.

When averaging the amount of time spent in each cluster across all participants, fallers spent 12.05% of the time in cluster 1, 25.94% in cluster 2, 42.83% in cluster 3, and 19.18% in cluster 4. Non-fallers spent 7.69% in cluster 1, 24.35% in cluster 2, 38.56% in cluster 3, and 29.43% in cluster 4. Differences in time spent in clusters 1 and 4 between fallers and non-fallers approaches significance, ($p = 0.055$, $p = 0.050$, respectively), but this was not the case for clusters 2 and 3.

Training models to classify fall risk from the different clusters of data revealed vastly difference performance between clusters. Considering the aggregation of 48 hours of data,

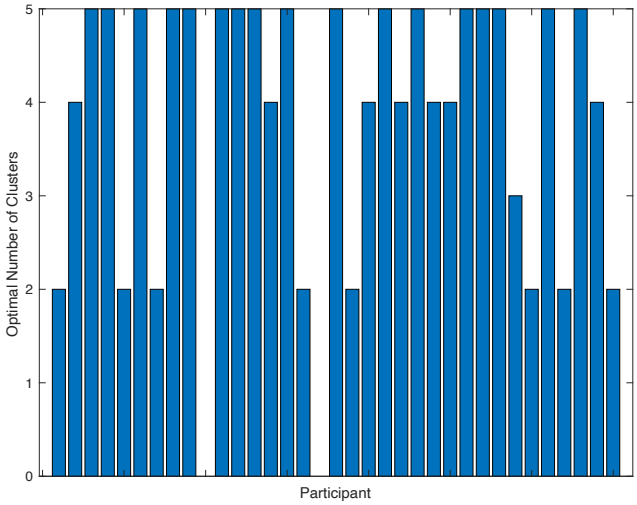

Fig 3. Optimal number of clusters for each participant for k-means clustering.

clusters 1-4 achieve AUCs of 0.57, 0.71, 0.53, and 0.32 respectively, as shown in Table IV. SVM models were found to perform best for clusters 1-3, while a logistic regression model provided the best performance for cluster 4. Overall, cluster 2 exhibits the strongest fall classification performance. Cluster 1 has a strong unaggregated performance, AUC 0.73 with the highest observed accuracy, sensitivity, and F1 score, however cluster 1 has a strong class imbalance toward fallers, which is corrected for by aggregation resulting in the AUC of 0.57. Details regarding class balances, additional model performance measures, and significance tests for model results are provided in Table IV.

IV. Discussion

The purpose of this work was to introduce postural sway as a remote digital biomarker. In doing so, we compared postural sway features from free-living data to those computed from a lab standing assessment, computed correlations to PRMs, and trained fall classification models to establish clinical significance. In these analyses, we explored the impact of selecting subsets of data by clustering compared to considering all free-living data.

When comparing lab and remotely collected postural sway, all features were found to be significantly different with larger IQRs observed in remote data in many cases. Interestingly, many of the features with high effect sizes were related to sway path and power (e.g., Path, RMS, Pwr), suggesting that perhaps sway patterns are more variable at home. These findings suggest that modeling approaches need to be trained using data from the targeted use environment. Similar observations were made in remotely collected gait in PwMS [21].

Our investigation of clinical significance finds several significant correlations between PRMs and remote sway features. The FD feature provided the strongest correlations with EDSS, ABC, and MSWS, while Area provided the strongest correlation with MFIS. In our previous studies, we have found few significant relationships between standard eyes-open standing and PRMs in the lab [15]. The strongest in-lab correlation we observed was r=-0.37 between Dist and MFIS. In this analysis, we not only find a stronger relationship between MFIS and Area (r=0.45). We also find a correlation of r=-0.62 between FD and MSWS when considering remotely collected sway. Based on these findings, remote sway parameters are clinically relevant because they show relationships with patient reported measures of impairment.

Using the remote and lab measures to train fall classification models, however, we find that considering all the remote data is highly variable and noisy when trying to classify fall risk, highlighting the need for some level of preprocessing such as clustering. The in-lab features were able to achieve an AUC of 0.79 in an eyes-open balance assessment compared to 0.52 with remote features. Investigating the feature importance of the lab model reveals the most important features from this set in the lab are RMS, Path, MV, Area, and Pwr. These findings are different from those previously found in MS that suggest the three domains to explain balance variance are sway amplitude and velocity and sway frequency and jerk in the anterior-posterior (AP) and medial-lateral (ML) directions [35]. These features may not arise as important in this analysis because we do not separate into AP and ML features.

When clustering methods were applied to the remote data, we found differing relationships with each cluster. Based on rank-sum tests, cluster 2 provided the fewest significant differences when compared to the lab standing, followed by cluster 1. All features were significantly different for clusters 3 and 4. When correlating these home and clustered features with the lab-derived features, we find most features are not correlated, meaning lab performance is not indicative of real world standing. Additionally, the features Range, FD, and CF had negative correlations between lab and remote data, suggesting that those who have less sway in the lab assessment have larger sway ranges at home. This may reflect an increase in confidence and movement in those who are less impaired. We also found the overall highest correlations when using data from cluster 2 between ABC and FD (r=0.64).

TABLE III
Association with Patient Reported Measures by Cluster

All Data						
Feature	RMS	Range	Area	CF	FD	LyExp
EDSS	-0.38	-0.14	-0.13	0.18	-0.26	0.14
ABC	0.24	0.25	-	-0.21	0.25	-0.07
Cluster 1						
EDSS	-0.18	-0.22	-	0.24	-0.38	-0.21
ABC	-	0.51	-0.21	-0.57	0.55	0.33
Cluster 2						
EDSS	-0.30	-0.25	-	0.41	-0.55	0.10
ABC	0.20	0.40	-	-0.55	0.64	-0.12
Cluster 3						
EDSS	-0.61	-0.19	-0.38	0.22	-0.27	0.30
ABC	0.42	0.33	0.14	-0.21	0.20	-0.16
Cluster 4						
EDSS	-	-	0.17	0.17	-0.16	0.11
ABC	-0.12	-	-0.22	-0.19	0.19	-

EDSS: Expanded Disability Severity Scale; ABC: Activities-Specific Balance Confidence. Spearman correlation between postural sway features from each cluster and patient reported measures. (α = 0.05).

TABLE IV
FALL RISK CLASSIFICATION PERFORMANCE BY CLUSTER

Data	Model	Input Size	AGG	ACC	SPE	SEN	AUC	F1	p-val
Lab	LR	F: 3008	None	0.69	0.74	0.64	0.74	0.67	**<0.001**
		NF: 3196	Med	0.76	0.82	*0.69*	**0.79**	0.69	**0.003**
All Home	LR	F: 2337	None	0.62	0.39	0.62	0.52	0.62	**0.034**
		NF: 2308	Med	0.45	0.65	0.56	0.44	0.50	0.417
C1	SVM	F: 200	None	**0.77**	0.55	**0.87**	0.73	**0.84**	**<0.001**
		NF: 93	Med	0.60	0.57	0.63	0.57	0.63	0.299
C2	SVM	F: 643	None	0.69	0.69	0.69	0.72	0.71	**<0.001**
		NF: 555	Med	0.73	***0.86***	0.63	0.71	0.71	**0.018**
C3	SVM	F: 1065	None	0.58	0.73	0.43	0.53	0.51	**0.003**
		NF: 998	Med	0.53	0.79	0.31	0.53	0.42	0.111
C4	LR	F: 495	None	0.50	0.44	0.58	0.31	0.50	**<0.001**
		NF: 658	Med	0.40	0.29	0.50	0.32	0.47	0.151

C1 – C4: Cluster 1 – Cluster 4; LR: Logistic Regression; SVM: Support Vector Machine; AGG: Aggregation; Med: Median; ACC: Accuracy; SPE: Specificity; SEN: Sensitivity; AUC: Area Under Curve. p-val: p-values of permutation test. Bold values signify the highest performance in column or significance of P-values (α = 0.05, uncorrected). Note: results tested to be significantly different from random, results lower than AUC of 0.5 are tested for a significant performance reduction compared to random.

Interestingly, however, the strongest correlation to EDSS was found between RMS in cluster 3 (r=0.61). This suggests that the clusters may capture different relationships within the data.

Perhaps the most interesting finding is that when the data from the clusters were used to train fall risk models, cluster 2 was able to achieve performance near that of the lab assessment. These clustering results show promise that accurate assessments can still be made with remote data when appropriate data are selected for analysis. Herein, the clustering was simply used as a method to select different sets of unique data. The improved performance and correlations observed when doing so motivates using similar unsupervised methods to remove unwanted data or select data of interest in future remote analyses. For example, a similar approach may have been able to explain the differences between fall classification performance of PwMS from gait from the lab and home [21], [36].

There are some limitations to this study. First, our analysis was based on a relatively small sample of 33 PwMS. There is a lack of demographic and regional diversity in this sample. Additionally, our methods do not distinguish between AP and ML direction features, which may impact the resulting conclusions and their agreement with prior work. Finally, our inclusion criteria limit these studies to lower impairment individuals.

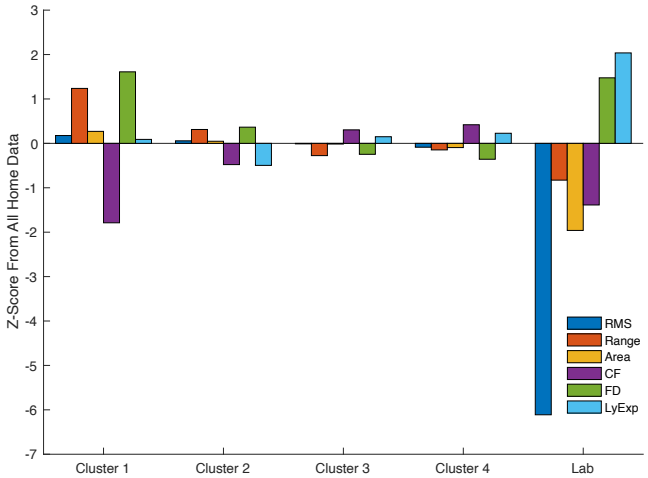

Fig 4. Z-score differences for each feature between that derived from all remote data and that derived from clustered remote data or lab data.

We expect a more impaired sample would result in stronger signals, however, that remains untested with our current dataset.

Despite these limitations, we were still able to provide strong results and motivations for the remote assessment of postural sway. Future studies need to be done to determine if these same clustering methods can be applied to improve deep learning classification of fall risk. Studies should also be done to determine if similar clustering methods provide meaningful findings in other activities measured remotely, such as gait.

V. CONCLUSION

Herein we examined the use of postural sway as a remote digital biomarker. We demonstrated that sway measures collected in the lab are significantly different from those collected remotely and that remote data demonstrated stronger correlations with PRMs. However, lab sway features were able to accurately assess fall risk while unclustered remote measures were unable to do so. To address this, we applied a clustering method to identify similar data at home and found differing relationships to PRMs and fall risk within each cluster. The best performing cluster was able to achieve similar performance to lab collected sway and provided stronger correlations than both the lab and all home data. Our results motivate the inclusion of postural sway as an analysis method in future remote studies.

REFERENCES

[1] C. Walton *et al.*, "Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition," *Mult. Scler. Houndmills Basingstoke Engl.*, vol. 26, no. 14, pp. 1816–1821, Dec. 2020, doi: 10.1177/1352458520970841.

[2] M. H. Cameron and Y. Nilsagard, "Balance, gait, and falls in multiple sclerosis," *Handb. Clin. Neurol.*, vol. 159, pp. 237–250, 2018, doi: 10.1016/B978-0-444-63916-5.00015-X.

[3] S. Coote, L. Comber, G. Quinn, C. Santoyo-Medina, A. Kalron, and H. Gunn, "Falls in People with Multiple Sclerosis," *Int. J. MS Care*, vol. 22, no. 6, Art. no. 6, 2020, doi: 10.7224/1537-2073.2020-014.

[4] S. M. Rao, *Neurobehavioral Aspects of Multiple Sclerosis*. Oxford University Press, 1990.

[5] J. Dunn, "Impact of mobility impairment on the burden of caregiving in individuals with multiple sclerosis," *Expert Rev. Pharmacoecon. Outcomes Res.*, vol. 10, no. 4, Art. no. 4, Aug. 2010, doi: 10.1586/erp.10.34.

[6] L. E. Powell and A. M. Myers, "The Activities-specific Balance Confidence (ABC) Scale," *J. Gerontol. Ser. A*, vol. 50A, no. 1, pp. M28–M34, Jan. 1995, doi: 10.1093/gerona/50A.1.M28.

[7] "Berg Balance Scale," *Shirley Ryan AbilityLab*. https://www.sralab.org/rehabilitation-measures/berg-balance-scale (accessed Dec. 13, 2022).

[8] F. B. Horak, D. M. Wrisley, and J. Frank, "The Balance Evaluation Systems Test (BESTest) to Differentiate Balance Deficits," *Phys. Ther.*, vol. 89, no. 5, pp. 484–498, May 2009, doi: 10.2522/ptj.20080071.

[9] M. Mancini *et al.*, "ISway: a sensitive, valid and reliable measure of postural control," *J. NeuroEngineering Rehabil.*, vol. 9, no. 1, p. 59, 2012, doi: 10.1186/1743-0003-9-59.

[10] B. Chen, P. Liu, F. Xiao, Z. Liu, and Y. Wang, "Review of the Upright Balance Assessment Based on the Force Plate," *Int. J. Environ. Res. Public. Health*, vol. 18, no. 5, Art. no. 5, Jan. 2021, doi: 10.3390/ijerph18052696.

[11] J. Swanenburg, E. D. de Bruin, D. Uebelhart, and T. Mulder, "Falls prediction in elderly people: a 1-year prospective study," *Gait Posture*, vol. 31, no. 3, pp. 317–321, Mar. 2010, doi: 10.1016/j.gaitpost.2009.11.013.

[12] R. Sun, K. L. Hsieh, and J. J. Sosnoff, "Fall Risk Prediction in Multiple Sclerosis Using Postural Sway Measures: A Machine Learning Approach," *Sci. Rep.*, vol. 9, no. 1, p. 16154, Dec. 2019, doi: 10.1038/s41598-019-52697-2.

[13] R. Sun *et al.*, "Assessment of Postural Sway in Individuals with Multiple Sclerosis Using a Novel Wearable Inertial Sensor," *Digit. Biomark.*, vol. 2, no. 1, pp. 1–10, 2018, doi: 10.1159/000485958.

[14] A. J. Solomon, J. V. Jacobs, K. V. Lomond, and S. M. Henry, "Detection of postural sway abnormalities by wireless inertial sensors in minimally disabled patients with multiple sclerosis: a case-control study," *J. Neuroengineering Rehabil.*, vol. 12, p. 74, Sep. 2015, doi: 10.1186/s12984-015-0066-9.

[15] B. M. Meyer *et al.*, "Chest-Based Wearables and Individualized Distributions for Assessing Postural Sway in Persons with Multiple Sclerosis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, pp. 1–1, 2023, doi: 10.1109/TNSRE.2023.3267807.

[16] A. Hadamus *et al.*, "Nonlinear and Linear Measures in the Differentiation of Postural Control in Patients after Total Hip or Knee Replacement and Healthy Controls," *Diagnostics*, vol. 12, no. 7, Art. no. 7, Jul. 2022, doi: 10.3390/diagnostics12071595.

[17] P. G. Monaghan, A. S. Monaghan, A. Hooyman, B. W. Fling, J. M. Huisinga, and D. S. Peterson, "Utilizing the ISway to Identify and Compare Balance Domain Deficits in People with Multiple Sclerosis," *Arch. Phys. Med. Rehabil.*, pp. S0003-9993(23)00153–3, Apr. 2023, doi: 10.1016/j.apmr.2023.02.018.

[18] I. Carpinella *et al.*, "Balance Impairments in People with Early-Stage Multiple Sclerosis: Boosting the Integration of Instrumented Assessment in Clinical Practice," *Sensors*, vol. 22, no. 23, p. 9558, Dec. 2022, doi: 10.3390/s22239558.

[19] 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, Art. no. 18, Sep. 2022, doi: 10.3390/s22186982.

[20] F. A. Storm, K. P. S. Nair, A. J. Clarke, J. M. Van der Meulen, and C. Mazzà, "Free-living and laboratory gait characteristics in patients with multiple sclerosis," *PloS One*, vol. 13, no. 5, Art. no. 5, 2018, doi: 10.1371/journal.pone.0196463.

[21] 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, Art. no. 10, Oct. 2022, doi: 10.1371/journal.pdig.0000120.

[22] 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.

[23] M. Ullrich *et al.*, "Detection of Unsupervised Standardized Gait Tests From Real-World Inertial Sensor Data in Parkinson's Disease," *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. PP, Oct. 2021, doi: 10.1109/TNSRE.2021.3119390.

[24] W. Wang and P. G. Adamczyk, "Analyzing Gait in the Real World Using Wearable Movement Sensors and Frequently Repeated Movement Paths," *Sensors*, vol. 19, no. 8, Art. no. 8, Jan. 2019, doi: 10.3390/s19081925.

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

[26] J. C. Hobart, A. Riazi, D. L. Lamping, R. Fitzpatrick, and A. J. Thompson, "Measuring the impact of MS on walking ability: The 12-Item MS Walking Scale (MSWS-12)," *Neurology*, vol. 60, no. 1, Art. no. 1, Jan. 2003, doi: 10.1212/WNL.60.1.31.

[27] "Modified Fatigue Impact Scale," *Shirley Ryan AbilityLab*, Jun. 16, 2020. https://www.sralab.org/rehabilitation-measures/modified-fatigue-impact-scale (accessed Jun. 16, 2020).

[28] 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. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 30, pp. 1213–1222, 2022, doi: 10.1109/TNSRE.2022.3169962.

[29] A. M. Sabatini, "Analysis of postural sway using entropy measures of signal complexity," *Med. Biol. Eng. Comput.*, vol. 38, no. 6, pp. 617–624, Nov. 2000, doi: 10.1007/BF02344866.

[30] J. Kędziorek and M. Błażkiewicz, "Nonlinear Measures to Evaluate Upright Postural Stability: A Systematic Review," *Entropy*, vol. 22, no. 12, Art. no. 12, Dec. 2020, doi: 10.3390/e22121357.

[31] X. Jin and J. Han, "K-Means Clustering," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds., Boston, MA: Springer US, 2010, pp. 563–564. doi: 10.1007/978-0-387-30164-8_425.

[32] J. MacQueen, "Classification and analysis of multivariate observations," in *5th Berkeley Symp. Math. Statist. Probability*, University of California Los Angeles LA USA, 1967, pp. 281–297.

[33] R. W. Motl and E. McAuley, "Symptom Cluster as a Predictor of Physical Activity in Multiple Sclerosis: Preliminary Evidence," *J. Pain Symptom Manage.*, vol. 38, no. 2, pp. 270–280, Aug. 2009, doi: 10.1016/j.jpainsymman.2008.08.004.

[34] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, Art. no. 2, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.

[35] P. G. Monaghan, A. S. Monaghan, A. Hooyman, B. W. Fling, J. M. Huisinga, and D. S. Peterson, "Using the Instrumented Sway System (ISway) to Identify and Compare Balance Domain Deficits in People With Multiple Sclerosis," *Arch. Phys. Med. Rehabil.*, vol. 0, no. 0, Apr. 2023, doi: 10.1016/j.apmr.2023.02.018.

[36] B. M. Meyer *et al.*, "Wearables and Deep Learning Classify Fall Risk from Gait in Multiple Sclerosis," *IEEE J. Biomed. Health Inform.*, pp. 1–1, 2020, doi: 10.1109/JBHI.2020.3025049.