2023

Applications of a Wearable Sensor Platform for Remotely Monitoring Impairment in Persons with Multiple Sclerosis

Brett Matthew Meyer
University of Vermont

APPLICATIONS OF A WEARABLE SENSOR PLATFORM FOR REMOTELY MONITORING IMPAIRMENT IN PERSONS WITH MULTIPLE SCLEROSIS

A Dissertation Presented

by

Brett M. Meyer

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Complex Systems and Data Science

August, 2023

Defense Date: July 07, 2023
Dissertation Examination Committee:

Ryan S. McGinnis, Ph.D., Advisor
Nick Cheney, Ph.D., Co-Advisor
Kimberly R. Bauerly, Ph.D., Chairperson
Chris Danforth, Ph.D.
David Jangraw, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College

ABSTRACT

Multiple sclerosis is an immune-mediated disease of the central nervous system that commonly results in impaired sensory function, balance, coordination, and fatigue. Persons with multiple sclerosis (PwMS) experience high rates of falls, with over half experiencing a fall in any three-month period. Presently, fall risk is assessed at biannual office visits, but symptoms are known to fluctuate and it is not clear that these assessments provide an adequate picture of a patient's fall risk. Remote monitoring of a patient's balance and mobility with wearable sensors may provide a way to better characterize fall risk, but technologies for making these measurements are just emerging. The purpose of this work was to advance the state of science in remote monitoring of balance and mobility, and to use the resulting technology to identify potential predictors of fall risk in PwMS.

The primary technical contribution of this work was a data analysis platform that allows for remote characterization of balance and mobility impairment. The platform detects walking and standing bouts from free-living wearable accelerometer data and computes metrics that describe how patients are engaging in these balance-challenging activities. This platform was leveraged to examine data from two cohorts of PwMS.

First, data from the platform were used to better understand the relationship between walking bout duration and measures that describe how a patient is walking. Walking metrics were significantly different between bouts of differing lengths, and between walking bouts observed in and out of the clinic. Long remote bouts were the closest to in-clinic measurements and were best able to identify PwMS at higher risk for falls using deep learning models. Interestingly, short remote bouts were best when using more traditional machine learning techniques.

Data from the platform were then used to investigate how much data is enough for capturing valid measures of balance and mobility impairment remotely. Analysis revealed only two days of data are needed to capture most measures of gait and postural sway in our cohorts of PwMS. In general, minimum wear duration was predicted by the daily variability of a measurement and number of daily observations.

Finally, data from the platform were used to further establish remote postural sway, measured by chest-worn accelerometer, as a digital endpoint for balance impairment. Chest-derived measures of sway were validated relative to gold-standard force platforms. A new analysis approach, which builds individualized distributions of each postural sway measure, was introduced that increased accuracy for classifying PwMS' risk for falls and the strength of associations with patient-reported measures (PRMs) of balance impairment. Remote measures of sway differed from these lab measures but had stronger associations to PRMs. A patient-specific clustering approach for analyzing remote sway further strengthened associations and enabled detection of PwMS at higher risk for falls.

Overall, this body of work addresses several key challenges of remote wearable sensor data analysis and introduces remote postural sway as a novel digital endpoint for balance impairment.

CITATIONS

Material from this dissertation has been published in the following form:

Meyer, B. M., Tulipani, L. J., Gurchiek, R. D., Allen, D. A., Solomon, A. J., Cheney, N., & McGinnis, R. S.. (2022). Open-source dataset reveals relationship between walking bout duration and fall risk classification performance in persons with multiple sclerosis. PLOS Digital Health, 1(10), Article 10. https://doi.org/10.1371/journal.pdig.0000120

Meyer, B. M., Depetrillo, P., Franco, J., Donahue, N., Fox, S. R., O'Leary, A., Loftness, B. C., Gurchiek, R. D., Buckley, M., Solomon, A. J., Ng, S. K., Cheney, N., Ceruolo, M., & McGinnis, R. S.. (2022). How Much Data Is Enough? A Reliable Methodology to Examine Long-Term Wearable Data Acquisition in Gait and Postural Sway. Sensors (Basel, Switzerland), 22(18), Article 18. https://doi.org/10.3390/s22186982

Meyer, B. M., Cohen, J. G., Donahue, N., Fox, S. R., O'Leary, A., Brown, A. J., Leahy, C., VanDyk, T., DePetrillo, P., Ceruolo, M., Cheney, N., Solomon, A. J., & McGinnis, R. S.. (2023). Chest-Based Wearables and Individualized Distributions for Assessing Postural Sway in Persons with Multiple Sclerosis. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 1–1. https://doi.org/10.1109/TNSRE.2023.3267807

AND/OR

Material from this dissertation has been submitted for publication to IEEE Journal of Biomedical and Health Informatics on August 9, 2023 in the following form:

Meyer, B. M., Cohen, J. G., DePetrillo, P., Ceruolo, M., Cheney, N., Solomon, A. J., & McGinnis, R. S.. (2023). Assessing Free-Living Postural Sway in Persons with Multiple Sclerosis

ACKNOWLEDGEMENTS

I would like to thank all my committee members for helping to shape this dissertation and bring it to life. Specifically, I would like to thank Dr. Ryan McGinnis for an immeasurable amount of help, guidance, and wisdom throughout our many years working together. Working with you beginning after freshman year through a dissertation has truly shaped my skills as an engineer, researcher, and data scientist and my growth as a person. I would like to thank Dr. Nick Cheney for sparking my interest in modeling and guiding me through the development and learning process. Exposure to your lab's work continually shows me how we can push the envelope with our work.

I would also like to thank Dr. Reed Gurchiek for helping shape my skills as an engineer and problem solver. You are truly inspirational, and I know will continue to do great things. I would like to thank Dr. Lindsey Tulipani for collecting and organizing a large amount of the data used in this dissertation with me. Additionally, I would like to thank all the undergraduates that participated in data collections and analysis, and specifically Dakota Allen.

Finally, I would like to thank my wife, Alyssa. You have supported me through this journey and provided much needed encouragement. This would not have been possible without you, so again, thank you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1: WEARABLES AND FALL RISK IN PERSONS WITH MULTIPLE SCLEROSIS

1.1. About Multiple Sclerosis

Multiple Sclerosis (MS) is characterized by progressive demyelination and axonal damage throughout the central nervous system [1], [2]. As a result, persons with MS (PwMS) experience in symptoms including debilitating fatigue and impaired coordination, muscle strength, and sensation [1]. Collectively, these symptoms lead to problems with balance and postural control, especially during dynamic activities, as well as altered movement and gait patterns [3]. MS is estimated to impact over 2.3 million people globally with a typical symptom onset of 30 years old.

1.2. Falls in MS

Despite the young age of symptom onset, over half of PwMS experience a fall in any given 3-month period, similar incidence to 80 year-old adults [4]. Over half of these falls result in injury, which increases fear of falling and decreases quality of life [5]. As a result of the young age of symptom onset coupled with the frequent falls, MS is a substantial long-term burden on the health care system, costing an estimated to be $85.4 billion in the United States alone [6].

1.3. Clinical Assessment of Falls in MS

PwMS are most commonly assessed by a clinician at biannual office visits; an observation frequency incapable of capturing the true time-varying nature of symptoms in MS [7]. Fall history is one of the most important predictors of fall risk in PwMS [8], but only 51% of falls are self-reported [9]. In clinical practice, fall prevention interventions are not often prescribed until recurrent falls have been reported to a healthcare provider. An

objective method of characterizing fall risk may enhance our ability to prescribe preventative interventions.

1.3.1. Patient Reported Measures

A common and accessible method to assess impairment and fall risk is to use a questionnaire, known as patient reported measures (PRMs) [8], [10]–[21]. There are several PRMs designed to capture differing aspects of MS. The most common PRMs are the Activities-Specific Balance Confidence (ABC) to capture balance confidence [22], Modified Fatigue Impact Scale (MFSI) to capture fatigue [23], [24], Multiple Sclerosis Walking Scale (MSWS) to capture walking impairment [25] and Falls Efficacy Scale to capture fear of falling [26]. The Expanded Disability Status Scale is a neurologist administered PRM that is designed to assess overall disability. Each of these has been shown to relate to fall risk in some respect; ABC has been shown to be the best predictor of falls in PwMS [27], [28]. However, these assessments are subjective and only capture a single observation and therefore cannot provide an objective assessment of PwMS's disease state.

1.3.2. Functional Assessments

Another means of assessing fall risk and impairment are functional assessments. These are simple tasks that are designed to allow a clinician to assess mobility and impairment. Examples include chair stand tests [29], [30], timed up and go [31], [32], walk tests [33], [34], dual tasking [35], [36], balance tests [37]–[40], and reaction time assessments [41]. While the results of these assessments may be related to fall risk [8], [21], there is little consensus on standardized threshold that would considered a PwMS at an elevated risk of falls [29]. Additionally, these assessments, such as the Berg Balance

Assessment, may only be sensitive to more severe disease states [42]. Lastly, functional assessments require a clinician and thereby only capture a one-time assessment and do not account for any variability in time of day or participant's schedules.

1.3.3. Clinical Interventions

Once an elevated risk of falls is identified, clinical interventions are typically prescribed. Common interventions include pharmacological treatments, exercise/physical therapy, education, home set up, and using assistive devices. Studies have found that these interventions can significantly increase quality of life in MS both with and without pharmacological treatment [43]–[45]. Exercise based interventions, however, have been shown to be more effective in earlier stages of disease, highlighting the need for earlier identification of fall risk [43].

1.4. Introduction to Wearable Sensing

Traditional gait and balance assessments require cumbersome equipment such as motion capture or force platforms. In addition to the prohibitive cost of these technologies, these methods also limit assessments to laboratory-based environments. Wearable inertial measurement units may provide an opportunity to quickly and unobtrusively capture the fall-related biomechanics of PwMS, and in clinical environments [46], [47]. To this end, several recent studies have demonstrated the ability of body-worn sensors to capture biomechanical measures associated with fall risk in PwMS (e.g., walking speed [48], stride time variability [49], and balance measures of postural sway [50], [51]) and demonstrate correlations with clinical assessments [51]–[53]. Recent advances in wearable technology pair improvements in battery life with conformal designs [54] allow studies to be deployed in the free-living environment. The advancing sensor design coupled with lower device

cost, compared to motion capture, make wearables-based assessments more accessible for clinicians and patients. As a result, sensors are being integrated into point of care; thus, algorithms and analysis techniques need to be designed to utilize these data.

1.5. Wearable Sensing Methods of Fall Risk Assessment

Objective fall risk assessments can also benefit from wearable sensors. While non-wearable objective measures exist [28], [55], [56], wearables provide a simple low-cost approach to measure motion and thus fall risk. Wearables allow fall risk assessment both in laboratory/clinic environments and in free-living conditions. [47].

1.5.1. Laboratory Based Assessment Methods

Gait in PwMS is becoming a well-studied area, with several laboratory-collected measures showing relationships to fall status [46], [52]. Studies assess various aspects of gait such as stride time variability [49], backwards walking [57], and measures of stability [55], [58]. The best performing models of fall classification utilize deep learning [27] methods to learn representations straight from the raw data instead of using manual feature extraction.

Wearables-based postural sway has not been used extensively in fall risk assessment up to this point. Sun et al. shows that sway measures can be used to classify fallers from non-fallers during a lab-based standing assessment, however, accuracy is poor at 53% [59]. Most standing and balance related assessments are still conducted using force plates [50] and analyze dynamic balance [40], [60]. As a result, several open questions exist, regarding the best method to compute sway features, which body locations should be measured, etc.

In addition to gait and sway assessments, there is a body of research aiming to augment the already utilized functional assessments with wearables to increase sensitivity. Studies have shown that adding sensors to chair stand tests [29], [61] and timed-up-and-go tests [61] increases our ability to classify fall risk. While this is a promising method to increase detection, these methods still require clinician supervision, thus limiting their deployment to broader communities.

1.5.2. Remote Assessment Methods

In order to increase accessibility of assessments, wearables are beginning to be leveraged in remote studies to monitor PwMS during their daily life [47]. Storm et al. demonstrated that gait in PwMS is significantly different during free-living conditions compared to the lab [62]. The impact of these results on remote gait fall risk assessment remains an open question that will be explored in this dissertation. Interestingly, studies have found that remotely collected chair stand tests actually provide greater predictability than those collected in the clinic [63]. Postural sway measures, however, have not been investigated remotely prior to the work presented herein.

1.5.3. Prior Work

Prior to the completion of this dissertation work, the author completed a study utilizing deep learning to classify fall risk from in lab walking [27]. In this analysis, one-minute of hallway walking was found to provide excellent fall classification performance, with an area under receiver operator characteristic curve (AUC) of 0.88, when using the data from a thigh and chest accelerometer as input to deep learning models. This approach outperformed other feature-based gait approaches and other clinical methods. However, when this model was applied to remotely collected data in PwMS, the results were very

poor, with an AUC of 0.40. This result provided the motivation for the work presented herein.

1.6. Purpose of Work

The purpose of this work is to address the challenges of remote monitoring to allow for robust assessment of gait and balance impairment related to fall risk in persons with MS. This will be investigated using a data analysis platform developed by the author to characterize gait and postural sway remotely. In doing so, several questions will be investigated. Does remote walking duration affect gait and fall risk assessment? How much data is sufficient for remote monitoring? Can metrics of postural sway be measured from the chest? Is there a more sensitive way to calculate measures of postural sway? And finally, can remotely collected postural sway classify fall risk and demonstrate clinical utility?

CHAPTER 2: OPEN-SOURCE DATASET REVEALS RELATIONSHIP BETWEEN WALKING BOUT DURATION AND FALL RISK CLASSIFICATION PERFORMANCE IN PERSONS WITH MULTIPLE SCLEROSIS.

2.1. Introduction

Over 50% of falls result in injury and 66% of first-time falls require a visit to the emergency department, reducing quality of life and yielding an estimated annual healthcare cost of $80 billion in the United States alone [5]. Of the 2.3 million PwMS globally, over half will experience a fall in any three-month period [64]. As MS is a chronic condition, injurious falls pose a substantial and long-term burden to patient quality of life and the healthcare system [65].

Given these impacts, effective fall prevention is critical. Fall risk in PwMS is difficult to assess as it is known to vary both within and across days. Fall risk may be elevated in the absence of an assistive device (e.g., walking sticks) [66] or during balance-challenging tasks, such as walking, position transfers, and changes of direction [67]. However, current clinical assessments often only occur once every six months; an observation frequency incapable of capturing the true time-varying nature of symptoms in MS, limiting the ability to prescribe preventative interventions [7]. There is a clear need for novel assessments that are sensitive to this inherent variability and that can capture the relationship between symptom fluctuations and fall risk. One approach is for assessments to incorporate continuous monitoring in free-living conditions, which provide far more than a twice-per-year snapshot of symptoms, and advanced machine learning techniques that can effectively capture the complex relationship between these movement data and fall

risk.

With the growing availability of wearable sensor data, it may now be possible to leverage machine learning, and particularly *deep learning* models, to learn high-level outcomes like fall risk directly from raw sensor data without manual feature engineering [68], [69]. Studies employing deep learning for time series classification tasks, such as our prior work classifying fall risk in PwMS from in-lab measurements [27] and work from others to detect falls and classify fall risk in non-MS populations with balance and mobility impairment [70]–[78], have found superior results when compared to machine learning techniques that rely on manually-constructed features. Notably, these results are achieved despite the significant amounts of data needed for training deep learning models. It is possible that given larger available datasets, performance of these models could improve further, but the accumulation of these large datasets remains a barrier to entry for many into the use of deep learning models for characterizing fall risk.

Remote gait monitoring in PwMS may enable continuous fall risk assessment and the deployment of personalized fall prevention interventions. In this approach, data from individual walking bouts could inform fall risk status instantaneously. This vision has motivated the development of fall risk classification models that require only wearable sensor data from a single gait bout as model inputs [27], [79], [80]. However, deploying these models remotely comes with additional challenges that may impact model performance. For example, it is well established in PwMS [62], [81], [82] and other populations [83]–[85] that gait observed in the clinic differs from gait observed remotely (especially for gait speed-dependent variables). Similarly, studies in older adults [86] and PwMS [62] have also discovered that gait parameters change with walking bout duration.

However, it is currently unclear how walking bout duration relates to fall risk in PwMS [66], [86], and this has not been evaluated in previous development of fall risk classification models [27], [79], [80].

The primary objective of this work is to share a new, open-source dataset that can help other research groups develop digital biomarkers of impairment and fall risk in PwMS. In service to this objective, we present a framework for remote gait analysis on this dataset and use it to examine how gait parameters and fall risk classification performance, based on feature-based machine learning and stride acceleration based deep learning methods, change in relation to walking bout duration in PwMS.

2.2. Materials and Methods

2.2.1. Dataset: Subjects and Protocol

A sample of 38 PwMS (21:17 fallers:non-fallers; 12:27 Male:Female, mean ± standard deviation age 51 ± 12 y/o), recruited from the Multiple Sclerosis Center at University of Vermont Medical Center participated in this study (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." All participants were asked to return for two additional identical study visits six-months and one-year following their initial visit. Of the 38 original cohort, 28 returned for a six-month follow-up (15:13 fallers:non-fallers; 8:20 Male:Female), and 15 returned for a one-year follow-up (6:9 fallers:non-fallers;6:9 Male:Female). Patients completed self-reported 6-month fall history each visit, allowing their fall status to change at subsequent visits. The

high attrition rate observed in this study was largely due to the COVID-19 pandemic, as 3 six-month and 11 one-year follow-ups were cancelled for this reason.

On the day of testing, subjects provided written informed consent to participate in the study. A neurologist with subspecialty expertise in MS completed the Expanded Disability Status Scale (EDSS) for each subject [87]. Subjects were asked to complete the our fall history survey: Fall Trips and Slips 6-month Survey, Activities-specific Balance Confidence Scale (ABC) [22], Modified Fatigue Impact Scale (MFIS) [23], Neurological Sleep Index (NSI) [88], and Twelve Item MS Walking Scale (MSWS) [25]. Two missing NSI entries in the clinical survey data were filled using k-nearest-neighbors (n=3) [89]. Table 1 reports demographics of the sample.

Table 1: Subject Demographics

Visit	Assessment	Fallers	Non-faller
Initial	N	21	17
	Age	56.0 (9.05)	45 (12.92)
	Sex	5M:16F	7M:10F
	ABC	75.0 (18.8)	91.4 (15.5)
	EDSS	3.3 (1.4)	2.3 (1.0)
	MFIS	39.8 (17.9)	29.2 (16.7)
	MSWS	55.0 (23.3)	27.5 (11.5)
	NSI	56.6 (17.2)	46.6 (22.2)
6-month	N	15	13
	Age	55.3 (10.3)	44.2 (14.0)
	Sex	4M:11F	4M:9F
	ABC	73.8 (14.1)	90.1 (11.5)
	EDSS	3.3 (1.3)	2.0 (0.8)
	MFIS	41.0 (17.2)	25.1 (19.5)
	MSWS	46.0 (22.3)	31.3 (14.0)
	NSI	58.9 (23.3)	42.6 (23.2)
1-year	N	6	9
	Age	57.0 (9.7)	49.9 (10.8)
	Sex	4M:2F	2M:7F
	ABC	72.3 (22.0)	79.3 (17.1)
	EDSS	3.6 (1.6)	2.3 (1.1)
	MFIS	38.5 (14.4)	35.9 (18.9)

	MSWS	59.7 (24.4)	38.7 (12.3)
	NSI	43.0 (23.4)	60.6 (13.0)

Mean (standard deviation) of survey results partitioned by fall status. ABC: Activity-Specific Balance Confidence; EDSS: Expanded Disability Status Scale; MFIS: Modified Fatigue Impact Scale; MSWS: MS Walking Scale; NSI: Neurological Sleep Index; N: number of subjects in group

Subjects performed several activities in the lab completed in the following order: right and left tibialis anterior maximum voluntary contraction, timed-up-and-go (TUG) [1], timed 25-foot walk test [90], 30-second chair stand test [91], lying to standing transition, three separate two-minute standing tests: tandem standing, feet shoulder-width apart eyes open, and feet shoulder-width apart eyes close, one-minute hallway walk at a self-selected pace including one turn, 30-second normal standing, 30-second upright sitting, 30-second slouch sitting, and 30 seconds each lying on back, left side, right side, and prone. During the lab visit, subjects were instrumented with MC10 BioStamp sensors. Accelerometer (31.25 Hz, ±16G) and electromyography (1000 Hz) were collected from the right and left tibialis anterior. Accelerometer (250 Hz, ±16G) and angular rate gyroscope data (250 Hz, ±2000º/s) were collected from the chest and lower back as well as bilaterally from the anterior thighs, proximal lateral shank, and dorsal aspect of the feet. Electromyography was collected to allow the investigation of foot drop, a common cause of falls in PwMS [92]. Detailed placement information can be found in Table 2. At the conclusion of the lab visit, the participants were sent home with two MC10 BioStamp sensors for 48 hours located on the medial chest and right anterior thigh measuring acceleration (31.25 Hz ± 16G) placed in accordance with Table 2. Data from these sensors were recorded throughout the subject's daily life. These deidentified data are available at https://simtk.org/projects/msense_ms_adls. This protocol was approved by the University

of Vermont's Institutional Review Board (CHRMS 18-0285). Portions of this dataset have been used previously to support the development of approaches for characterizing fall risk from lab-based gait and from in-lab and remotely tracked thirty-second chair-stand tests [27], [29], [63]. In these studies, raw gait data collected in lab and deep learning models were able to adequately classify fall risk, and chair-stand-tests conducted remotely and in lab provided similar levels of fall risk classification performance.

Table 2: Sensor Placement

Location	Sensing Modality 1	Sensing Modality 2	Placement Details
Medial Chest	Accel: 250 Hz, ±16G	Gyro: 250 Hz, ±2000º/s	Secured to sternum just below sternoclavicular joint.
Sacrum	Accel: 250 Hz, ±16G	Gyro: 250 Hz, ±2000º/s	Between or just above PSIS
Anterior Thigh (R/L)	Accel: 250 Hz, ±16G	Gyro: 250 Hz, ±2000º/s	Anterior aspect of thigh ~25% from knee to hip
Proximal Lateral Shank (R/L)	Accel: 250 Hz, ±16G	Gyro: 250 Hz, ±2000º/s	Secured to proximal lateral shank, ~4 fingers below fibular head
Tibialis Anterior (R/L)	Accel: 31.25 Hz, ±16G	EMG: 1000 Hz	Placed on muscle belly (widest part) of TA
Dorsal Foot (R/L)	Accel: 250 Hz, ±16G	Gyro: 250 Hz, ±2000º/s	Placed on metatarsals 2-4

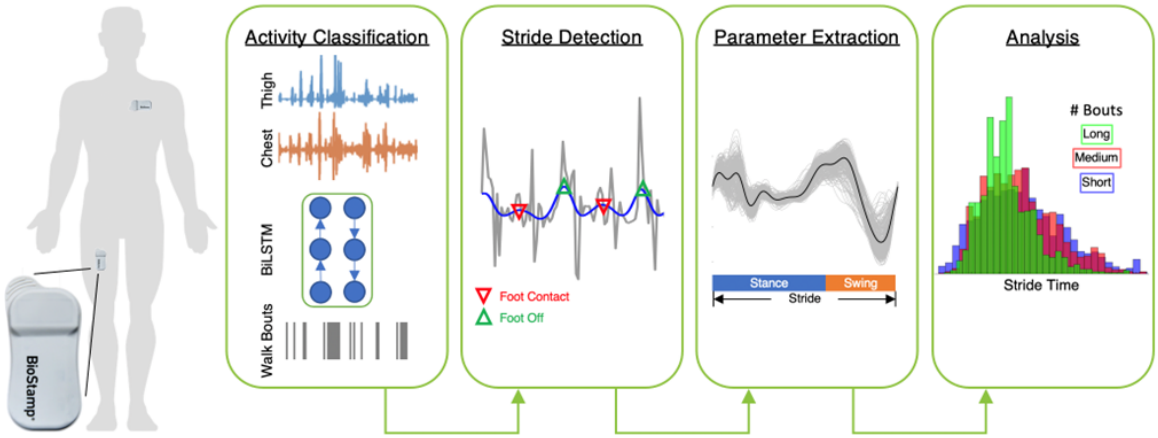
R/L: sensor placed symmetrically on right and left side of body; Accel: acceleration; EMG: electromyography; Gyro: gyroscopic data (angular velocity); PSIS: posterior superior iliac spine; TA: tibialis anterior.

2.2.2. Remote Gait Analysis

An overview of the remote gait analysis pipeline is presented in Figure 1. The depicted framework begins with acceleration gathered from the BioStamp sensors located on the thigh and chest followed by activity classification (e.g. finding walking), event detection within walking bouts, feature extraction, and finally analysis. Each aspect of this pipeline (gait bout identification, stride detection, parameter extraction, and analysis) are

discussed in more detail below. In terms of analysis, we examine the impact of context and bout duration on discriminating fallers from non-fallers, and on the performance of feature-based and deep learning methods for classifying fall risk. These analyses are only performed on the data from the initial study visit (n = 38).

Figure 1: Pipeline for free-living gait analysis from BioStamp nPoint wearable sensor data.



Pipeline for free-living gait analysis from BioStamp nPoint wearable sensor data. Activity classification is performed via deep neural network (BiLSTM architecture) on windows of accelerometer data sampled from the chest and thigh. Walking bouts are extracted from the resulting activity timeseries and gait events are identified using previously validated approaches to detect strides. Gait parameters are extracted from each walking bout and used for further analysis.

2.2.3. Activity Classification

Activity classification was carried out with wearable sensor data from the chest and thigh. Gait bouts were identified using a deep learning approach that leverages a Long Short Term Memory (LSTM), a type of recurrent neural network for analyzing time series data, architecture adapted from [93]. Specifically, the network is composed of a single Bidirectional LSTM (BiLSMT) layer with 215 hidden units [94], a 40% drop out layer [95], and Adam optimization [96]. This classifier was developed using 58% data from PwMS, 26% from healthy adults, and 16% from persons with Parkinson's Disease to

provide a wide variety of example gait and non-gait data for training. Data labeled as gait were sampled from prescribed slow, comfortable, and fast walking trials completed overground, as well on a treadmill for healthy adults. Data labeled as non-gait were sampled from standing, sitting, lying, running and stair ascent and descent. Ten-fold cross validation was conducted on the training set consisting of 20,000 4-second observations (50:50 gait:non-gait) yielding validation accuracy of 98.5%. Performance on a held-out test set consisting of 3,000 observations (50:50 gait:non-gait) was 98.4%, providing evidence that the classifier is well positioned to be used on new datasets. This network was then leveraged to identify all walking bouts completed by all subjects during the 48-hour free-living wear period. Walking bouts were identified by classifying 4-second segments of data, where consecutive walking segments were concatenated into a single bout.

2.2.4. Stride Detection

Following walking bout identification, strides were extracted using the method described and validated in [97], [98]. At a high level, this stride extraction method estimates step and stride frequency from the power spectral density of the thigh accelerometer signal. A filter bank based on these frequencies then provides the signals used to identify foot-off and foot-contact events from specific signal features. This algorithm has been validated on a wide range of walking speeds, 0.56-1.78 m/s [98], which covers the expected range of walking speeds for PwMS [99]. Bouts with fewer than two extracted strides were removed automatically before proceeding with the analysis that follows.

2.2.5. Gait Parameter Extraction

Following walking bout and stride identification, the following features were calculated for each stride and averaged for each bout; stance time, swing time, stride time,

coefficient of variation of stride time (stride time CV), duty factor, and coefficient of variation of duty factor (duty factor CV) [97]. The remaining features were calculated on the entire bout: Root mean square of the anterior-posterior acceleration from the chest sensor (RMS AP) [100], medial-lateral frequency dispersion of the chest sensor (Freqd ML) [100], and the entropy ratio between the thigh and chest [101]. Lyapunov exponent of the medial lateral (Ly ML) and anterior-posterior (Ly AP) chest sensor were calculated for gait bouts longer than 60 seconds [100].

The features mentioned above were selected based on previous literature that demonstrates their association with MS-induced gait impairment and fall risk. Stance time, swing time, and stride time have been shown to be significantly correlated with patient reported walking impairment in PwMS [102]. Stride time, duty factor [103], RMS AP, and Freqd ML have been shown to identify differences in walking impairment between PwMS and healthy controls [100]. Stride time CV has been shown to be strongly associated with fall risk in PwMS [104]. Non-linear measures, entropy ratio [101] and Lyapunov exponent in the ML and AP directions of chest acceleration [100], have been shown to capture gait stability in PwMS.

2.2.6. Walking Content and Bout Duration Analysis

Gait parameter data were grouped into one of three categories based on the duration of the walking bout from which they were extracted: short – 8 seconds or shorter; medium – 12-28 seconds; or long – 32 seconds or longer. These durations were based on results reported in other examinations of free-living gait [105]. Comparisons to gait parameters derived from lab-collected hallway-walking data and combined home data, grouped as all, were also made, see Table 4. Bouts where strides could not be identified or with

physiologically impossible values, defined as two standard deviations above and below the population average stride time, were deleted (496 removed in total). Gait parameters for each walking bout in each duration were summarized using mean, median, max, min, standard deviation, 5^{th} percentile, and 95^{th} percentile for each subject.

Group differences in each of the gait parameters were identified using Wilcoxon Rank Sum tests between fallers and non-fallers at each bout duration and between in-lab and free-living contexts. A significance threshold of $\alpha = 0.05$ was used for all statistical testing.

2.2.7. Feature-based Fall Risk Classification

Statistical models that require extracted features for discriminating between individuals at high and low risk for falls were trained and tested on five different feature-sets: gait parameters calculated on short, medium, and long gait bouts, all free-living gait bouts, and in-lab gait data. These feature-sets contained one entry per identified valid walking bout. Classifier performance was established using leave-one-subject-out cross validation (LOSO-CV). In this approach, data from all but one participant (N = 37) were partitioned into a training dataset while data from the remaining subject was used for testing. This process was repeated until data from each subject had been included in the test set. The LOSO-CV approach ensures the model was tested on subjects it had not previously seen, which provides a realistic estimate of how the model would perform during real-world use. The normalized posterior probabilities, known as the decision scores, assigned to the held-out subject were combined to calculate an overall model performance by considering the area under the receiver operating characteristic curve (AUC). AUC was chosen as the main performance metric because it provides a

comprehensive measure of how well a classifier is able to discriminate between groups without selecting a specific threshold and allows the results to be compared to other studies.

Features were normalized using z-scores then reduced using principal components analysis (PCA) within each iteration of the LOSO-CV. Prior to feature reduction, short, medium, and all-bouts have 8 features per input, long bouts have 9 features per input, and lab bouts have 11 features per input. To explain the discrepancy in the number of features, note that Entropy Ratio is computed for the long bouts and Entropy Ratio, Lyapunov Exponent AP-direction, and Lyapunov Exponent ML-direction are computed for lab walking. The principal components that explained 95% of the variance of these reduced feature sets were extracted, resulting in approximately 6 principal components for each home walking duration and 7 principal components for lab data. The reduced feature sets were then used to train Logistic Regression (LR) [106], Support Vector Machine (SVM) [107], Decision Tree [108], K-Nearest Neighbors (KNN) [109], and Ensemble of Trees (ENS) [108] binary statistical classification models to discriminate between subjects at high and low fall risk. A variety of model types were used to capture different relationships in the feature space, as each model excels with different shaped feature spaces [110]. Similar modeling approaches have been used previously to assess fall risk, as the fall risk of non-fallers is considered low and fallers high [27], [111]. Model hyperparameters were optimized with MATLAB's Optimize Hyperparameters feature, with no access to test data, for each input feature set to provide the highest classification performance in terms of AUC.

2.2.8. Deep Learning Fall Risk Classification

Based on previous literature [27], we also developed deep learning models for classifying walking fall risk. As used previously, we leveraged Long Short-Term Memory (LSTM) networks for this analysis. In our prior work, we demonstrated that the best classification performance was achieved considering four strides of data per input to the model, and showed that model performance changed with the number of strides considered [27]. For our analysis, we first optimized our networks to provide the best performance using four strides per input. This was done by extracting every walking bout with four or more strides and concatenating every consecutive, non-overlapping, four strides into a model input. These inputs contain three channels of raw acceleration from both the thigh and chest sensor from sequential strides. These data were arranged as a 6xN cell array, where the six represents the number of acceleration channels from both sensors and N represents the lengths of each stride summed. In the example case of a four-stride input, each input consisted of the thigh and chest acceleration from extracted stride 1 concatenated with the data from stride 2, then 3 and 4. Model outputs were a decision score for each input representing the posterior probability that the input belonged to a given class. Models were trained using LOSOCV, where n = 36 for training, n = 1 for validation, and n = 1 for testing for each training iteration (n = 35). A modified LOSOCV procedure was used for the deep learning methods to include an additional validation set to investigate the impacts of 18djustting the number of training epochs; note, this method ensures that all data from a given subject is only included in one of the training, validation, or test sets. Using four stride inputs, we optimized our model over the number of LSTM or Bidirectional LSTM (BiLSTM) layers, training epochs, and number of hidden units based

on the validation performance. The best two models were then selected and used to train inputs with one through twenty-two strides. The model referred to as LSTM 2 consisted of the following layers: an LSTM layer with 290 hidden units, 30% dropout, BiLSTM layer with 10 hidden units, 40% dropout, a fully connected layer, and softmax. The model referred to as LSTM 3 consisted of the following layers: an LSTM layer with 85 hidden units, 55% dropout, an LSTM layer with 85 hidden units, 55% dropout, an LSTM layer with 235 hidden units, 45% dropout, a fully connected layer, and softmax. The models were trained for 55 and 125 epochs, respectively, and both utilized Adam optimization. Model denoted as ABC contained the subjects' ABC score in the model inputs. Performance was assessed using area under the receiver operator curve (AUC) from the held-out test set for individual input predictions and for an aggregated model performance using the median classification from each subject.

2.3. Results

A total of 15,097 free-living walking bouts were analyzed, with 9,135 (61%) identified as short, 4,840 (32%) as medium, and only 1,122 (7%) as long. Gait parameters differed considerably between bout lengths (Table 3). Notably, stride time CV, swing time, duty factor CV, RMS AP, and Freqd ML were significantly different between all bout durations. Stride time CV and RMS AP increased, and Freqd ML decreased with increasing duration. The increase in stride time CV at home may indicate greater stride to stride variability. Swing time of short and medium bouts was similar and greater than that observed during long bouts. Collectively, the increase in motion in the direction of travel and decrease in lateral motion implies that PwMS walk with greater stability during longer walking bouts.

Table 3: Difference of Medians Testing for Free-Living Gait Parameters from Differing Bout Lengths

Comparison	Feature	Median 1	Median 2	p-value
4-8 vs. 12-28	Stride Time	1.16	1.16	0.460
	Stride Time CV	0.067	0.085	***< 0.001***
	Stance Time	0.71	0.71	0.644
	Swing Time	0.44	0.45	***0.012***
	Duty Factor	0.62	0.62	***0.025***
	Duty Factor CV	0.042	0.054	***< 0.001***
	RMS AP	0.14	0.14	***< 0.001***
	Freqd ML	0.62	0.60	***< 0.001***
4-8 vs. 32+	Stride Time	1.16	1.12	***< 0.001***
	Stride Time CV	0.067	0.079	***< 0.001***
	Stance Time	0.71	0.69	***< 0.001***
	Swing Time	0.44	0.42	***< 0.001***
	Duty Factor	0.62	0.62	***0.005***
	Duty Factor CV	0.04	0.051	***< 0.001***
	RMS AP	0.14	0.15	***< 0.001***
	Freqd ML	0.62	0.55	***< 0.001***
12-28 vs. 32+	Stride Time	1.16	1.12	***< 0.001***
	Stride Time CV	0.085	0.079	***< 0.001***
	Stance Time	0.71	0.69	***< 0.001***
	Swing Time	0.45	0.42	***< 0.001***
	Duty Factor	0.62	0.62	0.081
	Duty Factor CV	0.054	0.051	***0.011***
	RMS AP	0.14	0.15	***< 0.001***
	Freqd ML	0.61	0.55	***< 0.001***

Rank sum test with level of significance $\alpha = 0.05$, significant results bolded and italicized. Stride, swing, and stance time in seconds; Freqd ML in Hz; RMS AP in √g; Duty Factor is unitless.

Significant differences between home and lab walking were found for all bout durations (Table 4). Freqd ML was significantly higher in free-living than in-lab conditions for all walking durations, with the shorter durations showing the largest differences. Stride time was also increased in free-living gait, with significant differences found in short, medium, and combined walking durations. As expected, these results imply that longer free-living walking bouts are the most similar to those completed in the lab, however, significant differences in the longer bouts remain. Specifically, the long free-living bouts

have significantly higher entropy ratios, and Lyapunov exponents in the AP direction than those completed in the lab – each of which indicates a decrease in stability in free-living situations.

Table 4: Difference of Medians Testing for Free-Living and in Lab Gait Parameters from Differing Bout Lengths

Comparison	Feature	Median In-lab	Median Free-living	p-value
Lab data vs 4-8	Stride Time	1.11	1.16	***< 0.001***
	Stride Time CV	0.06	0.067	***0.030***
	Stance Time	0.69	0.71	***0.001***
	Swing Time	0.42	0.44	***0.045***
	Duty Factor	0.62	0.62	0.473
	Duty Factor CV	0.045	0.042	***< 0.001***
	RMS AP	0.14	0.14	0.641
	Freqd ML	0.55	0.62	***< 0.001***
Lab data vs. 12-28	Stride Time	1.11	1.16	***< 0.001***
	Stride Time CV	0.06	0.085	***< 0.001***
	Stance Time	0.69	0.71	***0.001***
	Swing Time	0.42	0.45	**0.014**
	Duty Factor	0.62	0.62	0.979
	Duty Factor CV	0.045	0.054	***0.050***
	RMS AP	0.14	0.14	0.538
	Freqd ML	0.55	0.60	***< 0.001***
Lab data vs. 32+	Stride Time	1.11	1.12	0.160
	Stride Time CV	0.060	0.079	***< 0.001***
	Stance Time	0.69	0.69	0.206
	Swing Time	0.42	0.42	0.312
	Duty Factor	0.62	0.62	0.776
	Duty Factor CV	0.045	0.051	***0.050***
	RMS AP	0.14	0.15	***0.042***
	Freqd ML	0.55	0.55	0.205
	Lyapunov AP	0.006	0.014	***< 0.001***
	Lyapunov ML	0.013	0.012	0.194
	Entropy Ratio	1.78	2.55	***< 0.001***
Lab data vs All home data	Stride Time	1.11	1.16	***< 0.001***
	Stride Time CV	0.060	0.064	0.821
	Stance Time	0.69	0.71	***0.002***
	Swing Time	0.42	0.44	***0.037***
	Duty Factor	0.62	0.62	0.673
	Duty Factor CV	0.045	0.042	0.195
	RMS AP	0.14	0.14	0.948
	Freqd ML	0.55	0.61	***< 0.001***

Unit of stride, swing, and stance duration is seconds. Unit Freqd ML is Hz, RMS AP is √g, and Duty Factor is unitless. All p-values found using a rank sum test using a significance threshold of 0.05, significant results are bolded and italicized.

Significant differences between the gait parameters of fallers and non-fallers were observed for short and long walking bouts as seen in Table 5. Notably, in short walking bouts, we see fallers have a lower RMS AP, signifying higher impairment as expected [100]. This suggests short and long walking bouts are more sensitive to fall risk compared to medium duration walking bouts. Fall classification models trained on the gait parameters explored in this study performed best on lab walking bouts and short walking bouts when considering home walking only (see AUC of knn for 8-seconds or less in Fig. 2).

Figure 2: Fall Risk Classification Model AUC for Short Home, Medium Home, Long Home, All Home, and In-Lab Walking Bouts

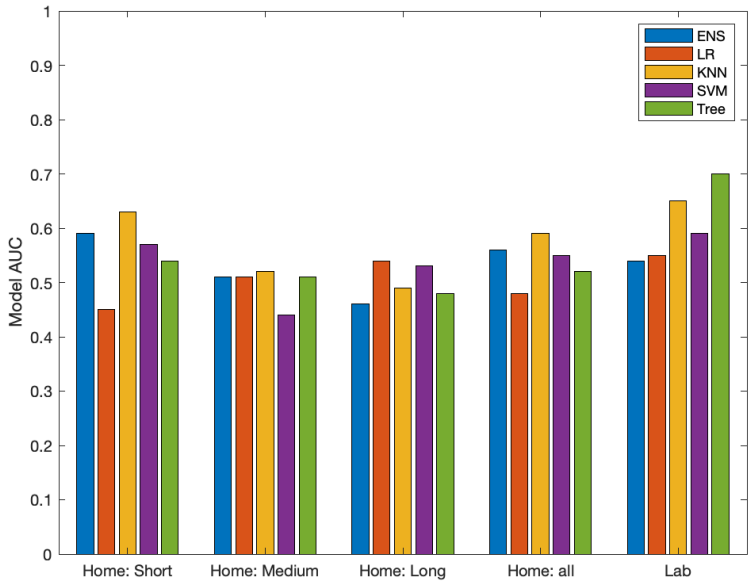


Table 5: Significant Differences of Medians of Gait Parameters for Fallers vs Non-Fallers from Differing Bout Lengths

Bout Duration	Feature	Median Faller	Median Non-faller	p-value

4-8	Max RMS AP	0.25	0.36	***0.012***
	Med RMS AP	0.13	0.15	***0.019***
	Mean RMS AP	0.14	0.15	***0.019***
	95th P RMS AP	0.20	0.22	***0.037***
12-28	No significant differences			
32+	5th P Duty Factor	0.57	0.58	***0.017***
	Mean Duty Factor	0.61	0.62	***0.017***
	Med Duty Factor	0.61	0.63	***0.018***
	Max Duty Factor	0.65	0.66	***0.029***
	Max Stance Time	0.79	0.86	***0.033***
	95th P Duty Factor	0.64	0.65	***0.038***
	Min Entropy Ratio	1.06	0.92	***0.039***
	Med Swing Time	0.43	0.43	***0.040***

Unit of step, swing, and stance duration is seconds. Unit Freqd ML is Hz, RMS AP is $\sqrt{g}$, and Duty Factor is unitless. All p-values found using a rank sum test using a significance threshold of 0.05, significant results are bolded and italicized. 5th or 95th P: 5th or 95th percentile; Med: Median.

The best overall feature-based fall classifier was a decision tree model using lab walking bouts. Performance of this model was characterized by an AUC of 0.70. The best performing feature-based home fall classification model was a KNN with short bout inputs achieving an AUC of 0.63. The KNN model also performed best for medium walking bouts, and all home walking bouts, providing AUCs of 0.52 and 0.59 respectively. The best performing feature-based model on long home walking bouts was the LR model, with an AUC of 0.54. The best performing deep learning model was the LSTM 2 trained on inputs with 22 strides with ABC for all walking bouts using the median aggregation of all classifications at home with an AUC of 0.76. The best performing non-aggregated model was LSTM 3 with ABC trained on input with three strides from all walking bouts. Detailed performance of the models can be found in Table S1, located in the appendix. Figure 3 reveals that when using the median aggregation, the performance of the medium bouts sees a notable improvement compared to the other bout lengths, suggesting that the aggregation may be reducing some of the noise inherent in that walking duration. Figure 4 shows the

performance of each model relative to its input size, which seems to show that short, medium, and long bouts continue to increase their performance with dataset size. In contrast, the all-bouts models seem to achieve stable performance levels as dataset size is increased.

Figure 3: Fall risk classification model AUC for best performing deep learning model from short, medium, long, and all walking bouts for 1-5 inputs per stride without aggregation (left) and with median aggregation (right).

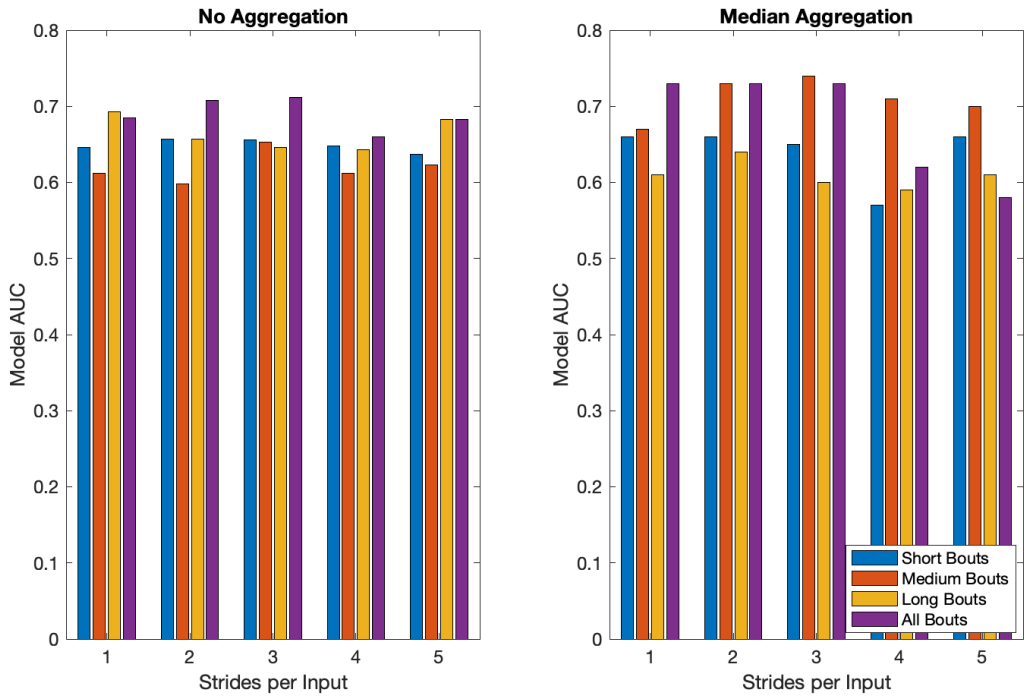
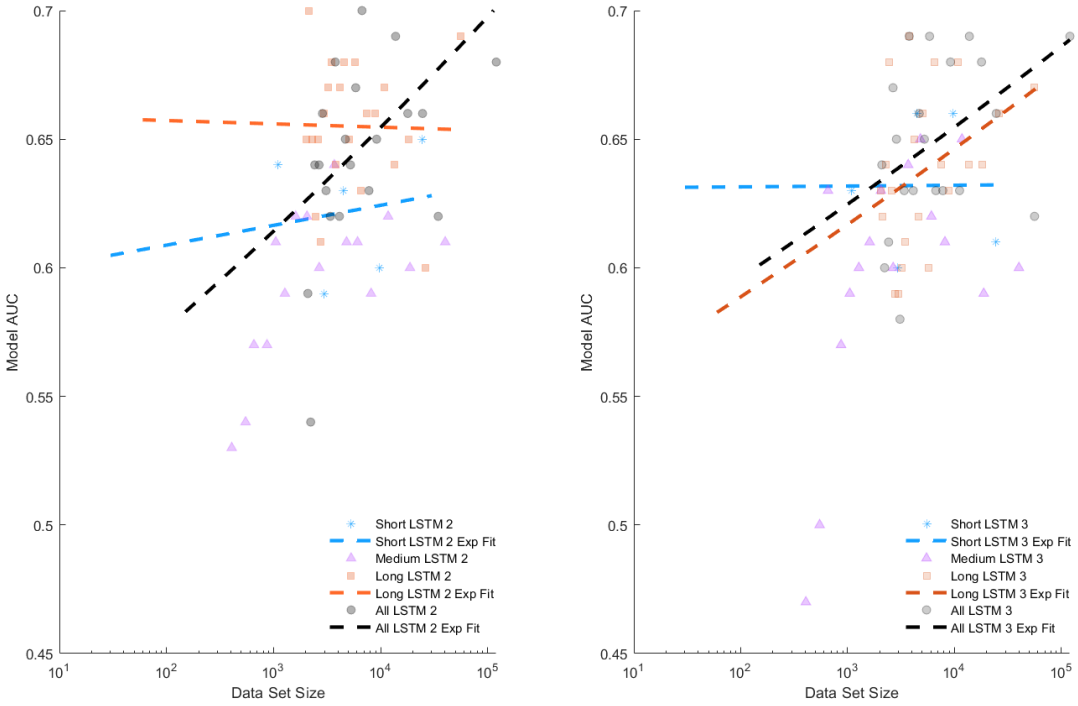


Figure 4: Fall risk classification model AUC for LSTM 2 ABC and LSTM 3 ABC for all stride durations colored by bout length, short (blue), medium (pink), long (red), and all (black), plotted against the training set size for each model.



Fall risk classification model AUC for LSTM 2 ABC and LSTM 3 ABC for all stride durations colored by bout length, short (blue), medium (pink), long (red), and all (black), plotted against the training set size for each model showing increasing performance, increasing exponential fits, for several model/bout configurations with data set size. Notice the stronger increasing trends in the right LSTM 3 plots in all and long bouts compared to the LSTM 2 plot. Additionally notice the increase in slope of short LSTM 2 compared to short LSTM 3. This suggests that the larger models are needed to capture variability in longer bouts and smaller models perform better with shorted bouts. Note, the medium trend (not shown) was strongly increasing for both LSTM 2 and 3.

The impact of these results is twofold. First, considering the feature-based methods, these models show that overall fall risk is best predicted by lab walking and that for free living gait fall risk is best predicted by considering short-duration walking bouts. Second, we show that deep learning models trained on raw stride data perform better on home data when considering all bouts and using a larger number of strides per input. As the strides per input increase, the gait is likely more similar to steady-lab walking than variable free-living walking. With this hypothesis, both the feature-based models and deep learning

modeling reach a similar conclusion (supported by Table 4), namely that many consecutive clean strides are needed to classify fall risk using this framework. Figure 4, however, shows that the performance of both models using medium, and all bouts seems to increase with dataset size. Short bouts using the LSTM 2 model also appear to show and increasing performance with more data, however, the limited range of data set sizes for small data limits the ability to find trends. Performance using long bouts is better captured using a larger model such as LSTM 3 which shows improvement with increasing data set size compared to the smaller LSTM 2 model where this trend does not exist. These trends, however, suggest that the addition of more data, and perhaps models that can better account for the variability may provide better performance.

2.4. Discussion

In this paper we present a novel wearable sensor dataset collected from PwMS. This dataset includes data from a supervised laboratory visit, neurologist assessments, patient reported measures, and an unsupervised monitoring period for each PwMS. Novel findings from the in-lab period of this study have found walking and 30-second chair stand tests to be indicative of fall risk [27], [29]. Analysis of free-living 30-second chair stand tests and posture transitions have also revealed relationships with fall risk and impairment [63]. Herein, we presented a preliminary analysis of walking in the free-living environment as it relates to fall risk and differing lengths of walking bouts.

The main finding from this study is that both gait bout length and environment influence wearables-based fall classification in PwMS. Specifically, the best performance overall was observed for classifiers that use lab data or long, steady walking bouts that are similar to the lab (Fig. 2 and Table A1). The best performing feature-based model on free-

living data was trained on short walking bouts, suggesting that short free-living bouts may be worth further exploration with a more nuanced feature-set. Our best un-aggregated deep learning model was trained on 3-stride inputs from all bouts. We hypothesize this performed best because deep learning models require a large amount of data to train and considering all bouts allows the model access to far more data than just the short bouts.

Compared to other fall risk classification studies, the performance of our remote fall risk classifier is on par with many lab-based studies, but still lags behind the best approaches. In-lab studies have achieved AUCs between 0.73 and 0.79 in older adults [112]. In PwMS an in-lab study using the dynamic gait index achieved an AUC of 0.80 [55] and our prior work, where a deep learning model was used on walking data, achieved an AUC of 0.88 [27]. The difference between our previous lab-based fall risk performance of 0.88 and the performances presented herein highlights a key challenge in using deep learning methods on remote data. Namely, that the model must be able to reconcile the additional variability in gait observed under free living conditions. Performance was observed to increase with increasing dataset size in figure 4, indicating that deep learning approaches may be able to learn appropriate representations of the data to account for this variability, but the dataset considered here is likely not large enough. By open-sourcing these data, we aim to allow future researchers to realize the promise of deep learning for fall risk classification in PwMS.

Our finding that bout length and environment influence discrimination of fallers from non-fallers is in agreement with similar gait-based classification applications in patients with neurological disorders. For example, one study found that the features that best discriminate between PwMS and healthy controls were different when using lab data

and home data [113]. Similarly, other studies demonstrate that shorter walking bouts provide better discriminative power when trying to identify a person with Parkinson's Disease versus healthy controls as well [105], and pace is different in free-living walking compared to in-lab for PwMS [62].

The influence of bout length and environment on fall classification is likely related to the observed differences in the various gait descriptors used as features in the classification models (Tables 3 and 4). This finding contributes more generally to the growing body of evidence that controlled in-lab observations of gait are not representative of free-living conditions. In the current study, this discrepancy was more pronounced for short and medium walking bouts than for long; a finding which is likely due to the fact that the in-lab walking bout was, by our definition, a long walking bout (one-minute long). Differences observed between gait parameters calculated at differing bout lengths (see Table 3) show that stride, stance, and swing time decrease as bout duration increases. This likely means that PwMS are increasing their cadence for longer walking bouts. The observed decrease in ML frequency dispersion with increasing bout length also suggests PwMS walk more steadily, with less lateral motion for long duration walking bouts. These results are consistent with Storm et al., who found that gait pace significantly increased and variability significantly decreased with increasing bout length [62]. Karle et al. found little correlation between an in-lab 2-minute walk test and free-living walking [81]. In older adults, Najafi et. Al observed significantly different walking strategies between short and long walks [86]. The reason for this change in gait is unknown, however, it can be speculated that shorter walking bouts may elicit more goal-direction actions towards activities other than walking while longer bouts are more purposeful [105]. Further

expanding on the involuntary nature of shorter walking bouts, subjects may be more likely to be dual-task walking, in other words focused on more than just walking, and may be more impacted by the start-up and stopping strides [114]. This conjecture aligns with research on dual-task walking in PwMS that shows dual-task walking is more discriminative of impairment than single task walking [57].

The distribution of bout length in free-living gait from the current sample (61% short, 32% medium, 7% long) is comparable to what has been observed in Parkinson's disease [105]. Preliminarily, this consistency across populations may suggest a phenomenon that is representative of free-living gait more generally. This raises important questions concerning remote gait analysis more broadly to be investigated in future research. For example, does bout length explain the free-living vs. in-lab discrepancy in various gait descriptors consistently observed across multiple populations? If the observed distribution of bout lengths does generalize, then free-living gait is generally short-bout and less purposeful while long, purposeful walking is rare. Further, given that in-lab investigations of gait are controlled and supervised by a clinician or researcher, they may naturally elicit more purposeful walking from the subject (even over short distances) and be less prone to the impacts of fatigue inherent in daily-life. Thus, differences in free-living and in-lab gait may be explained by the fact that aggregated metrics of free-living data (e.g., average gait speed in a 24-hour period) are dominated by those characteristic of short-duration gait bouts (> 50%) and is influenced to a far lesser extent by metrics characteristic of long-duration and purposeful gait bouts (< 10%).

There are several limitations to our study. First, our relatively small sample with moderate to low impairment may not generalize to a larger population of PwMS,

particularly PwMS with EDSS greater than six, who were not represented in this study. Other studies utilize different sensing modalities that provide gait speed, which was not available with our data collection set up. Additionally, our analysis methods require a four second window to be classified as non-walking to denote separate bouts. This definition of what defines a separate bout may impact certain gait quantity metrics, however, our study uses gait quality metrics which have been shown to be independent of temporal gait bout definitions [115]. Lastly, symptoms in PwMS are known to fluctuate over differing time scales and thus, 48 hours may not have been a long enough collection time to provide an accurate depiction of each participants overall mobility status [7]. Future work will be needed to determine how gait parameters vary in PwMS on longer time scales.

With the presented dataset, we hope to alleviate one of the most challenging issues related to human subject research with wearables: not having enough data. Publicly available datasets gathered from PwMS are largely related to medical imaging [116]–[118] and medication [119]. One dataset tackles a related issue: remote fall detection in PwMS [120], however, it is lacking data from PwMS who have yet to become recurrent fallers, preventing the investigation of gait as it relates to distinguishing fallers from non-fallers and potentially fall-risk prediction. Utilizing the presented data, potentially with other collected or open-source data, researchers may be able to leverage deep learning to enhance the performance of their digital biomarkers and phenotypes, and particularly for detecting fall risk in PwMS in both lab and free-living environments. With that said, the vision of real-time monitoring fall risk monitoring comes with challenges such as when and how to alert the user to an elevated fall risk, how or if to integrate with their comprehensive care, and these data need to be protected. These are all challenges that will need to be addressed

and researched in the future as we move towards a preventative care paradigm for falls in PwMS and other populations with balance and mobility impairment.

2.5. Conclusion

Herein, we introduce a new open-source dataset featuring activities of daily living and functional assessments from a lab environment as well as two days of free-living data in PwMS. This dataset features data from PwMS with lower impairment, including approximately half that do not yet have recurrent fall histories. As an example use case, we present a study of gait in the free-living environment. In this study, we explored differences in gait parameters calculated on short, medium, and long duration walking bouts. Specifically, we investigated the significant differences between durations of home walking and in-lab walking and fall classification performance using features calculated from differing walking durations. Several significant differences were found between the gait parameters at differing durations. We also demonstrated that fall risk classification performance using gait changes based on walking bout duration. Short walking bouts, 8 seconds or less, were found to be the most discriminative, providing significant differences between fallers and non-fallers and providing the best free-living fall classification performance in the feature-based models. Additionally, we demonstrated that in-lab walking gait parameters are significantly different from free-living walking, at all durations, and that fall risk models used on remote data should be trained with remote data. While future studies are required to assess the reliability of these findings over a longer time period, these results suggest that remote gait analysis may benefit from focusing on short walking bouts in future analysis.

CHAPTER THREE: HOW MUCH DATA IS ENOUGH? A RELIABLE METHODOLOGY TO EXAMINE LONG-TERM WEARABLE DATA ACQUISITION IN GAIT AND POSTURAL SWAY

3.1. Introduction

Wearable sensors are increasingly common, with a vast number of uses including health research [47], [56], [63], [121]–[128], and fitness tracking [129]–[131]. Laboratory-based studies of features of gait and postural sway have contributed important foundational knowledge to the field of wearable sensor-based movement tracking [59], [122], [132]–[136]. However, they also indicate that movement characteristics measured in the lab often do not reflect those displayed during daily life, and thus only capture a limited picture of balance and mobility impairment [123], [137], [138]. Recent advances in wearable technology pair improvements in battery life with conformal designs [54], allowing studies to be deployed in the free-living environment. Work in this emerging area has focused largely on demonstrating feasibility [139] and identifying reliable measures of gait performance [121], [137].

Many free-living studies of balance and mobility have been conducted in older adults, however, these studies may be more informative in certain clinical populations, such as those with neurological disorders [47], [140]. For example, persons with multiple sclerosis (PwMS) experience symptom fluctuations due to disease. As a result, a bi-annual clinic visit or in-lab assessment may not capture an accurate picture of their impairment, nor its variability [141]. In contrast, remote observation with wearable sensors could enable a more accurate assessment of balance and mobility impairment that is sensitive to variability over time and is captured while patients are engaging in their everyday lives. However, it

is not yet clear how long we must monitor these patients to capture an accurate picture of their impairment and its variability. Prior studies have found that the necessary wear duration for capturing measures of mobility impairment depends on the activity, metric, and population being considered. For example, two to three days of data are required to capture gait speed, and four days are required for capturing daily step counts in healthy adults [138], [142]. Three days are needed for remotely monitoring chair stand tests in healthy adults and persons with Parkinson's disease [143]. Between two and seven days are required for physical activity metrics (e.g., Actigraphy) depending on which metric is being considered [144]–[149]. While these studies recommend sensor wear durations, there is not an established method for arriving at these conclusions. Some studies rely on intra-class correlations (ICC) for this analysis [138], [143]. Others use analysis of variance [146], generalizability theory [148] (similar to ICC), or a combination of ICC and difference testing [147]. Each of these approaches considers different aspects of the data, leading to slightly different conclusions. Moreover, these methods do not consider how wear duration impacts the relationship between sensor-derived parameters and other important variables such as patient-reported measures of impairment. This burgeoning field of research has only considered a small subset of potential wearable-derived metrics for characterizing balance and mobility impairment and a standardized and rigorous process for evaluating necessary wear duration remains an unmet need.

For PwMS, prior work has identified key laboratory-based measures of balance and mobility impairment that can be derived from wearable sensor data. Spatiotemporal gait metrics, such as gait speed and stride time, have been associated with disease severity [137], [150] and fall risk [122]. Similarly, postural sway metrics have been associated with

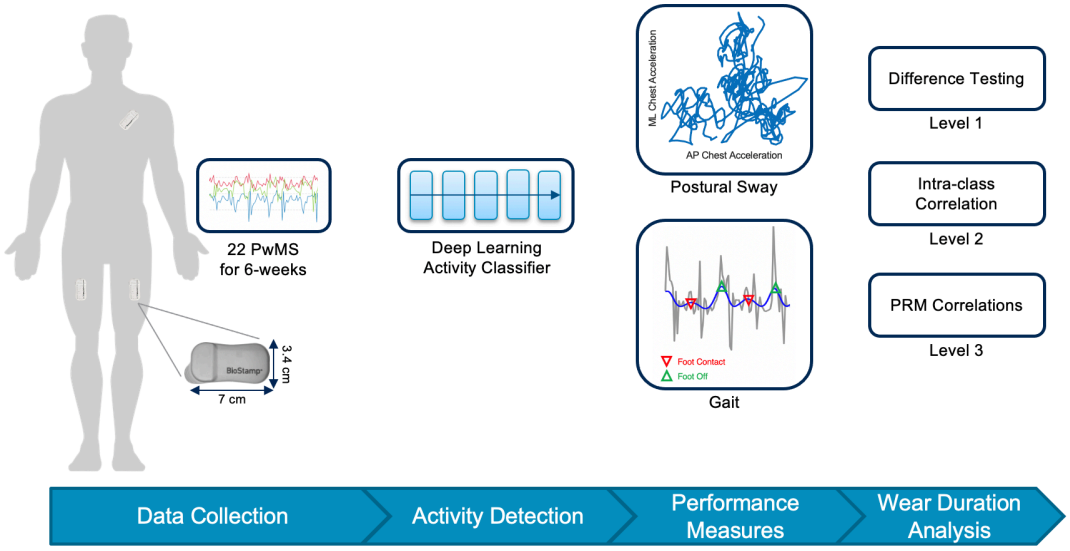
fall risk and balance impairment in PwMS [56], [151]. Given these findings, it is likely that remote monitoring of gait and postural sway could be important in this population. However, it remains unclear how long PwMS must be monitored in the free-living environment to provide reliable measurements of these parameters.

The purpose of this work is to demonstrate a comprehensive and reproducible approach for determining the wearable monitoring duration needed to capture an accurate picture of impairment and its variability. We aim to establish this minimum monitoring period to balance patient burden, convenience, and cost. We apply this approach to study the impact of wear duration on postural sway and gait measures in PwMS.

3.2. Materials and Methods

An overview of the approach used for studying the wear duration required for capturing gait and balance impairment and its variability with wearables in a sample of PwMS is provided in Figure 5. As depicted, we remotely collected data from PwMS, used a classification model to identify period of walking and standing, computed metrics of gait and sway, and then performed our analysis of wear duration. The sample and associated experimental protocol are detailed in Section 3.2.1. The framework for detecting walking and standing bouts and extracting associated balance and mobility performance parameters is presented in Section 3.2.2. Finally, the three-stage wear duration analysis is presented in Section 3.2.3 along with details for how it was implemented in this study.

Figure 5: Overview of approach for wear duration analysis.



Accelerometer data were collected from BioStamp nPoint® (Medidata) devices worn on the chest and bilaterally on the thighs by a sample of persons with multiple sclerosis (PwMS) for six-weeks. Sensors were worn for all hours of the day during monitoring periods. A deep learning approach was used to detect bouts of walking and standing from which performance measures were extracted. Measures were analyzed through a three-stage process to determine the number of days of wear required to capture an accurate picture of impairment and its variability.

3.2.1. Participants and Protocol

Herein we consider data from 22 PwMS (5:16 Male:Female, mean ± standard deviation age 51 ± 9 y/o) recruited from the Multiple Sclerosis Center at University of Vermont Medical Center and from the University of Vermont's IDEAL for MS Program (inclusion: no condition affecting balance and mobility other than MS, ambulatory without aid, no known skin hypersensiveity to adhesives or hydrogel, not preg-t or breastfeeding).

Participants were asked to complete biweekly at-home sensor wear for 12 weeks, yielding six weeks of sensor data for analysis. All participants completed at least two-weeks of monitoring (n=22), 21 participants completed five weeks and 19 completed all six weeks of monitoring. During the sensor wear weeks, participants were asked to complete a daily 30-second chair stand test, a one-minute walk, and a 30-second standing

balance assessment (upright and still for 30-seconds). Each evening participants were asked to complete a daily falls survey and the Acitivity-Specific Balance Confidence assessment (ABC) [22]. At the end of each non-sensor wear week, participants were asked to complete the Modified Fatigue Impact Scale (MFIS) [23] and 12 item Multiple Sclerosis Walking Scale (MSWS) [25]. This timing was chosen because these surveys ask participants to recall the past two weeks. During active sensor wear weeks, participants were instrumented with BioStamp nPoint sensors for all hours of the day located on the left upper chest, and bilaterally on the anterior aspect of each thigh collecting acceleration (31.25 Hz, ±16G) and electromyography (250 Hz) data. This sensor system is FDA cleared and details of these sensors has been perviously disucssed [152]. Data from the sensors was saved locally on the sensors and then uploaded to the nPoint cloud via a provided dock following a daily sensor change. Participants were also asked to fill out the Patient-Determined Disease Steps (PDDS) following the completion of their monitoring period [153]. Due to the remote nature of the study, patient disability was assessed with PDDS instead of a neurologist-conduct Expanded Disability Severity Assessment [154]. The mean (std) survey results for our cohort are as follows: PDDS 0.88 (1.05); ABC 77.6 (21.9); MFIS 28.3 (16.1); MSWS 19.1 (7.1). This protocol was approved by the University of Vermont's Institutional Review Board (CHRMS 21-0401).

3.2.2. Remote Analysis Pipeline

Periods of walking and standing were identified in the wearable accelerometer data from each participant using a classification model described previously [123]. Briefly, this model uses a deep learning approach where windows of raw accelerometer data from the thigh and chest are classified as walking, standing, sitting, lying or other using a model

with two Long-Short-Term-Memory layers [123], [155]. This model was trained on over 100,000 four-second observations of acceleration from a different cohort of PwMS, healthy adults, and persons with Parkinson's disease resulting in a validation accuracy over 96%.

Following activity classification, gait events were identified using the thigh acceleration-based method described in previous work [98], [121]. Walking bouts with two or more valid strides were used for analysis. Temporal, stability, and asymmetry measures of gait were extracted from each bout. The temporal gait parameters (computed for each stride and averaged across the bout) considered were stride duration, stance duration, swing duration, duty factor and double support duration [121]. The gait asymmetry parameters considered were duty factor asymmetry (normalized using the L1-norm, $Asymmetry = \frac{|Right-Left|}{0.5(|Right|+|Left|)}$), an affine transformation of the correlation between right thigh and left thigh raw acceleration such that a result of one corresponds to a correlation of zero with the transformation ($Correlation\ Asymmetry = 0.5 * (1 - corr(right, left))$), and asymmetry of an ensemble average of stride acceleration between the right and left leg normalized by the L1-norm method (Acceleration Asymmetry) [121]. The gait stability parameters considered were the root-mean-square (RMS) of the anterior-posterior (AP) acceleration (RMS AP), frequency dispersion of the media-lateral (ML) acceleration [135], entropy ratio between the trunk and thigh [156], and Lyapunov Exponent of the AP and ML directions [135]. The entropy ratio was only calculated for walking bouts longer than 30 seconds, and the Lyapunov Exponent was only calculated for walking bouts longer than one minute. Entropy ratio asymmetry normalized using the L1 method was also calculated.

3.2.3. Wear-time Analysis

We propose a three-stage process for analyzing the wear duration required for capturing impairment and its variability with wearables. The stages include 1) difference testing, 2) intra-class correlation (ICC), and 3) correlation to established clinical measures. These three stages fit nicely into the emerging framework for identifying digital medicine technologies that are fit for purpose [140], [157]. Specifically, Stages 1 and 2 are key aspects of Analytical Validation, which aims to establish the performance of algorithms that translate raw sensor data into measures of human physiology or behavior. Stage 3 is a key aspect of Clinical Validation, which aims to demonstrate that a digital medicine technology captures the phenotype of interest in the intended clinical population. To inform the proposed three-stage analysis, the researcher must first partition their dataset into candidate wear durations (e.g., 1 hour, 1 day, 1 week, 1 month) and choose a duration (baseline) they aim to compare against. The baseline should be a wear duration that is expected a priori to capture impairment and its variabilty for the measures being considered. In Stage 1, statistical difference testing (e.g., ranksum tests) is used to identify the wear durations that demonstrate significant differences in the median, 95^{th} percentile, or variability (coefficient of variation – CV) relative to baseline. In Stage 2 intra-class correlation analysis is used to identify the reliability of measures extracted from each wear duration relative to baseline. We recommend conducting this analysis with the median, 95^{th} percentile, and variability of each measure, to capture the central tendency and edges of the distribution, using a 'C-k' or similarity type correlation, as done previously, with the addition of CV to capture variability [138]. Wear durations that demonstrate an ICC of 0.7 and higher provided reliable measures. In Stage 3, correlation (e.g., Spearman rank

correlation) between measures extracted from each wear duration and estabished clinical measures are examined. Wear durations that yield significant correlations with established clinical measures are said to capture the clinical phenotype of interest in the intended patient population.

For the dataset of PwMS we considered here, one week of data served as the baseline, and we examined wear durations of one, two, or three days, and two, three, four, five, and six weeks in Stages 1 and 2. For Stage 3, we examined wear durations of between one and fourteen days. For the statistical analysis, we leveraged non-parametric ranksum testing with a significance threshold of 0.05 for Stage 1 and 'C-k' type ICCs (threshold of 0.7 indicating a strong ICC) for Stage 2. For Stage 3, we computed Spearman rank correlations between sensor-derived measures (median, 95^{th} percentile, and CV) at each wear duration and the the ABC (median over the two weeks) and MFIS (sampled at the end of the two weeks). The ABC captured balance confidence and has been shown to relate to fall risk and other functional-assessments [158], [159]. The MFIS captured fatigue [160], which has been shown to relate to fall risk [159], and is widely used clinically [161]. We considered the emergence of a significant correlation followed by similar strength correlations to be a reliable estimate for required wear duration. A power analysis was performed on selected features to determine the stability of these findings. This was done using a bootstrap with 1,000 replicates and comparing the synthetic data between two durations [162].

The outlined three-stage wear duration analysis allows us to identify the number of days of wear required for capturing impairment and its variability with wearables. However, it is important that we also understand what factors may impact wear duration

so that we can better predict how many days may be required. To this end, we employed regression analysis to investigate our hypothesis that longer wear periods are required for measures with relatively few observations or with high variability during a given day. We operationalized this hypothesis by defining the wear duration as number of days required to yield no statistically significant differences relative to baseline and strong ICCs for a given physiological measure. Number of observations was captured by considering the average number of times a measure was computed per day and variability was captured with the CV of the measure per subject across a two-day period.

3.3. Results

Examining the outcome of the three-stage wear duration analysis, we first report results of the difference testing (Diff) and intra-class correlation (ICC) analysis for measures of gait (Section 3.3.1) and postural sway (Section 3.3.2) across the wear durations noted above. In Section 3.3.3 we then examine the correlations between the gait and sway measures and PRMs of balance confidence and fatigue for wear durations ranging from one day to two weeks. Finally, in Section 3.3.4, we present findings from the regression analysis.

The results of the Diff and ICC stages of the wear duration analysis for the gait and sway measures are summarized in Table 6 for a baseline wear duration of one week. This table reports a percentage of features that have strong ICCs, in the ICC column, and percentage of features that do not have significant differences in the Diff column. Percentages were used to report the results in a simple frame of reference.

Table 6: Summary of reliability analysis including difference testing and intra-class correlation

Comparison	Gait ICC	Gait Diff	Sway ICC	Sway Diff

	Median	95th P	CV	Median	95th P	CV	Median	95th P	CV	Median	95th P	CV
1 Day v 1 Week (n=22)	100	80	60	100	100	80	38	46	8	100	100	77
2 Days v 1 Week (n=22)	100	100	90	100	100	100	100	92	92	100	100	100
3 Days v 1 Week (n=22)	100	100	100	100	100	100	100	100	100	100	100	100
2 Weeks v 1 Week (n=22)	100	100	100	100	100	100	100	100	85	100	100	100
3 Weeks v 1 Week (n=21)	100	100	100	100	100	100	100	100	100	100	100	100
4 Weeks v 1 Week (n=21)	100	90	80	100	100	100	100	100	100	100	100	100
5 Weeks v 1 Week (n=21)	100	90	90	100	100	100	100	100	85	100	100	100
6 Weeks v 1 Week (n=19)	100	100	100	100	100	100	100	100	85	100	100	100
Weekday v Weekend (n=22)	100	90	90	100	100	100	100	100	100	100	100	100

[7] Summary results of ranksum difference testing (Diff) and intra-class correlation (ICC) for gait (10 parameters) and sway (13 parameters) measures in persons with multiple sclerosis (MS). For Diff, values are percentage of features that did not have a significant difference across wear durations. For ICC, values are percentage of features that had strong (≥0.70) ICC between the compared wear durations. Lyapunov Exponent AP, Lyapunov Exponent ML, Entropy Ratio, and Entropy Ratio Asymmetry are not included in this summary analysis. Data used to inform this table are reported in Appendix B.

3.3.1. Difference Testing and Intra-Class Correlation for Gait Measures

The results of Table 6 suggest that an adequate median of stabillity, asymmetry, and temporal gait measures can be obtained from one day of data in this sample, as there were not significant differences between gait measures at those wear durations and all ICCs were strong. However, for 95th percentile only 80% of the gait measures have strong ICCs (RMS_AP and stance duration are weak, see Table B3 for detailed results). For CV we observed only 60% of gait mesures have strong ICC from one day (Double Support Duration, Stance, Stride, and Swing Duration are weak) and one significant difference (Swing Suration). Increasing the wear duration to two days eliminated these weak ICCs, aside from one weak CV of Swing Duration ICC, and the significant difference for the temporal, stability and asymmetry gait measures relative to baseline in this sample of

PwMS. Notably, as we examined the comparisons of wear durations of one week and longer, we saw that some gait measures exhibit reduced ICC strength (significant differences are not detected) for wear durations of 4 and 5 weeks. This reduction in ICC is still observed in bootstraped samples, suggesting a reduction in sample size is an unlikely explanation.

Interestingly, we found that at least two days of monitoring are needed for entropy ratio measures (Entropy Ratio, and Entropy Ratio Asymmetry) and that one full week is needed for Lyapunov exponent measures (Lyapunov Exponent AP, Lyapunov Exponent ML) of gait stability in PwMS (see Table B3). These gait measures likely required longer wear durations because they can only be computed from walking bouts of at least 30 (entropy) or 60 (lyapunov exponent) seconds, effectively limiting the number of observations of these measures each day. Comparing weekdays to weekend, we observed strong agreement between timeframes, however, we saw the Lyapunov exponent features and double support duration did not exhibit strong ICCs (see Table B3). With the presented information we cannot speculate on why these differences were obervered.

3.3.2. Difference Testing and Intra-Class Correlation for Postural Sway Measures

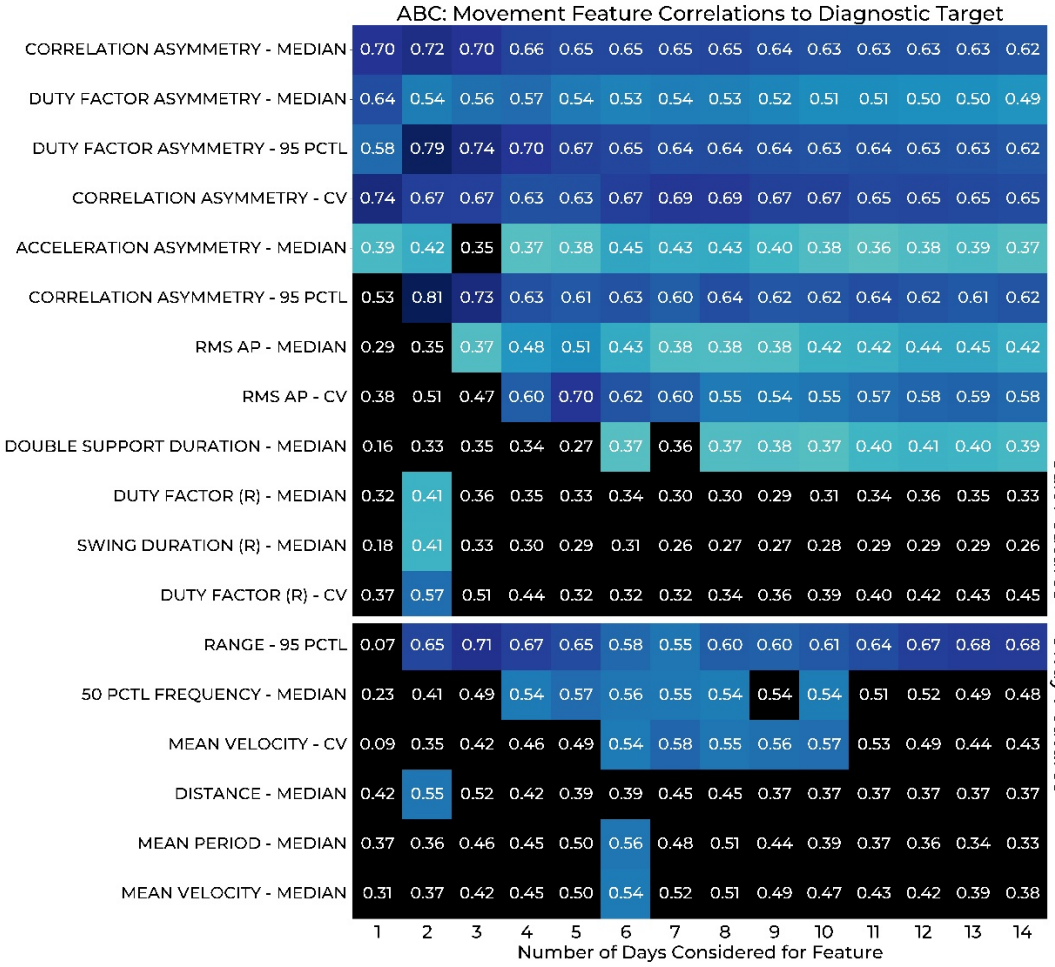
For postural sway measures, the results of Table 6 indicate that sway may require longer durations than were required for gait. Only 38% of measures show strong ICCs for their median (46% for 95th percentile and 8% for CV) and 23% show differences in their CV for one day of data. Two days of data improved the results (no significant differences, median ICCs all strong, 92% of 95th percentile and CV ICCs are strong), and all measures had strong ICCs and no significant differences after three days. The features that fail to show a strong ICC for two days of data are 95th percentile of Sway Area and CV of the Mean

Period. Moving to longer comparisons, including three days, and one week to longer periods, we found the data remains consistent with no significant differences aside from some features demonstrating weak ICCs for CV comparing one-week to longer durations. Unlike the gait measures, we did not oberserve any difference between weekdays and weekends. We speculate that postural sway measures require more data than gait measures because the sway features have a wider distribution of values and/or we only consider standing bouts that were 30 seconds or longer, thereby reducing our observations of these features. The bootstrapped power analysis suggests that the observed ICCs would slightly increase (<10%) with a larger sample size, however this small increase would not change our conclusions regarding the number of days required.

3.3.3. Correlation of Gait and Sway Features to PRMs

The results of Stage 3 of the wear duration analysis are reported in the heat maps of Figures 6 and 7 where gait and postural sway measures are correlated with PRMs of balance confidence (Figure 6) and fatigue (Figure 7). Significant correlations are indicated with colored boxes while those in black do not reach significance. Only features that display significance for at least one wear period are depicted.

Figure 6: ABC: Movement Feature Correlations to Diagnostic Target



Spearman correlation between Activity-Specific Balance Confidence (ABC) and measures of gait and postural sway by number of sensor-wear days. Cells shaded in blue represent significant correlations and cells shaded in black were not significant. PCTL: Percentile; CV: Coefficient of Variation; RMS: Root Mean Square; AP: Anterior-Posterior; ML: Medial-Lateral.

Considering the correlations between the measures of gait and postural sway and balance confidence (Figure 6), we saw the amount of time required to establish a steady significant correlation varies by feature. Both the median and 95th percentile of gait asymmetry measures provided reliable correlations with only one day of data and the strongest correlations were seen within the first two days of data for most. In the sway features, we saw two days were required to establish a significant correlation with Range,

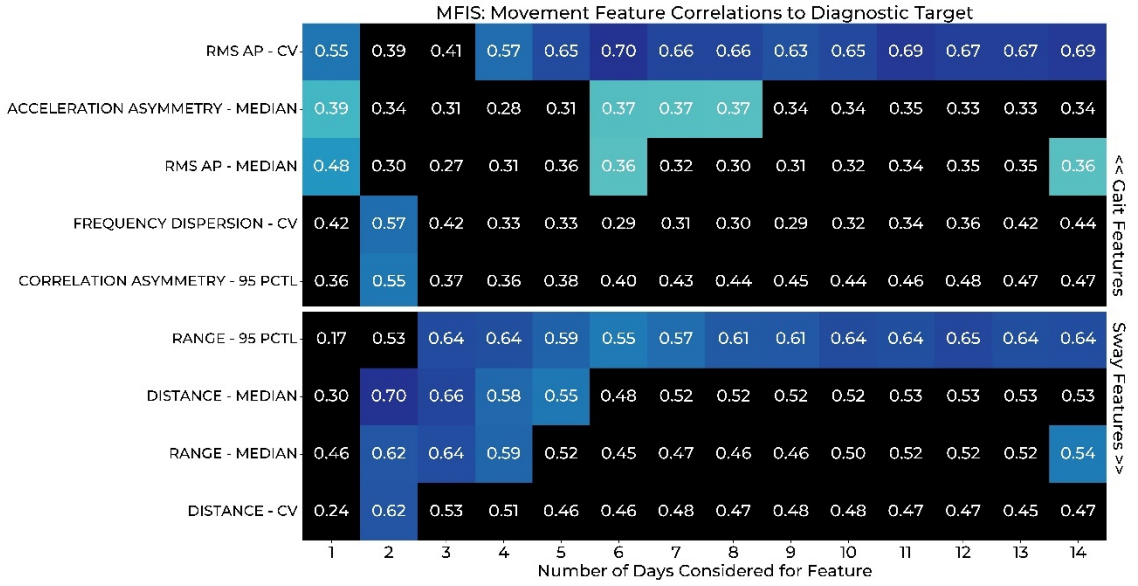
and four days were required to estalish significant correlations with the 50th percentile frequency. The fluctuation of significance seen in other features may suggest these features were not as consistent across a longer period of time. Interestingly, most significant correlations with the ABC survey occurred with the median or 95th percentile of the features, not the CV. This may suggest that the ABC survey is not sensistive to variation in these gait and sway features and instead is related to more extreme values and typical values of participants during walking and standing activities.

In contrast to the PRM of balance confidence, for the correlations between the measures of gait and postural sway and PRM of fatigue (Figure 7), we did not find a steady nor significant relationship between MFIS and gait asymmetry. As with the ABC, Range is reliably established with only two days of data. Two new consistent relationships that emerged from this analysis are sway Distance and the gait feature RMS AP, which provide reliable correlation after at least two and four days of data, respectively.

Considering the correlation results to both PRMs, unsurprisingly we find different results for each survey. This idea makes sense because we would not expect all features to have the same relationship between balance confidence and fatigue, or other clinical comparisons. Taking these differences into account, we considered a feature reliable when a significant relationship appears with either survey (ABC or MFIS) and the strength remains similar in the following days. With these criteria, we would consider the gait asymmetry features to be clinically valid with one day of data based on the ABC PRM results, even though the MFIS relationships were less clear. An example of a relationship we would not consider valid is the correlation between the 95th percentile of correlation

asymmetry and MFIS, which is significant at two days and then drops out of significance with an approximate 0.18 drop in the correlation coefficient as well, as seen in Figure 3.

Figure 7: MFIS: Movement Feature Correlations to Diagnostic Target



Spearman correlation between Modified Fatigue Impact Scale (MFIS) and measures of gait and postural sway by number of sensor-wear days. Cells shaded in blue represent significant correlations and cells shaded in black were not significant. PCTL: Percentile; CV: Coefficient of Variation; RMS: Root Mean Square; AP: Anterior-Posterior; ML: Medial-Lateral.

3.3.4. Analysis of Factors Impacting Wear Duration

Results from the proposed three-stage wear duration analysis suggest that wear duration is dependent on both the physiological measure being considered (analytical validation aspects) and the underlying disease state (clinical validation) that was being assessed. Specifically, in considering the results of Table 6, there seemed to be support for our hypothesis that wear duration is impacted by how often one observes a given measure during a day and the inherent variability of the measure. For example, the Lyapunov exponent-based measure of gait stability has very few observations per day relative to more traditional gait measures and takes almost a week of data rather than only two days.

Similarly, postural sway measures inherently have more variability than gait measures and require more days of data. However, to address our hypothesis more directly, we also present the regression results of Table 7. Here, we regress variability (log-transformed CV), number of observations (count), and their interaction on wear duration. All predictors, including an intercept, were at least trend-level significant and the model explains over 40% of the variance in wear duration. Based on the model coefficients, one can see that our hypothesis is supported. Controlling for count, an increase in measure variability will yield a subsequent increase in required wear duration. Similarly, controlling for variability, a decrease in the number of daily observations will yield a subsequent increase in required wear duration. Interestingly, the trend-level interaction term suggests that as variability increases, the relative impact of the number of daily observations on wear duration decreases.

Table 7: Wear duration time regression analysis

Coefficient	Estimate	Standard Error	P-Value
Intercept	4.41	0.53	**< 0.01**
Log CV	0.86	0.41	**0.047**
Count	-0.011	0.0029	**< 0.01**
Interaction (Log CV * Count)	-0.0040	0.0020	0.054
R-Squared: 0.46; Adjusted R-Squared: 0.39			
Number of Observations: 27			

[7] Regression analysis of days required for each feature using the log of the coefficient of variation (CV) and feature count computed from two days as predictors: Days Required ~ 1 + Log CV + Count + Interaction. Bolded p-values represent significant predictors at a 0.05 significance threshold.

3.4. Discussion

We have proposed a three-stage analysis for determining the wear duration required for capturing impairment and its variability with wearables that aligns with current best

practices for developing digital medicine technologies that are fit for purpose. The process was used to examine data from a sample of PwMS who have mild to moderate balance and mobility impairment. We now discuss these results, place them in the context of existing literature, and suggest next steps for future researchers.

Results from the proposed three stage wear duration analysis are more nuanced than expected and illustrate that each stage provides unique insights. The results of Stages 1 and 2 (difference testing and ICC, Table 6) suggest two days of monitoring are sufficient for most gait and sway features in this sample of PwMS. These results align with prior work in older adults that suggests that two to three days of data are required for measuring gait speed [138]. However, some measures, like Lyapunov exponent-based gait stability, can require up to a week of data in this sample. The results of Stage 3 (correlation with PRMs) suggest that two to three days are needed to find relationships with balance confidence and a week may be needed to find significant relationships with fatigue. Collectively, these results suggest that wear duration is dependent on both the measure being considered and the underlying disease state being addressed. This likely means that studies that use only one method to determine the necessary wear duration (e.g., [138], [143]) may not be capturing the full picture and could either be collecting data for too long, wasting resources and increasing patient burden, or not long enough, yielding unreliable measures or missing key relationships with the underlying disease state being monitored.

Regression results support our hypothesis that wear duration is impacted by how often a given measure is observed during a day and the inherent variability of the measure. These results could be used to inform deployment of wearables for characterizing digital biomarkers of impairment in several ways. First, given estimates of the number of

observations of a measure and its variability, the model could be used to predict an estimate of necessary wear duration. These parameters could be estimated from a small study (two days of remote wear based on the results presented here), or potentially from values reported in the literature. For example, remote studies of gait often report number of observed walking bouts each day and variability of the associated gait measures (e.g., see supplementary material of [138]) which could inform this estimate. This result could also be used to inform protocol changes to reduce the number of required days. For example, if a measure like Lyapunov exponent-based gait stability is critical for a given application, the protocol could include asking participants to engage in a certain number of longer duration walking bouts each day to increase the number of observations.

The proposed approach is in line with developing best practices for ensuring that digital medicine technologies are fit for purpose. As we push to realize the promise of digital medicine enabled by remote patient monitoring with cutting edge wearables, it is critical that the associated measures of impairment are appropriately validated and in the intended patient populations. As demonstrated by the results of Table 6 and Figures 6 and 7, the wear duration analysis presented herein is a key aspect of this validation as the resulting conclusions are impacted by the wear duration considered. Moreover, this multifaceted approach is important as each aspect provides slightly different information and can lead to different conclusions. Importantly, this analysis impacts every application of wearables in digital medicine and can inform the use of these technologies for informing clinical care or as novel endpoints in clinical trials. In the specific studies of gait and postural sway, these results suggest that studies aiming to capture remote assessment of variability likely need only 2-3 days to adequately assess these parameters per monitoring

period. This finding may alter study design to include several 2-3 day snap shots over a long period of time instead of one long block of monitoring.

There are several limitations to this study, including the limited sample size, and lack of demographic and geographic diversity of the PwMS. We suspect some of the correlations that alternate between significant or not may resolve to a cleaner relationship and firmer conclusions with a larger sample size. Additionally, we do not how long the two-to-three-day monitoring period is valid for. Particularly in PwMS where we observe symptom fluctuations, studies should be conducted to analyze the change between several two-to-three-day monitoring periods occurring throughout the year. Lastly, our approach requires the collection of a long baseline monitoring period, which may be prohibitive in some populations. In these circumstances, using the regression approach to determine necessary wear time may be more appropriate.

3.5. Conclusion

Herein, we present an analysis framework designed to establish a minimum duration of wearable sensor data needed to estimate features in the free-living environment. This approach combines previously used methods, difference testing and intra-class correlation, with an analysis of correlations to PRMs. In the present study, we employ this method to find that the intra-feature variance between two days in PwMS compared to one week of data is low, however, if the desired outcome of the study is strong correlations with clinical assessments and surveys, a longer monitoring period is likely needed for optimal results. Regression results also reveal necessary wear time is significantly related to number of observations and variability.

CHAPTER FOUR: CHEST-BASED WEARABLES AND INDIVIDUALIZED DISTRIBUTIONS FOR ASSESSING POSTURAL SWAY IN PERSONS WITH MULTIPLE SCLEROSIS

4.1. Introduction

Balance and mobility impairment affect a wide variety of individuals including older adults and those living with musculoskeletal and neurological disorders [163]–[167]. Persons with multiple sclerosis (PwMS) are particularly impacted [168]–[170]. Multiple sclerosis (MS) is associated with progressive demyelination of central nervous system axons which leads to delayed or altered nerve communication and subsequent sensory impairment, motor impairment, fatigue, and reduction in postural control. These impairments are often most impactful during balance challenging activities [171]. As a result, over half of PwMS fall in any given three-month period, similar incidence to 80 year-old adults [172], however symptoms of MS typically manifest around age 30 years [173]. This heightened risk for falls impacts quality of life and creates a long-term care burden [174].

Balance impairment is typically assessed with subjective patient reported measures (PRM) [22], non-instrumented balance assessments [175], [176] and/or balance assessments using force platforms [177], [178]. Force platforms have emerged as the gold standard for postural sway analysis, which considers objective movement features captured during a period of standing for characterizing balance impairment [177]. Studies utilizing force platforms have been able to distinguish between impaired individuals and healthy controls [177] and classify the fall risk of older adults [179] and PwMS [59]. 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 accelerometer [177], [180], [181]. Sacral sensor-derived postural sway measures have been used to classify fall risk in PwMS [180], distinguish between disease states [59], [177], [182] and to augment current assessment techniques [183], thereby achieving similar clinical utility to the force platform.

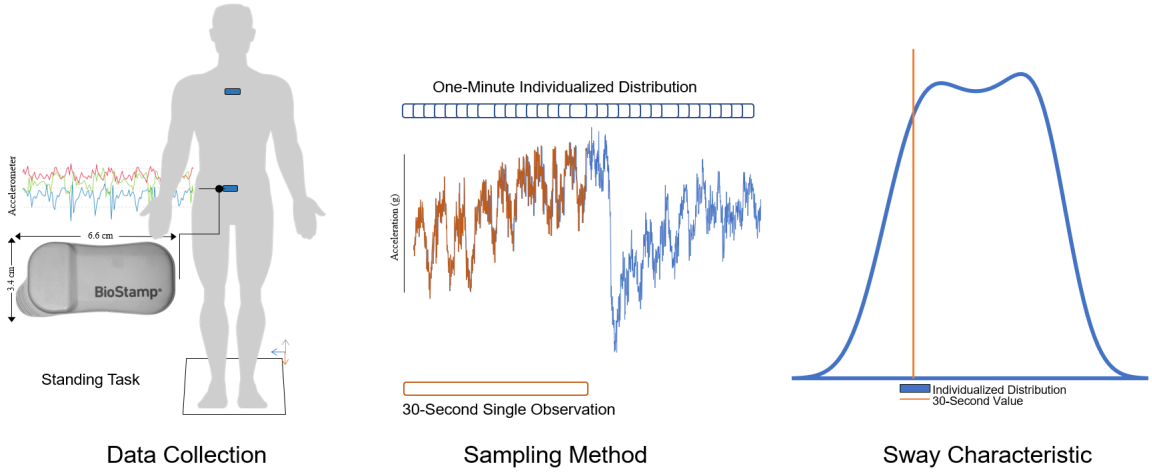
New chest-worn sensor patches are increasingly being used clinically to capture cardiac electrophysiology over long periods of time, replacing more cumbersome Holter monitors [152], [184]. In addition to being less cumbersome than Holter monitors, these chest-worn sensors may be easier to apply in remote settings compared to a sacral sensor when used for movement analysis. Cardiac measures, such as heart rate, have been shown to relate to disease severity and fatigue in PwMS [185] and may inform new measures of fatiguability [186]. Wide adoption of these devices, which often include accelerometers, may also provide an opportunity for expanding our ability to capture measures of postural sway using these devices, but chest-worn accelerometers have not yet been validated for this purpose. Validation of chest-based sway measures would allow postural sway to be assessed from a growing class of sensors and in studies not conducive to traditional force plate or sacrum measurements.

The purpose of this study is to validate the measurement of postural sway features gathered from chest-worn accelerometers. We compare chest-based measures to current measurement techniques to establish current validity. Then we compare the clinical significance of chest and sacrum-based measurements. We also introduce a new method of computing postural sway features by generating a distribution of feature values, which allows more nuanced relationships to be found.

4.2. Methods

In addressing these objectives, we first establish concurrent validity of postural sway measures derived from a chest-worn accelerometer by comparing them to gold-standard force platform measures. We then examine the clinical significance of these measures by investigating relationships to PRMs and determining which measures can detect differences in tasks and between PwMS at high and low risk for experiencing a fall based on their prior fall history. We first present the data processing methodology leveraged for computing postural sway measures and then describe how we establish concurrent validity and clinical significance.

Figure 8: Process overview of Individualized Distributions (ID) and a 30-second single observation (SO) methods.



Data were collected using wearable sensors located on the chest and sacrum during various standing tasks. Features were computed using ID or SO method. The resulting feature displays the value of an example feature computed using the standard SO method on top of the distribution obtained from the ID method.

4.2.1. Data Processing for Postural Sway Parameters

Wearable accelerometer data were gathered from the chest and sacrum locations using Biostamp nPoint® (Medidata) sensors (62 Hz ± 16G, see Figure 8). The chest sensor

was secured to the sternum, just below the sternoclavicular joint and the sacrum sensor was placed between or just above the posterior superior iliac spine. Following a standing calibration activity which allowed projection of accelerometer measurements along anatomical directions, the acceleration data were down sampled to 31.25 Hz and a 4^{th} 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. Thirty-second epochs of these data were used to compute 15 postural sway features. These features included thirteen features from Mancini et al [177]: 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) [187], and Lyapunov Exponent (LyExp) [182], [188].

4.2.2. Concurrent Validity

4.2.2.1. Subjects and Protocol

For establishing concurrent validity, we considered data from 16 PwMS (4:12 Male:Female, mean (standard deviation) age 50.6 (10.5) 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 several functional assessments in a supervised laboratory setting. For this analysis, we consider data from a standing task where participants stood on a force platform for 30 seconds with their eyes open and with their feet a comfortable width apart.

Wearable sensor data were collected as described above and ground truth center of pressure (COP) data were collected with an AMTI Force Platform (1000 Hz) that was down sampled to 100 Hz and filtered using a 4^{th}-order, zero-phase Butterworth low-pass filter with a cutoff frequency of 10 Hz. Participants were also asked to complete the following PRMs: Activity Specific Balance Confidence (ABC) [22], Modified Fatigue Impact Scale (MFIS) [23], Multiple Sclerosis Walking Scale (MSWS) [25], and Patient-Determined Disease Steps (PDDS) to characterize the sample [153]. The mean (standard deviation) of the PRMs for this cohort were as follows: ABC 80.4 (19.6); MFIS 31.7 (16.5); MSWS 20.3 (8.1); PDDS 0.93 (1.1). This protocol was approved by the UVM Institutional Review Board (CHRMS 21-0401).

4.2.2.2. Statistical Analysis

Gold standard values of the aforementioned postural sway parameters are computed from force platform COP data. From the same trial, we also computed the sway features from accelerometer data collected from the previously validated sacrum location and our proposed chest location. Sacrum findings will be used to contextualize results and to enable comparison to previous sacrum sensor validation studies.

Relationships between the force plate and sensor features are established using Spearman correlation (significance level of 0.05) as done in previous validation studies [177]. Sensor-derived features demonstrating a significant association with the gold standard parameters are considered valid at the recorded location.

Spearman correlation between features within a measurement type and location are also considered to characterize the independence of the postural sway feature set. We considered features to be linearly dependent at significant correlation of |0.7| or higher.

4.2.3. Clinical Significance

4.2.3.1. Subjects and Protocol

For examining clinical significance, we consider data from a separate, larger sample of 39 PwMS (21:18 fallers:non-fallers; 12:27 Male:Female, mean ± standard deviation age 51 ± 12 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." This study has been previously described in detail and the data are publicly available [189]. Participants were asked to complete various activities of daily living, several PRMs, and a neurologist administered Expanded Disability Status Scale (EDSS) [154]. In this analysis we utilize chest and sacrum acceleration data from three two-minute standing trials, eyes-open comfortable standing, eyes-closed standing, and tandem standing, where participants were asked to stand with their feet in a straight line. We also utilize the ABC, MFIS, and MSWS PRMs. The mean and standard deviation of these assessments is broken down between fallers and non-faller in Table 8. This protocol was approved by the University of Vermont's Institutional Review Board (CHRMS 18-0285).

Table 8: Subject Demographics

Assessment:	N	Age	Sex	ABC	EDSS	MFIS	MSWS
Fallers	21	56.0 (9.1)	5M 16F	75.0 (18.8)	3.3 (1.4)	39.8 (17.9)	55.0 (22.3)
Non-Fallers	18	45.0 (12.9)	7M 11F	91.4 (15.5)	2.3 (1.0)	29.2 (16.7)	27.5 (11.5)

Mean (standard deviation) of assessments and patient reported measures partitioned by fall status. ABC: Activity-Specific Balance Confidence; EDSS: Expanded Disability Status

Scale; MFIS: Modified Fatigue Impact Scale; MSWS: MS Walking Scale; N: number of subjects in group.

4.2.3.2. Statistical Analysis

Following traditional approaches, we calculate the outlined features of sway based on sensor data from the first 30 seconds of each two-minute balance assessment trial (Single Observation - SO). We also considered postural sway parameters derived from subject-specific distributions created by computing each feature for a sequence of 30 second windows offset by five samples (0.16 seconds) over one-minute of data from each balance assessment (Individualized Distribution Method - ID). We then extracted the 5th percentile, 25th percentile, median, 75th percentile, 95th percentile, and standard deviation to characterize each distribution. These approaches result in one feature per sway metric in the SO method and six distribution features of each sway metric for ID. For this reason, we consider a significant result for a sway measure if one or more of the ID features demonstrate significance. This process is outlined in Figure 8, where the difference between the two computational approaches is evident.

Wilcoxon rank sum tests [190] were used to detect differences in medians between normal and eyes-closed standing, normal and tandem standing, and eyes-closed and tandem standing. Significant differences indicate that a given sensor-derived measure of postural sway is sensitive to expected differences in balance performance (i.e., eyes open > eyes closed > tandem). This same approach was used to determine if these sensor-derived measures of postural sway were significantly different between fallers and non-fallers in the eyes-open, eyes-closed, and tandem standing tests. Effect size of differences is assessed with Cohen's D [191]. Finally, Spearman rank correlation is used to identify significant associations between sensor-derived measures of postural sway and PRMs. These were

examined for features derived using both the SO and ID approaches to determine how sensitive each approach is to MS-specific balance impairment.

4.3. Results

4.3.1. Concurrent Validity

Five chest- and sacrum-derived postural sway features were significantly and positively correlated with the force platform (Table 9). Five postural sway features were also significantly correlated between chest and sacrum sensors. Jerk had the strongest correlation among all comparisons (r = 0.71-0.91), followed by Range (r = 0.60-0.94) and Distance (r = 0.58-0.74). Significant, moderate correlations were also found for FD, MF, and ApEn.

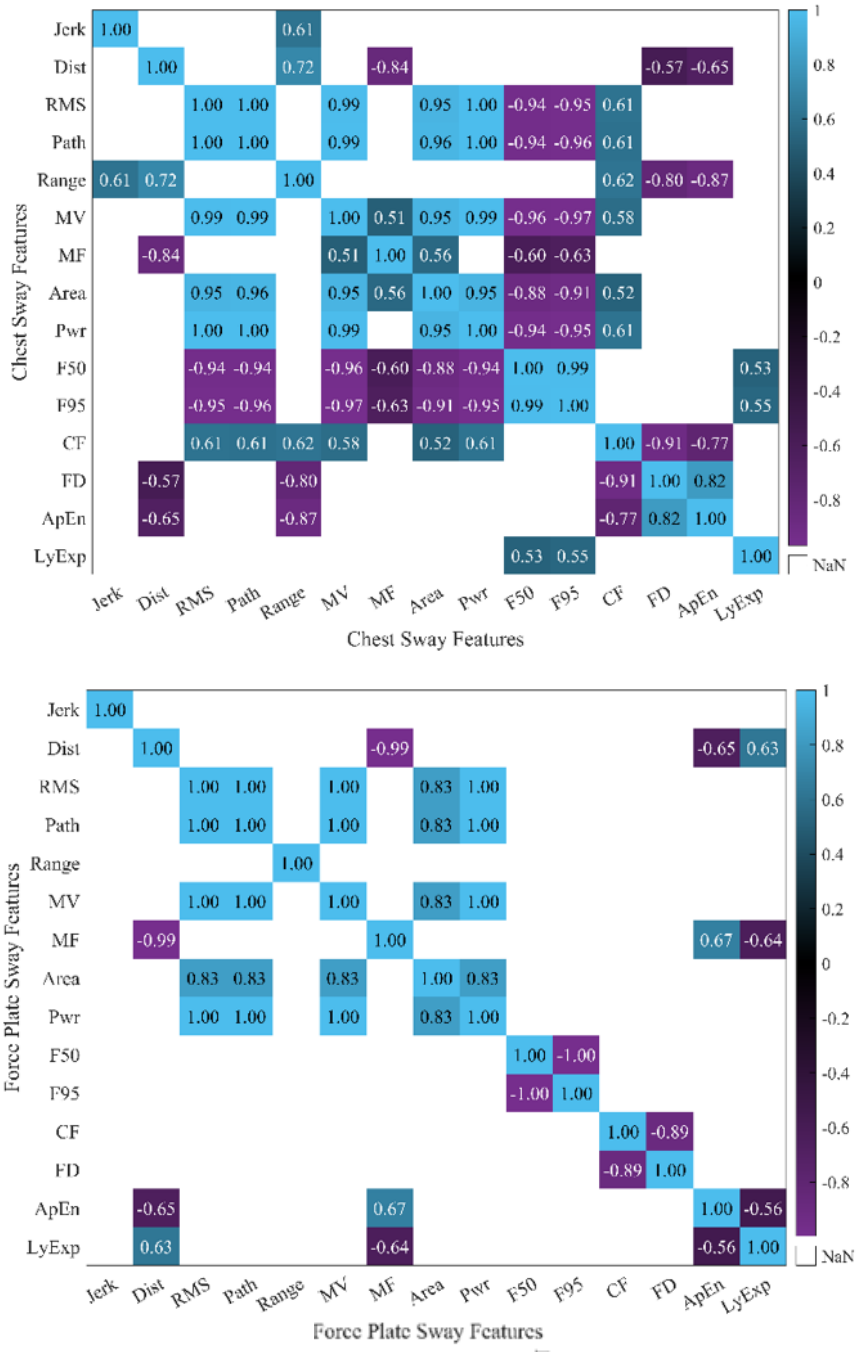
Chest-derived and sacrum-derived postural sway features have more collinearity between features compared to the force plate derived features (see Figure 9 for heatmaps illustrating significant associations between features measured with chest sensors – top, and a force plate – bottom). In the chest sensor, sacrum sensor, and force platform data, RMS, Path, MV, Area, and Power are highly positively correlated and Dist and MF are highly negatively correlated. In the chest and sacrum sensor data, F50 and F95 are highly positively correlated whereas they are highly negatively correlated in the force plate data. In the chest sensor data, FD and ApEn are positively correlated and in the chest and sacrum sensor data, CF and Range are positively correlated. The high number of features that are colinear within all three instrumentation setups suggests that while this feature set is well established in the literature [50], [59], [134], [180], [192] many of the features may contain very similar information, and the number of features reported may be able to be reduced, at least in this cohort.

Table 9: Eyes Open Postural Sway Feature Correlation

	Chest & FP		Sacrum & FP		Chest & Sacrum	
Feature	r	p	r	p	r	p
Jerk	0.88	< 0.01	0.71	< 0.01	0.91	< 0.01
Dist	0.71	<0.01	0.74	< 0.01	0.58	0.02
RMS	-	-	-	-	-	-
Path	-	-	-	-	-	-
Range	0.60	0.01	0.74	< 0.01	0.94	< 0.01
MV	-	-	-	-	-	-
MF	0.78	< 0.01	-	-	-	-
Area	-	-	-	-	-	-
Pwr	-	-	-	-	-	-
F50	-	-	-	-	-	-
F95	-	-	-	-	-	-
CF	*0.44*	*0.09*	0.59	0.02	*0.43*	*0.10*
FD	0.62	0.01	0.63	0.01	0.51	0.04
ApEn	-	-	-	-	0.53	0.03
LyExp	-	-	-	-	-	-

Significant correlations of postural sway features amongst sensors and force platform (FP) comparisons. Results approaching significance ($0.05 < p < 0.10$) are italicized.

Figure 9: Heat maps illustrating significant correlations between postural sway features derived from the chest sensor (top) and force platform (bottom).

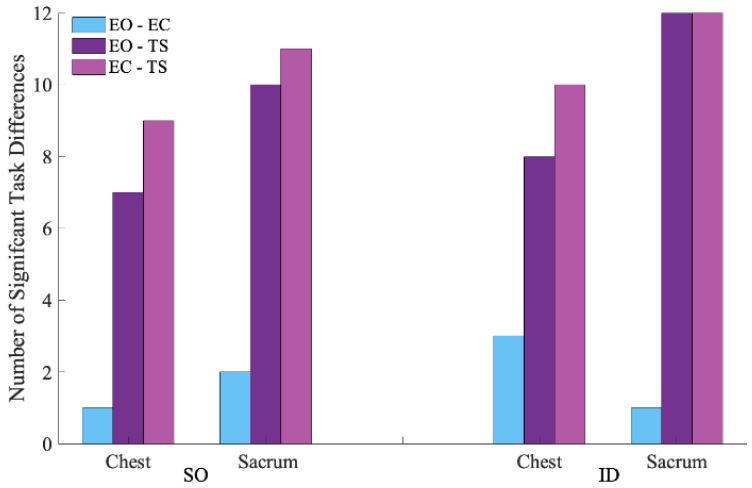


Correlations of features from chest and force plate data to themselves. Boxes are colored based on correlation strength and direction (positive/negative). Non-significant correlations are shown in white. $\alpha = 0.05$.

4.3.2. Clinical Significance

As shown in Figure 10, the ID method found more significant differences in features than the SO method between tasks for both chest and sacrum sensor locations. These tasks were expected to elicit varying levels of balance performance in this cohort such that performance during the eyes open task would be significantly better than during the eyes closed task and both would be significantly better than during the tandem standing task. As expected, significant differences in postural sway features were observed between each of the tasks. The largest number of significant differences were found between the eyes-open and tandem standing tasks and between the eyes-closed and tandem standing tasks. In contrast, few significant differences were found between the eyes-open and eyes-closed tasks. The effect sizes of these differences range from 0.42-0.80 when comparing eyes open to eyes closed, 0.21 to 1.91 comparing eyes open to tandem standing, and 0.21 to 1.81 when comparing eyes closed to tandem standing. The larger number of differences found in the ID method mean the effect sizes range from comparable to those of the SO method for the strongest effects to much weaker in the more nuanced differences.

Figure 10: Number of significant differences in features between tasks in chest and sacrum sensors for the Single Observation (SO) and the Individualized Distributions (ID) methods.



Tasks: eyes-open (EO), eyes-closed (EC), and tandem standing (TS), are expected to elicit significantly different postural sway performance. For the ID method, multiple significant differences within each postural sway feature's statistics are not included (i.e., significant median and standard deviation differences of jerk count as one significant correlation, not two). $\alpha = 0.05$.

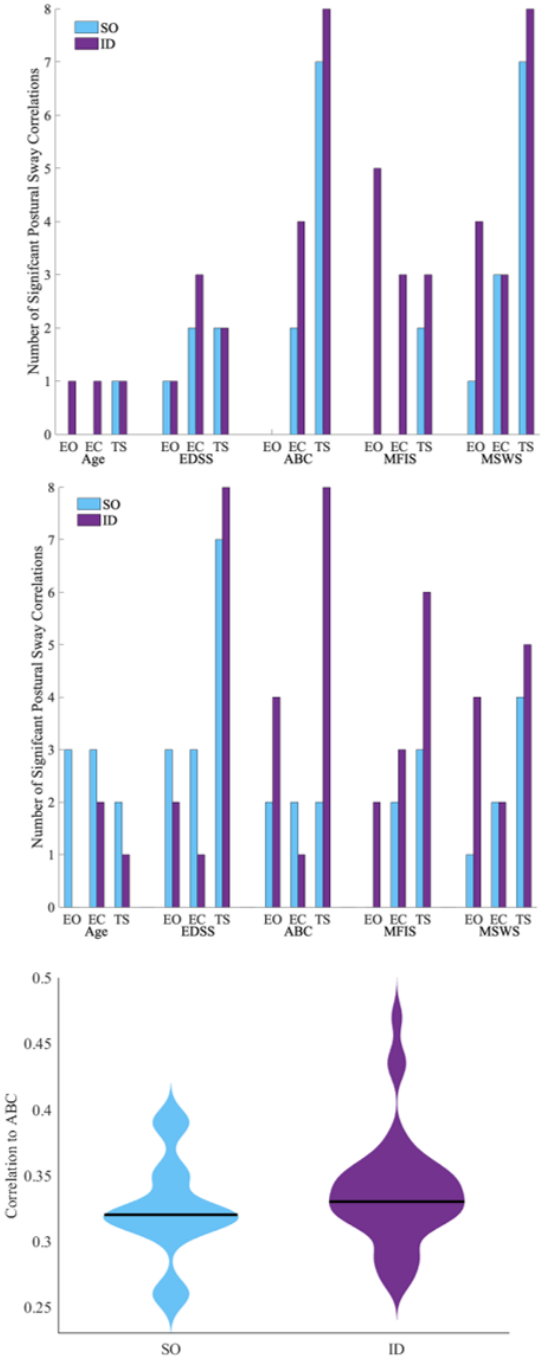
Considering faller and non-faller groups, the most differences were detected with the ID method during the tandem standing task with the chest sensor location. Jerk, MV, Area, CF, and FD all demonstrated statistically significant differences with moderate effect sizes ranging from 0.29 to 0.60. Additional details on these results can be found in Table C5. Using the ID method for the eyes closed and tandem standing tasks, Range and MV were found to be approaching significance for the sacrum and chest sensor locations. Only Range computed from the sacrum sensor during the eyes-closed task was found to be approaching a significant difference in the SO method.

The greatest number of significant or trend level relationships between postural sway features and PRMs were consistently found from the ID method. In the ID method, the edges of the distribution are important, as demonstrated by our strongest correlations coming from the 5^{th} percentile of these distributions. The chest sensors with the ID method

found significant and approaching significant relationships in at least one of the six statistics between MF and four of the five PRMs (Figure 11, Tables C1-C4). In the SO method, no significant relationships were found between MF and the PRMs.

Considering correlations of postural sway features from both computational methods, the chest sensor features had more associations with MSWS while the sacrum sensor had more associations with EDSS and MFIS (see Figure 11). For example, during the tandem standing task the sacrum sensor had seven significant EDSS correlations using the SO method and eight using the ID method, while the chest sensor had two significant EDSS correlations using both methods. While the sacrum sensor had several additional relationships between postural sway features and PRMs, as compared to the chest sensor, the overall level of association was similar between the two sensor locations.

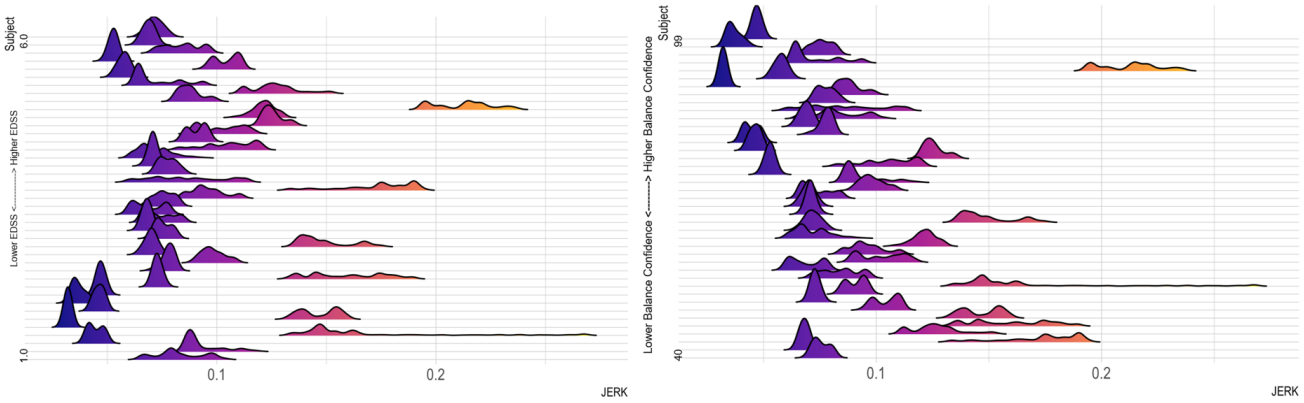
Figure 11: Significant correlations between age, EDSS, ABC, MFIS, and MSWS and postural sway features.



Number of significant correlations between age, EDSS, ABC, MFIS, and MSWS and postural sway features for chest (top) and sacrum (middle) sensor data for the Single Observation (SO - blue) and the Individualized Distributions (ID - purple) methods. Range of correlations to ABC during tandem standing with median indicated by black line (bottom). For SSD, multiple significant correlations within each postural sway feature's statistics are not included. Tasks are indicated as eyes-open (EO), eyes-closed (EC), and tandem standing (TS). $\alpha = 0.1$.

Comparing standing tasks, tandem standing consistently demonstrated the greatest number of significant relationships between postural sway features and PRMs across both computational methods. For example, considering the chest sensor location, the ID method had eight postural sway features significantly associated with ABC in the tandem standing task, only four from the eyes-closed task, and no significant associations from the eyes-open task. The bottom of Figure 11 demonstrates the range of correlations between ABC and tandem standing for each computational method and shows the increased median and overall stronger correlations observed using the ID method. As a result of the ID method, the strongest correlation increased by 21%. The remaining correlation information is located in Tables C1-C4. One of these highly correlated features was jerk. Figure 12 depicts jerk ID computed for each subject during the tandem standing task. From these distributions, it is clear that there is significant variability in the Jerk across the collection of 30-second epochs sampled from the one-minute of tandem standing task data. There is also a clear relationship between ID-computed Jerk and EDSS (left) and ABC (right) such that subjects with a higher EDSS and lower ABC score (less balance confidence) show higher jerk. The strongest correlations observed were between the 5^{th} percentile of Jerk with EDSS and ABC were 0.54 and 0.47, respectively.

Figure 12: Individualized distributions of Jerk with subject sorted in order of EDSS on left and ABC on right.



Individualized distributions of Jerk found using the chest sensor data for the tandem standing task with subjects sorted in order of Expanded Disability Severity Score (EDSS) shown on the left and Activities Specific Balance Confidence Score (ABC) shown on the right. Higher EDSS and balance confidence subjects are at the top of the figures. The strongest correlations were 0.54 and 0.47 for EDSS and ABC with the 5th percentile.

4.4. Discussion

The purpose of this work was to validate the measurement of postural sway features from a chest-worn accelerometer and to advance a new approach for considering postural sway data that better captures expected variability in sway performance. In doing so, we investigated the concurrent validity of this approach by comparing to gold standard force platform derived measures. We then further explored the significance and sensitivity of sway features from chest and sacrum sensors by exploring differences between standing tasks, differences between fallers and non-fallers, and correlations to PRMs in a sample of PwMS.

Our investigation of concurrent validity suggests valid measures of postural sway can be gathered from a chest accelerometer. We found that five chest-, and five sacrum-computed features were significantly correlated with force platform-derived features (Table 9). These results align with prior validations of sensor-derived measures of postural

sway, but from a new body location that has emerged as having broad clinical utility and acceptance for wearable device use, both in clinic and at home. For example, a previous study investigating the concurrent validity of a sacrum accelerometer for measuring sway found eight measures significantly correlated using a slightly larger sample with more pronounced balance impairment [177].

Our investigation of clinical significance indicates that chest acceleration-derived measures of postural sway are related to PRMs of balance confidence (ABC), fatigue (MFIS), and walking impairment (MSWS) as well as neurologist-administered measures of disability (EDSS). Significant associations were observed for features derived from each of the tasks, with the tandem standing task providing the largest number of significant relationships. This is likely because tandem standing was the most balance challenging task performed. Prior studies have shown that a person's ability to respond to postural perturbations and engage in balance challenging tasks, like tandem standing, is clinically relevant and is associated with characteristics like fall risk [193], [194].

Additionally, our analysis suggests that conducting a one-minute trial and creating subject specific distributions is more sensitive to MS-related disability than a single 30-second epoch of data as considered in traditional postural sway analysis. This was manifested in task differences (Figure 10), faller-non-faller comparisons, and PRM correlations (Figure 11). This is likely a result of the ID method's inherent ability to resolve the variability and extremes of balance performance observed during balance assessments. This is particularly apparent in the PRM correlations for data collected during the eyes-open task. Using the traditional SO method, very few significant correlations were observed (two from chest, six from sacrum), which suggests that the task may not be

providing information relevant to MS-related balance or mobility impairment. However, the ID method reveals several additional significant relationships - 10 from chest and 13 from sacrum.

The increased sensitivity of the ID method also leads to stronger relationships to PRMs. As shown in Figure 11, the ID approach yields a 21% increase in correlation strength above the SO approach. This is a substantial increase for a method that looks at the same postural sway data, just in a slightly different way by sliding a window instead on computing one value. This increase in correlation strength has a number of important impacts. Notably, the sample size required to resolve a correlation of 0.39 with 80% power is N=36. In contrast, only N=24 participants are needed to resolve a correlation of 0.47 with 80% power. Fewer required participants reduces study burden (cost, time) for participants and researchers alike.

Considering the results from Figure 9, there is significant collinearity in this set of postural sway features, despite extensive use in the literature. Combining the results of Figure 9 with Tables C1 and C2 suggests that this feature set can be reduced while capturing the same relationships. Within the set of highly correlated features, Path, MF, RMS, Power, Area, F95, and F50, we found Path, RMS, Power, and Area were only significantly related to PRMs when computed from the chest and MF, F50, and F95 were only significantly related to PRMs when computed from the sacrum. If we were to down select for future work the features from these sets that are related to the most PRMs, we would keep Area for chest measurements and F50 for sacrum. Following a similar procedure as discussed above, we could continue to reduce the feature set to obtain a minimal set of Jerk, Dist, Range, Area, F50, ApEn, and LyExp.

Consistently, we observed the features Jerk and Range to provide the strongest and most significant relationships to PRMs in addition to being able to distinguish fallers from non-fallers. However, based on our previous findings [192], the strength of these relationships is poor compared to those observed on postural sway data collected during daily life (e.g., Range and ABC had a correlation of 0.71 from free living data compared to the 0.36 observed using the ID method). This difference may be related to the well documented issue that laboratory based tests are not able to adequately capture the variability with which people, and particularly PwMS whose symptoms are known to change dramatically from day to day, move in free living environments [189], [195]. Considering the results presented herein and our previous findings, postural sway features computed from chest accelerometer data are valid, which allows for easy sensor placement if deployed in a remote setting. Collectively, these results highlight the need for an in-depth investigation of wearable-derived postural sway measures in the free-living environment.

There are several limitations to this study. First, our analysis of concurrent validity is based on a relatively small sample of 16 PwMS, compared to the 39 PwMS used to determine clinical significance, but this size is similar to those used in other postural sway validations (e.g., [134], [177]). There is a lack of demographic and regional diversity in this sample. Additionally, our proposed ID approach requires at least one minute of data and therefore may not be compatible with existing datasets or ongoing study designs where only 30 seconds of postural sway are captured. Finally, our inclusion criteria limit these studies to lower impairment individuals. We expect a more impaired sample would result in stronger signals, however, that remains untested with our current dataset.

Despite these limitations, due to the ease of placement, ability to be utilized on a growing class of wearables and validity relative to gold standard measures, chest acceleration derived features of postural sway show promise to serve as a powerful tool for understanding impairment in PwMS in a clinical setting and at home. We find these methods have similar sensitivity to previously validated measures, thereby allowing sensitive assessments of postural sway in study designs not conducive to force platforms or sacrum sensors. Moreover, the presented ID approach will likely increase the utility of these measures. Further studies are needed to assess the ability of this approach to identify fallers from non-fallers and these relationships need to be investigated in the home environment to test the viability of these methods outside of the clinic to further explore how these interesting findings may be extended to impact comprehensive care of PwMS and potentially to other conditions that impact balance and mobility.

4.5. Conclusion

We examine the use of a chest-worn accelerometer to capture measures of postural sway. We utilized a force platform to establish the concurrent validity of chest-based postural sway measures and then analyzed the clinical significance of these measures. We found sway measures from the chest were able to distinguish between standing tasks and fall status and were significantly related to PRMs of balance confidence, fatigue, and walking impairment, and a neurologist administered measure of MS-related disability in a sample of PwMS. Additionally, we presented a new means of processing postural sway features that results in greater sensitivity to task difference and fall status and exposes more associations with PRMs. These findings support the use of chest-worn accelerometers for characterizing postural sway.

CHAPTER FIVE: ASSESSING FREE-LIVING POSTURAL SWAY IN PERSONS WITH MULTIPLE SCLEROSIS

5.1. Introduction

Multiple sclerosis (MS) is a immune mediated disorder leading to the demyelination of central nervous system axons affecting an estimated 2.8 million people worldwide [196]. Subsequently, 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 disfunction and over 50% experience a fall in any given 3-month period [4], [197]. 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 [198], [199].

Postural instability and balance impairment are typically assessed with subjective patient reported measures (PRM) [22], non-instrumented balance assessments [175], [176] and/or balance assessments using force platforms [177], [178]. 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 [177]. Studies utilizing force platforms have been able to distinguish impaired individuals from controls [177] and classify the fall risk of older adults [179] and PwMS [59]. 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 [51], [177], [180], [181]. Sensor-derived postural sway measures have been used to classify fall risk in PwMS [180], distinguish between disease states [51], [59], [177], [182], [200] and to augment current assessment techniques [183],

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. Therefore, an investigation into remotely assessed postural sway is needed. Recent studies of chest accelerometer-based postural sway have found stronger relationships to PRMs from remotely collected measures compared to in-clinic assessment [51], [53]. 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 the remote parameters are significantly different and have higher variability compared to those from an in-clinic assessment [201], [202]. 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 to provide consistent context for analysis. For example, this approach has been applied to 30-second chair stand tests [203] and ten-meter walk tests [204], 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 [205]. 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

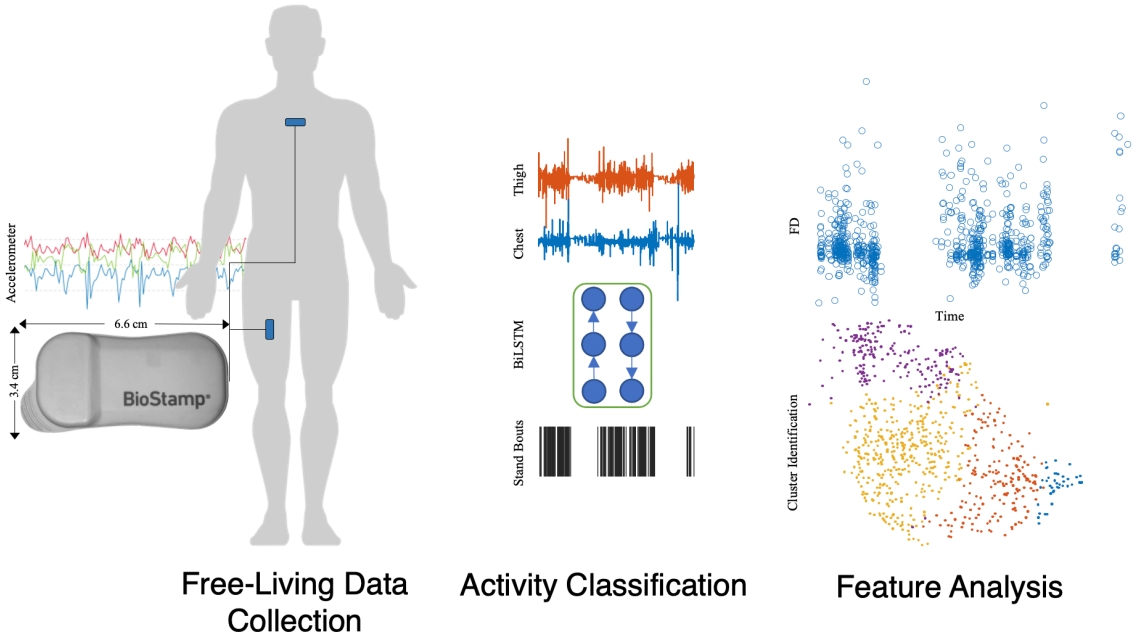
analyze in a fully unsupervised manner, but these approaches have not yet been developed.

The purpose of this work is to introduce remote postural sway as a potential biomarker for balance impairment in PwMS. In doing so, we compare postural sway features computed in-clinic and remotely, demonstrate relationships to PRMs, investigate the ability of remote postural sway to classify fall risk in PwMS, and show how data-driven clustering can help identify which remote postural sway observations to analyze.

5.2. Methods

To address the goal of introducing remotely collected postural sway as a biomarker, we collected free-living data from PwMS as shown in the first section of Figure 13. We then utilize our activity classification model to determine was activities participants are performing at home using raw acceleration, depicted in the middle section of Figure 13. first present a comparison of features to the commonly used laboratory assessment. Then we examine postural sway features computed from the standing bouts and determine which data are suitable for analysis, depicted in the last section of Figure 13. Lastly, we utilize our selected measures of remote postural sway to correlated to PRMs, classify fall risk, and attempt to contextualize our findings.

Figure 13: Data Processing Overview



Data processing overview. Free-living data collected from thigh and chest accelerometer 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.

5.2.1. 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 [202]. Participants were asked to complete several PRMs, several activities of daily living, and a neurologist administered Expanded

Disability Status Scale (EDSS) assessment [206] during a laboratory visit. Participants were then asked to complete a 48-hour monitoring period of daily life immediately following the laboratory visit. The PRMs utilized in this analysis were Activities-Specific Balance Confidence (ABC) [22], Multiple Sclerosis Walking Scale (MSWS) [207], and Modified Fatigue Impact Scale (MFIS) [208]. The in-laboratory assessment used in this analysis was a 2-minute standing balance assessment where participants were instructed to stand with their feet should 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 and the thigh sensor was on the anterior aspect of the right thigh ~25% from the knee to hip.

5.2.2. Data Processing

5.2.2.1. 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 was classified using a previously described classification framework that identifies bouts of walking, standing, sitting, and lying [53], [209]. 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 length of the typical in balance lab assessment [51], [177]. The

first minute of the laboratory standing balance assessment was used for in-lab analysis. Data was processed using the individualized distributions [51] approach where a 30-second window is slid 5 samples over the trial to create a distribution of each sway parameter.

5.2.2.2. 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 [177]: 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) [187], and Lyapunov Exponent (LyExp) [182], [188].

5.2.2.3. Data Clustering Methodology

Clustering methods provide a way to sort through the data in an unsupervised manner to find similar data. A popular method of clustering is called ‘K-Means.’ This method begins with an initial but not optimal clustering, then relocates each point to its new nearest center, updates the clustering centers by calculating the mean of the member points, and repeats the relocating-and-updating process until convergence criteria are satisfied [210], [211]. These techniques may be able to sort through the different type of data that may exist in remote analyses 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 in PwMS [212]. We will examine whether these methods increase relationships found in remote postural sway herein.

To filter the noise that is inherent in remotely collected data, we employed the aforementioned kmeans clustering [210], [211] to identify similar subsets of data. The optimal number of clusters was chosen using MATLAB's evalclusters function with DaviesBouldin criterion and Euclidean distance for 1 to 5 clusters. This was performed for each participant and resulted in an average of best number of clusters of 4. Then using the z-scores of the reduced feature set, four clusters were identified for each participant. The clusters were assigned based on the sorted centroid of the feature FD because this feature is strongly related to impairment and is a measure of the range of movements in a standing bout. Cluster 1 contains all the data from the cluster with the highest centroid of FD, cluster 4 contains the data with the lowest FD, and clusters 2 and 3 fall in between sequentially.

5.2.3. Statistical Analysis

With the postural sway features extracted from both the lab and home standing, first we used a Ranksum difference test to check for differences in medians. Effect sizes of these differences were reported by Cohen's D. Median and inter-quartile range (IQR) are also reported for each feature. Then, we computed the spearman correlations between the postural sway features and PRMs. The multiple observations of features per participant were aggregated using the following summary statistics: 5^{th} percentile (P5), 25^{th} percentile (P25), median (Med), 75^{th} percentile (P75), 95^{th} 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 [51] cross-correlations between the features, we then selected a reduced feature-set for remote analysis containing features that demonstrate correlations to PRMs and 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; however, the reduced remote feature set contained the following features: RMS, Range, Area, CF, FD, and LyExp.

To examine the clinical significance of the clusters, we computed spearman correlations to both ABC and EDSS for the data from each cluster and compared the features with a Ranksum test to lab features. 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 95^{th} percentile of each feature. For comparison, these same methods have been applied to the non-clustered data.

5.2.4. Fall Risk Classification

Using the features extracted from both in lab and remote data, the six-month fall history was used to inform classification models of fallers and non-fallers. Logistic regression (LR) and support vector machine (SVM) classification models were trained, optimized, and tested separately on both the lab and remote features. 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. Lastly, the weights of the LR model were then used to measure feature importance.

We also created fall risk classification models with features from each cluster. Both LR and SVM models were trained and evaluated using LOSOCV, and performance of each cluster was again assessed using AUC, accuracy, sensitivity, specificity, and F1 score, both with and without aggregation. Hyperparameters of both models 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 SMO solver. The model weights from the lasso regression were also considered for feature importance. The number of faller and non-faller observations for each cluster was also reported. A permutation analysis was conducted to compare the model AUC against random chance, this was done using 100 run average of classification results compared to 1000 replicates of permuted labels.

5.3. Results

All fifteen sway features computed were significantly different between the lab and all remote data as seen in Table 10. Very high effect sizes were also observed for Pwr, Path, and RMS. When correlating the remote data to PRMs, seen in Table 11, we found the strongest relationships across all PRMs with the feature 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 10: Difference Testing Between Lab and Remote Sway

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 ranksum difference testing and Cohen's D effect size ($\alpha = 0.05$)

Table 11: Correlations of Remote Sway Features to Patient Reported Measures

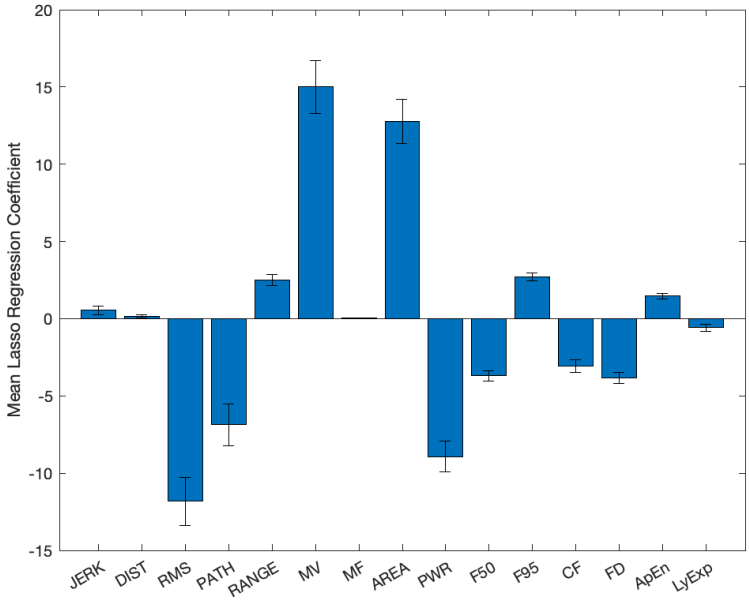
Feature	EDSS		ABC		MFIS		MSWS	
Jerk	-	-	-	-	-	-	-	-
Dist	-	-	-	-	-	-	-	-
RMS	-	-	-	-	P75	0.34	-	-
Path	-	-	-	-	P75	0.32	-	-
Range	-	-	STD	0.36	-	-	-	-
MV	-	-	-	-	Med	0.33	-	-
MF	-	-	-	-	-	-	-	-
Area	-	-	P75	0.39	*STD*	*0.45*	STD	0.30
Pwr	-	-	-	-	P75	0.34	-	-
F50	-	-	-	-	-	-	-	-
F95	-	-	-	-	-	-	-	-
CF	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	-	-	-	-

Strongest spearman correlations observed between sway features and patient reported measured. EDSS: Expanded Disability Severity Scale; ABC: Activities-Specific Balance Confidence; MFIS: Modified Fatigue Impact Scale; MSWS: Multiple Sclerosis Walking Scale; P5: 5th Percentile; Med: Median; P75: 75th Percentile; STD: Standard Deviation; Strongest correlation for each patient reported measure in italics. ($\alpha = 0.10$).

Measures of postural sway were found to be able to assess fall risk when computed on in-lab data. We observed an AUC of 0.74 fall risk classification which increased to 0.79 when we took the median of each participants classifications. This was found using a logistic regression model and the weights of the model for a feature importance are

depicted in Figure 14. To establish a baseline remote data performance, we first fit models using all the collected data. The best performing model was a logistic regression with AUC of 0.52 before aggregation and 0.44 after, suggesting the model is unable to perform any better than guessing likely due to noisy data. More model performance details can be found in Table 13.

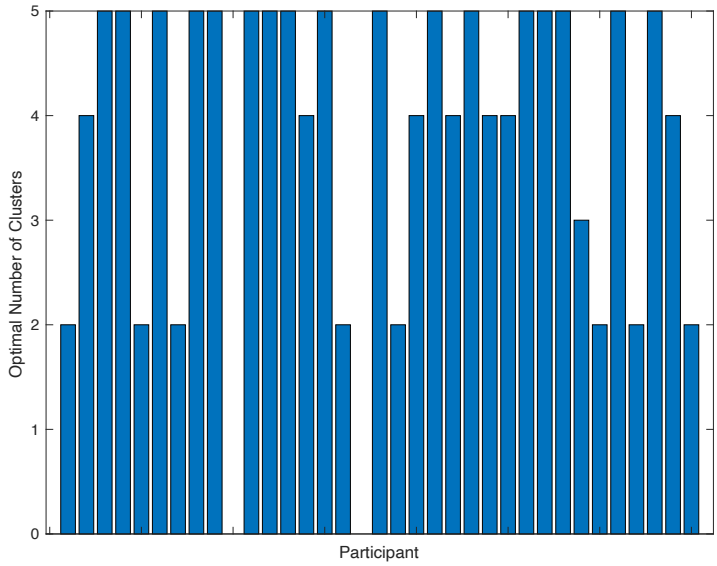
Figure 14: Feature importance of logistic regression model for in-lab fall risk classification.



Feature importance of postural sway features computed from in-lab fall risk classification.

Clustering methods were then applied to investigate whether selecting subsets of data would provide an increase over all data performance by potentially removing noise and identifying similar data. The average optimal number of clusters was found to be 4, the optimal number of clusters for each participant is depicted in Figure 15.

Figure 15: Optimal number of clusters for each participant for kmeans clustering.

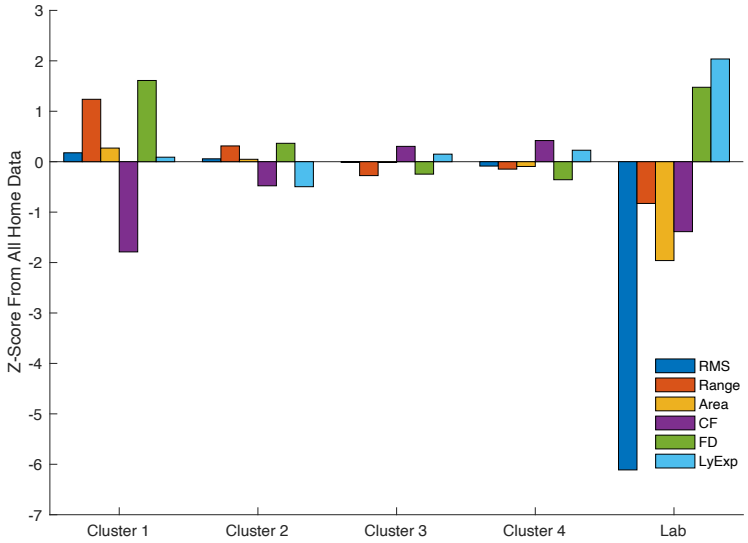


Optimal number of clusters by participant for kmeans clustering using MATLAB's evalclusters function.

Each cluster was found to establish differing relationships to PRMs. As seen in Table 12, 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 features. The feature RMS has a correlation of 0.36 and 0.43 with the median of clusters 2 and 3 respectively. The feature CF was also significantly correlated with cluster 2 ($r = -0.46$). Interestingly, while not significant, the features Range and FD, are negatively correlated with the lab data for all home data and clusters 1-3. All other features, including those from all home data, were not significantly correlated to lab-derived features. When averaging the amount of time spent in each cluster across all of the 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), however, this was not the case for clusters 2 and 3. Figure 16 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 home.

Figure 16: Z-score differences of clustered data and lab data for selected features.



Differences between clustered data and all home data. Z-scores computed from difference between clustered/lab feature and all home data.

Table 12: Correlations of Clustered Features to EDSS and ABC

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. ($\alpha = 0.05$).

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, clusters 1-4 achieve AUCs of 0.57, 0.71, 0.53, and 0.32 respectively, as shown in Table 13. 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 13.

Table 13: Fall risk classification performance by input data source of postural sway features.

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

Fall risk classification performance by input data source of postural sway features. C1 – C4: Cluster 1 – Cluster 4; LR: Logistic Regression; SVM: Support Vector Machine; AGG: Aggregation; F: Fallers; NF: Non-fallers; 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. 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.

5.4. Discussion

The purpose of this work was to introduce postural sway as a remote digital biomarker. In doing so, we compared the metrics 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 many cases. Interestingly, many of the features with high effect sizes were related to sway path, 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 [202].

Our investigation of clinical significance finds several significant correlations between PRMs and remote sway features. The strongest correlations were observed with FD for EDSS, ABC, and MSWS, while Area provided the strongest correlations to MFIS. In our previous studies, we have found few significant relationships between standard eyes-open standing and PRMs in the lab [51]. The strongest in-lab correlation we observed was -0.37 between Dist and MFIS while in this analysis we not only find a stronger relationship to MFIS with Area (r=0.45), we also find a correlation of -0.62 between FD and MSWS

when considering remotely collected sway. Based on these findings, it is clear that remote sway demonstrates clinical significance in relation to PRMs.

Using the remote and lab measures to train fall classification models, however, we find that considering all the remote data is not clinically meaningful 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 remotely. Investigating the feature importance of the well fit 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 [213]. These features may not arise as important in this analysis because we do not require our participants to perform a lying calibration and therefore cannot distinguish between AP and ML features.

When clustering methods were applied to the remote data, we found differing relationships with each cluster. Based on ranksum 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, features like CF, FD, and Range have negative correlations, 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). 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 motives using similar unsupervised methods to remove unwanted data or select data of interest in future remote analyses. A similar approach may have been able to explain the differences between fall classification performance of PwMS from gait from the lab and home [122], [202].

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 provide different results when doing so. Finally, our inclusion criteria limit these studies to lower impairment individuals. 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 deep learning methods of fall

risk classification. Studies should also be done to determine if similar clustering methods provide meaning findings in other fields of remote analysis, such as for gait.

5.5. Conclusion

Herein we examine 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 stronger correlations are found with remote data. However, lab sway features were able to accurately assess fall risk while 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. The results presented herein motivate the inclusion of postural sway as an analysis method in future remote studies.

CHAPTER SIX: CONCLUSION

6.1. Summary of Developments

The goal of this dissertation was to address challenges of remote monitoring to allow for robust assessment of gait and balance impairment related to fall risk in persons with multiple sclerosis. These challenges and questions were examined with a data analysis platform that was developed for remote analysis. This platform detects the participants' activities, such as walking and standing, and then computes the appropriate gait, sway, or other activity features and outputs the results for further analysis. This platform was designed to function with data from several studies and patient cohorts to facilitate remote analysis for others and in future work.

The introduction provided in Chapter 1 motivates and discusses falls in MS, current techniques of fall risk assessment, and how wearable sensing can be used to improve the standard of care. Specifically, this includes discussion of how remote studies may be more appropriate and accessible, however, there are few studies investigating fall risk remotely, particularly in the field of postural sway. Further, prior work is discussed and how the excellent performance achieved using in-lab assessments techniques did not translate when applied to remote data. Based on these challenges, this work aimed to answer the following questions: Does walking duration affect gait and fall risk assessment? How much data is enough? Can we measure postural sway from the chest? Is there a better way to compute sway metrics? Can remotely collected postural sway classify fall risk and demonstrate clinical utility?

Chapter two addresses the challenges regarding walking duration. The major contribution of this work is the demonstration that features of gait do change in PwMS

based on walking duration. This builds upon on the work on Storm et al. that demonstrated significant differences between lab and remote gait parameters [137]. As a result, walking duration also has an impact on fall risk classification models. In this analysis, it was found that longer walks were most similar to in-lab walking. These longer walks also provided the input data to the best performing fall risk classification models. In addition to these findings, the data set used to train these models was released publicly to address the common lack of data in clinical populations.

Chapter three addresses a common question asked by reviewers when reviewing a remote study: how much data is enough? To answer this question, this study utilized six weeks of data and compared features of gait and sway computed from shorter durations of wear. Searching through literature on methods to determine sufficient monitoring time, there was a lack of a defined methodology to do so. Therefore, this study proposed a standardized three-part method to determine an adequate monitoring period. First difference testing and intra-class correlation were used to establish analytical validation. Then clinical significance was established by correlating to PRMs. Applying this analysis to gait and sway features, we found that most measures were analytically valid in two to three days of monitoring and in some cases, longer periods of monitoring were needed to establish clinical significance. Overall, the number of days required was found to be significantly related coefficient of variation of the feature and number of daily observations.

Chapter four addresses two challenges in the field of postural sway. First, can sway features be obtained from the chest for more convenient remote monitoring. Second, is there a better method to compute sway features than first proposed in Mancini et al. where

features are computed once based on a 30-second standing period [50]. In this analysis, sway measures computed from the chest were found to be comparable to the previously validated sacrum accelerometer location. Once concurrent validity was established, chest measures were also found to relate to PRMs and demonstrate differences between fallers and non-fallers. This study also proposed a novel method of computing postural sway features called Individualized Distributions (ID). This approach required an additional 30-seconds of standing data and used a sliding window approach to create a distribution of each feature rather than a single value. The resulting distribution was found to capture more differences between fallers and non-fallers and have more numerous and stronger correlations to PRMs.

Chapter five builds upon the finding that postural sway can be assessed from the chest and only requires two days of monitoring. This study presents an analysis of remotely collected postural sway and shows its relationship to PRMs, fall risk, and different groups of data that exists in remote data. Measures of remote sway were found to provide stronger correlations to PRMs when compared to lab-derived features, however, lab-derived features far outperformed remote features when classifying fall risk. To explain this performance discrepancy, clustering was employed to identify clusters of similar remote data that may provide improved predictive power. In doing so, one cluster was able to obtain fall classification performance nearing in-lab performance while also improving upon the remote correlations to PRMs.

6.2. Implications of Work

This work answers foundational questions and challenges required to successfully design, collect, and analyze wearable sensors data in remote settings and also provides a

platform for remote analysis. Chapter two explains that the duration of the activity, gait in this case, has an impact on participant performance and prediction power. Future studies need to consider this notion when performing remote analyses. Chapter two also outlines an open-sourced dataset that was released along with the paper. This serves to help fill the clinical population data shortage and encourage other groups to share study data as well.

Chapter three has several implications for future study design. The finding that two-three days is sufficient to monitor features of gait and sway suggests longer studies may be a waste of resources. More frequent shorter observation periods may be more useful when trying to track aspects of disease like symptom fluctuations. The number of required days to adequately measure a feature was found to be related to variability and number of observations. This knowledge can be applied during feature engineering and development to ensure the selected features are appropriate for the desired study length. Additionally, this study also proposed a framework that can be used to perform similar analyses in other populations or activities.

Chapter four demonstrates that sway measures can be obtained from the far more accessible chest sensor location compared to a sacrum sensor. This allows these measures to be obtained the growing number of cardiac-focuses wearables and alleviates the remote monitoring of postural sway. In addition, this study also demonstrates that sway measures computed using individualized distributions provide greater predictive power compared to those computed in the standard approach.

Lastly, Chapter 5 builds upon previous findings and demonstrates that postural sway is a clinically significant remote digital endpoint. This is the first study to formally introduce remote sway in PwMS and other populations. The use of clustering to find

stronger relationships also suggests similar methods may be used in other fields of remote analysis. The remote analysis framework used to perform these analyses was licensed for use by Medidata Solutions and is being broadly applied to other clinical populations. Overall, this work shows clinically significant measures can be obtained remotely when proper processing and analysis is performed.

References

[1] D. Podsiadlo and S. Richardson, "The timed 'Up & Go': a test of basic functional mobility for frail elderly persons," *J. Am. Geriatr. Soc.*, vol. 39, no. 2, pp. 142–148, Feb. 1991.

[2] Y. Nilsagård, Cecilia 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.

[3] S. L. Kasser, J. V. Jacobs, M. Ford, and T. W. Tourville, "Effects of balance-specific exercises on balance, physical activity and quality of life in adults with multiple sclerosis: a pilot investigation," *Disabil. Rehabil.*, vol. 37, no. 24, pp. 2238–2249, Nov. 2015, doi: 10.3109/09638288.2015.1019008.

[4] 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.

[5] E. W. Peterson, C. C. Cho, L. von Koch, and M. L. Finlayson, "Injurious Falls Among Middle Aged and Older Adults With Multiple Sclerosis," *Arch. Phys. Med. Rehabil.*, vol. 89, no. 6, pp. 1031–1037, Jun. 2008, doi: 10.1016/j.apmr.2007.10.043.

[6] B. Bebo *et al.*, "The Economic Burden of Multiple Sclerosis in the United States: Estimate of Direct and Indirect Costs," *Neurology*, vol. 98, no. 18, pp. e1810–e1817, May 2022, doi: 10.1212/WNL.0000000000200150.

[7] J. Veldhuijzen van Zanten, M. R. Douglas, and N. Ntoumanis, "Fatigue and fluctuations in physical and psychological wellbeing in people with multiple sclerosis: A longitudinal study," *Mult. Scler. Relat. Disord.*, vol. 47, p. 102602, Jan. 2021, doi: 10.1016/j.msard.2020.102602.

[8] M. H. Cameron, E. Thielman, R. Mazumder, and D. Bourdette, "Predicting falls in people with multiple sclerosis: fall history is as accurate as more complex measures," *Mult. Scler. Int.*, 2013, Accessed: Jan. 02, 2020. [Online]. Available: https://link.gale.com/apps/doc/A377777201/HWRC?u=vol_b92b&sid=HWRC&xid=552f58cd

[9] 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; quiz 632, Jul. 2011, doi: 10.1016/j.pmrj.2011.04.015.

[10] C. Hentzen *et al.*, "Are falls in people with multiple sclerosis related to the severity of urinary disorders?," *Ann. Phys. Rehabil. Med.*, vol. 64, no. 4, p. 101452, Jul. 2021, doi: 10.1016/j.rehab.2020.10.006.

[11] J. E. Zelaya, C. Murchison, and M. Cameron, "Associations Between Bladder Dysfunction and Falls in People with Relapsing-Remitting Multiple Sclerosis," *Int. J. MS Care*, vol. 19, no. 4, pp. 184–190, 2017, doi: 10.7224/1537-2073.2016-049.

[12] H. Choobsaz, S. ShahAli, R. Salehi, S. Noorizadeh Dehkordi, and S. Shanbehzadeh, "Discriminative ability of fall efficacy scale international in Iranian people with multiple sclerosis," *Mult. Scler. Relat. Disord.*, vol. 42, p. 102083, Jul. 2020, doi: 10.1016/j.msard.2020.102083.

[13] P. D. Hoang *et al.*, “Fall risk in people with MS: A Physiological Profile Assessment study,” *Mult. Scler. J. - Exp. Transl. Clin.*, vol. 2, p. 2055217316641130, Dec. 2016, doi: 10.1177/2055217316641130.
[14] R. van Vliet, P. Hoang, S. Lord, S. Gandevia, and K. Delbaere, “Falls efficacy scale-international: a cross-sectional validation in people with multiple sclerosis,” *Arch. Phys. Med. Rehabil.*, vol. 94, no. 5, pp. 883–889, May 2013, doi: 10.1016/j.apmr.2012.10.034.
[15] 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,” *Mult. Scler. Houndmills Basingstoke Engl.*, vol. 21, no. 1, pp. 92–100, Jan. 2015, doi: 10.1177/1352458514538884.
[16] R. Mazumder, W. E. Lambert, T. Nguyen, D. N. Bourdette, and M. H. Cameron, “Fear of Falling Is Associated with Recurrent Falls in People with Multiple Sclerosis: A Longitudinal Cohort Study,” *Int. J. MS Care*, vol. 17, no. 4, pp. 164–170, 2015, doi: 10.7224/1537-2073.2014-042.
[17] G. Quinn, L. Comber, C. McGuigan, A. Hannigan, R. Galvin, and S. Coote, “Risk factors for falling for people with Multiple Sclerosis identified in a prospective cohort study,” *Clin. Rehabil.*, vol. 35, no. 5, pp. 765–774, May 2021, doi: 10.1177/0269215520973197.
[18] L. Prosperini and L. Castelli, “Spotlight on postural control in patients with multiple sclerosis,” *Degener. Neurol. Neuromuscul. Dis.*, vol. 8, pp. 25–34, 2018, doi: 10.2147/DNND.S135755.
[19] A. Kalron, R. Aloni, M. Dolev, L. Frid, U. Givon, and S. Menascu, “The relationship between gait variability and cognitive functions differs between fallers and non-fallers in MS,” *J. Neural Transm. Vienna Austria 1996*, vol. 125, no. 6, pp. 945–952, Jun. 2018, doi: 10.1007/s00702-018-1843-y.
[20] A. Kalron, R. Aloni, and G. Allali, “The relationship between depression, anxiety and cognition and its paradoxical impact on falls in multiple sclerosis patients,” *Mult. Scler. Relat. Disord.*, vol. 25, pp. 167–172, Oct. 2018, doi: 10.1016/j.msard.2018.07.029.
[21] D. Cattaneo, A. Regola, and M. Meotti, “Validity of six balance disorders scales in persons with multiple sclerosis,” *Disabil. Rehabil.*, vol. 28, no. 12, pp. 789–795, Jun. 2006, doi: 10.1080/09638280500404289.
[22] 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.
[23] “Modified Fatigue Impact Scale,” *Shirley Ryan AbilityLab*. https://www.sralab.org/rehabilitation-measures/modified-fatigue-impact-scale (accessed Jun. 16, 2020).
[24] L. B. Strober *et al.*, “Tired of not knowing what that fatigue score means? Normative data of the Modified Fatigue Impact Scale (MFIS),” *Mult. Scler. Relat. Disord.*, vol. 46, p. 102576, Nov. 2020, doi: 10.1016/j.msard.2020.102576.
[25] 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, pp. 31–36, Jan. 2003, doi: 10.1212/WNL.60.1.31.

[26] L. Yardley, N. Beyer, K. Hauer, G. Kempen, C. Piot-Ziegler, and C. Todd, "Development and initial validation of the Falls Efficacy Scale-International (FES-I)," *Age Ageing*, vol. 34, no. 6, pp. 614–619, Nov. 2005, doi: 10.1093/ageing/afi196.
[27] 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.
[28] E. Sebastião, Y. C. Learmonth, and R. W. Motl, "Mobility measures differentiate falls risk status in persons with multiple sclerosis: An exploratory study," *NeuroRehabilitation*, vol. 40, no. 1, Art. no. 1, 2017, doi: 10.3233/NRE-161401.
[29] 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.
[30] Y. Nilsagård, M. Andreasson, A. Carling, and H. Vesterlin, "Examining the validity and sensitivity to change of the 5 and 10 sit-to-stand tests in people with multiple sclerosis," *Physiother. Res. Int. J. Res. Clin. Phys. Ther.*, vol. 22, no. 4, Oct. 2017, doi: 10.1002/pri.1681.
[31] A. Kalron, M. Dolev, and U. Givon, "Further construct validity of the Timed Up-and-Go Test as a measure of ambulation in multiple sclerosis patients," *Eur. J. Phys. Rehabil. Med.*, vol. 53, no. 6, pp. 841–847, Dec. 2017, doi: 10.23736/S1973-9087.17.04599-3.
[32] G. Quinn, L. Comber, C. McGuigan, R. Galvin, and S. Coote, "Discriminative ability and clinical utility of the Timed Up and Go (TUG) in identifying falls risk in people with multiple sclerosis: a prospective cohort study," *Clin. Rehabil.*, vol. 33, no. 2, Art. no. 2, Feb. 2019, doi: 10.1177/0269215518793481.
[33] Y. Nilsagård, E. Westerdahl, A. Wittrin, and M. Gunnarsson, "Walking Distance as a Predictor of Falls in People With Multiple Sclerosis," *Physiother. Res. Int. J. Res. Clin. Phys. Ther.*, vol. 21, no. 2, pp. 102–108, Jun. 2016, doi: 10.1002/pri.1625.
[34] A. Kalron, S. Menascu, M. Dolev, and U. Givon, "The walking speed reserve in low disabled people with multiple sclerosis: Does it provide greater insight in detecting mobility deficits and risk of falling than preferred and fast walking speeds?," *Mult. Scler. Relat. Disord.*, vol. 17, pp. 202–206, Oct. 2017, doi: 10.1016/j.msard.2017.08.010.
[35] Y. Etemadi, "Dual task cost of cognition is related to fall risk in patients with multiple sclerosis: a prospective study," *Clin. Rehabil.*, vol. 31, no. 2, pp. 278–284, Feb. 2017, doi: 10.1177/0269215516637201.
[36] D. A. Wajda, R. W. Motl, and J. J. Sosnoff, "Dual task cost of walking is related to fall risk in persons with multiple sclerosis," *J. Neurol. Sci.*, vol. 335, no. 1–2, pp. 160–163, Dec. 2013, doi: 10.1016/j.jns.2013.09.021.
[37] L. E. Dibble, C. Lopez-Lennon, W. Lake, C. Hoffmeister, and E. Gappmaier, "Utility of disease-specific measures and clinical balance tests in prediction of falls in persons with multiple sclerosis," *J. Neurol. Phys. Ther. JNPT*, vol. 37, no. 3, pp. 99–104, Sep. 2013, doi: 10.1097/NPT.0b013e3182a18460.
[38] K. D. Mitchell, H. Chen, and S. P. Silfies, "Test-Retest Reliability, Validity, and Minimal Detectable Change of the Balance Evaluation Systems Test to Assess

Balance in Persons with Multiple Sclerosis," *Int. J. MS Care*, vol. 20, no. 5, pp. 231–237, 2018, doi: 10.7224/1537-2073.2016-118.
[39] J. V. Jacobs and S. L. Kasser, "Balance impairment in people with multiple sclerosis: preliminary evidence for the Balance Evaluation Systems Test," *Gait Posture*, vol. 36, no. 3, pp. 414–418, Jul. 2012, doi: 10.1016/j.gaitpost.2012.03.026.
[40] T. Hortobágyi *et al.*, "Beam Walking to Assess Dynamic Balance in Health and Disease: A Protocol for the 'BEAM' Multicenter Observational Study," *Gerontology*, vol. 65, no. 4, pp. 332–339, 2019, doi: 10.1159/000493360.
[41] M. Tijsma, E. Vister, P. Hoang, and S. R. Lord, "A simple test of choice stepping reaction time for assessing fall risk in people with multiple sclerosis," *Disabil. Rehabil.*, vol. 39, no. 6, Art. no. 6, 2017, doi: 10.3109/09638288.2016.1148784.
[42] S. Caselli *et al.*, "When 'good' is not good enough: a retrospective Rasch analysis study of the Berg Balance Scale for persons with Multiple Sclerosis," *Front. Neurol.*, vol. 14, May 2023, doi: 10.3389/fneur.2023.1171163.
[43] S. Beer, F. Khan, and J. Kesselring, "Rehabilitation interventions in multiple sclerosis: an overview," *J. Neurol.*, vol. 259, no. 9, pp. 1994–2008, Sep. 2012, doi: 10.1007/s00415-012-6577-4.
[44] J. Tollár *et al.*, "Exercise Effects on Multiple Sclerosis Quality of Life and Clinical-Motor Symptoms," *Med. Sci. Sports Exerc.*, vol. 52, no. 5, pp. 1007–1014, May 2020, doi: 10.1249/MSS.0000000000002228.
[45] S. Coote, N. Hogan, and S. Franklin, "Falls in people with multiple sclerosis who use a walking aid: prevalence, factors, and effect of strength and balance interventions," *Arch. Phys. Med. Rehabil.*, vol. 94, no. 4, pp. 616–621, Apr. 2013, doi: 10.1016/j.apmr.2012.10.020.
[46] Y. Moon *et al.*, "Monitoring gait in multiple sclerosis with novel wearable motion sensors," *PLOS ONE*, vol. 12, no. 2, p. e0171346, Feb. 2017, doi: 10.1371/journal.pone.0171346.
[47] M. L. Frechette, B. M. Meyer, L. J. Tulipani, R. D. Gurchiek, R. S. McGinnis, and J. J. Sosnoff, "Next Steps in Wearable Technology and Community Ambulation in Multiple Sclerosis," *Curr. Neurol. Neurosci. Rep.*, vol. 19, no. 10, p. 80, Sep. 2019, doi: 10.1007/s11910-019-0997-9.
[48] R. S. McGinnis *et al.*, "A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis," *PLoS ONE*, vol. 12, no. 6, Jun. 2017, doi: 10.1371/journal.pone.0178366.
[49] Y. Moon, D. A. Wajda, R. W. Motl, and J. J. Sosnoff, "Stride-Time Variability and Fall Risk in Persons with Multiple Sclerosis," *Mult. Scler. Int.*, vol. 2015, 2015, doi: 10.1155/2015/964790.
[50] M. Mancini *et al.*, "ISway: a sensitive, valid and reliable measure of postural control," *J. NeuroEngineering Rehabil.*, vol. 9, no. 1, Art. no. 1, 2012, doi: 10.1186/1743-0003-9-59.
[51] 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.
[52] E. Sebastião, Y. C. Learmonth, and R. W. Motl, "Mobility measures differentiate falls risk status in persons with multiple sclerosis: An exploratory study," *NeuroRehabilitation*, vol. 40, no. 1, pp. 153–161, 2017, doi: 10.3233/NRE-161401.

[53] 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.
[54] Y. Cheng, K. Wang, H. Xu, T. Li, Q. Jin, and D. Cui, “Recent developments in sensors for wearable device applications,” *Anal. Bioanal. Chem.*, vol. 413, no. 24, pp. 6037–6057, Oct. 2021, doi: 10.1007/s00216-021-03602-2.
[55] M. M. Mañago, M. Cameron, and M. Schenkman, “Association of the Dynamic Gait Index to fall history and muscle function in people with multiple sclerosis,” *Disabil. Rehabil.*, pp. 1–6, May 2019, doi: 10.1080/09638288.2019.1607912.
[56] 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.
[57] E. M. Edwards *et al.*, “Backward Walking and Dual-Task Assessment Improve Identification of Gait Impairments and Fall Risk in Individuals with MS,” *Mult. Scler. Int.*, vol. 2020, p. e6707414, Sep. 2020, doi: 10.1155/2020/6707414.
[58] S. Tajali, M. Mehravar, H. Negahban, J. H. van Dieën, M.-J. Shaterzadeh-Yazdi, and R. Mofateh, “Impaired local dynamic stability during treadmill walking predicts future falls in patients with multiple sclerosis: A prospective cohort study,” *Clin. Biomech.*, vol. 67, pp. 197–201, Jul. 2019, doi: 10.1016/j.clinbiomech.2019.05.013.
[59] 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.
[60] A. T. Peebles, A. P. Bruetsch, S. G. Lynch, and J. M. Huisinga, “Dynamic balance in persons with multiple sclerosis who have a falls history is altered compared to non-fallers and to healthy controls,” *J. Biomech.*, vol. 63, pp. 158–163, Oct. 2017, doi: 10.1016/j.jbiomech.2017.08.023.
[61] H. J. Witchel *et al.*, “Thigh-Derived Inertial Sensor Metrics to Assess the Sit-to-Stand and Stand-to-Sit Transitions in the Timed Up and Go (TUG) Task for Quantifying Mobility Impairment in Multiple Sclerosis,” *Front. Neurol.*, vol. 9, p. 684, 2018, doi: 10.3389/fneur.2018.00684.
[62] 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, p. e0196463, May 2018, doi: 10.1371/journal.pone.0196463.
[63] 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.
[64] S. Coote, J. J. Sosnoff, and H. Gunn, “Fall Incidence as the Primary Outcome in Multiple Sclerosis Falls-Prevention Trials,” *Int. J. MS Care*, vol. 16, no. 4, pp. 178–184, 2014, doi: 10.7224/1537-2073.2014-059.
[65] K. Berg, S. Wood-Dauphine, J. I. Williams, and D. Gayton, “Measuring balance in the elderly: preliminary development of an instrument,” *Physiother. Can.*, Apr. 2009, doi: 10.3138/ptc.41.6.304.
[66] S. L. Kasser, A. Goldstein, P. K. Wood, and J. Sibold, “Symptom variability, affect and physical activity in ambulatory persons with multiple sclerosis: Understanding

patterns and time-bound relationships," *Disabil. Health J.*, vol. 10, no. 2, pp. 207–213, Apr. 2017, doi: 10.1016/j.dhjo.2016.10.006.
[67] D. Cattaneo, C. De Nuzzo, T. Fascia, M. Macalli, I. Pisoni, and R. Cardini, "Risks of falls in subjects with multiple sclerosis," *Arch. Phys. Med. Rehabil.*, vol. 83, no. 6, pp. 864–867, Jun. 2002, doi: 10.1053/apmr.2002.32825.
[68] D. Yu and L. Deng, "Deep Learning and Its Applications to Signal and Information Processing [Exploratory DSP," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 145–154, Jan. 2011, doi: 10.1109/MSP.2010.939038.
[69] S. Hochreiter and J. Schmidhuber, "Long short-term memory.," *Neural Comput.*, vol. 9, no. 8, p. 1735, Nov. 1997.
[70] D. Giansanti, V. Macellari, and G. Maccioni, "New neural network classifier of fall-risk based on the Mahalanobis distance and kinematic parameters assessed by a wearable device," *Physiol. Meas.*, vol. 29, no. 3, pp. N11–N19, Mar. 2008, doi: 10.1088/0967-3334/29/3/N01.
[71] C. Tunca, G. Salur, and C. Ersoy, "Deep Learning for Fall Risk Assessment With Inertial Sensors: Utilizing Domain Knowledge in Spatio-Temporal Gait Parameters," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 7, pp. 1994–2005, Jul. 2020, doi: 10.1109/JBHI.2019.2958879.
[72] A. Nait Aicha, G. Englebienne, K. S. Van Schooten, M. Pijnappels, and B. Kröse, "Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry," *Sensors*, vol. 18, no. 5, p. 1654, May 2018, doi: 10.3390/s18051654.
[73] E. Torti *et al.*, "Embedded Real-Time Fall Detection with Deep Learning on Wearable Devices," in *2018 21st Euromicro Conference on Digital System Design (DSD)*, Aug. 2018, pp. 405–412. doi: 10.1109/DSD.2018.00075.
[74] I. Wayan Wiprayoga Wisesa and G. Mahardika, "Fall detection algorithm based on accelerometer and gyroscope sensor data using Recurrent Neural Networks," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 258, p. 012035, May 2019, doi: 10.1088/1755-1315/258/1/012035.
[75] M. Musci, D. D. Martini, N. Blago, T. Facchinetti, and M. Piastra, "Fall Detection using Recurrent Neural Networks," p. 7, Apr. 2018.
[76] F. Luna-Perejon *et al.*, "An Automated Fall Detection System Using Recurrent Neural Networks," in *Artificial Intelligence in Medicine*, D. Riaño, S. Wilk, and A. ten Teije, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 36–41. doi: 10.1007/978-3-030-21642-9_6.
[77] F. Luna-Perejón, M. J. Domínguez-Morales, and A. Civit-Balcells, "Wearable Fall Detector Using Recurrent Neural Networks," *Sensors*, vol. 19, no. 22, Art. no. 22, Jan. 2019, doi: 10.3390/s19224885.
[78] X. Yu, H. Qiu, and S. Xiong, "A Novel Hybrid Deep Neural Network to Predict Pre-impact Fall for Older People Based on Wearable Inertial Sensors," *Front. Bioeng. Biotechnol.*, vol. 8, Feb. 2020, doi: 10.3389/fbioe.2020.00063.
[79] Y. Zhou *et al.*, "Classification of Neurological Patients to Identify Fallers Based on Spatial-Temporal Gait Characteristics Measured by a Wearable Device," *Sensors*, vol. 20, no. 15, p. 4098, Jul. 2020, doi: 10.3390/s20154098.

[80] R. Z. U. Rehman *et al.*, "Gait Analysis with Wearables Can Accurately Classify Fallers from Non-Fallers: A Step toward Better Management of Neurological Disorders," *Sensors*, vol. 20, no. 23, p. 6992, Dec. 2020, doi: 10.3390/s20236992.
[81] V. Karle *et al.*, "The Two-Minute Walk Test in Persons with Multiple Sclerosis: Correlations of Cadence with Free-Living Walking Do Not Support Ecological Validity," *Int. J. Environ. Res. Public. Health*, vol. 17, no. 23, p. 9044, Dec. 2020, doi: 10.3390/ijerph17239044.
[82] S. Shema-Shiratzky *et al.*, "A wearable sensor identifies alterations in community ambulation in multiple sclerosis: contributors to real-world gait quality and physical activity," *J. Neurol.*, vol. 267, no. 7, pp. 1912–1921, Jul. 2020, doi: 10.1007/s00415-020-09759-7.
[83] S. Del Din, A. Godfrey, B. Galna, S. Lord, and L. Rochester, "Free-living gait characteristics in ageing and Parkinson's disease: impact of environment and ambulatory bout length," *J. Neuroengineering Rehabil.*, vol. 13, no. 1, p. 46, 12 2016, doi: 10.1186/s12984-016-0154-5.
[84] K. C. Foucher, L. E. Thorp, D. Orozco, M. Hildebrand, and M. A. Wimmer, "Differences in Preferred Walking Speeds in a Gait Laboratory Compared With the Real World After Total Hip Replacement," *Arch. Phys. Med. Rehabil.*, vol. 91, no. 9, pp. 1390–1395, Sep. 2010, doi: 10.1016/j.apmr.2010.06.015.
[85] N. Takayanagi *et al.*, "Relationship between Daily and In-laboratory Gait Speed among Healthy Community-dwelling Older Adults," *Sci. Rep.*, vol. 9, no. 1, p. 3496, Dec. 2019, doi: 10.1038/s41598-019-39695-0.
[86] B. Najafi, J. L. Helbostad, R. Moe-Nilssen, W. Zijlstra, and K. Aminian, "Does walking strategy in older people change as a function of walking distance?," *Gait Posture*, vol. 29, no. 2, pp. 261–266, Feb. 2009, doi: 10.1016/j.gaitpost.2008.09.002.
[87] A. Kalron and U. Givon, "Gait characteristics according to pyramidal, sensory and cerebellar EDSS subcategories in people with multiple sclerosis," *J. Neurol.*, vol. 263, no. 9, pp. 1796–1801, Sep. 2016, doi: 10.1007/s00415-016-8200-6.
[88] R. Mills, A. Tennant, and C. Young, "The Neurological Sleep Index: A suite of new sleep scales for multiple sclerosis," *Mult. Scler. J. - Exp. Transl. Clin.*, vol. 2, pp. 1–10, Apr. 2016, doi: 10.1177/2055217316642263.
[89] O. Kramer, "K-Nearest Neighbors," in *Dimensionality Reduction with Unsupervised Nearest Neighbors*, O. Kramer, Ed., in Intelligent Systems Reference Library. Berlin, Heidelberg: Springer, 2013, pp. 13–23. doi: 10.1007/978-3-642-38652-7_2.
[90] M. Kaufman, D. Moyer, and J. Norton, "The significant change for the Timed 25-foot Walk in the multiple sclerosis functional composite," *Mult. Scler. Houndmills Basingstoke Engl.*, vol. 6, no. 4, pp. 286–290, Aug. 2000, doi: 10.1177/135245850000600411.
[91] C. J. Jones, R. E. Rikli, and W. C. Beam, "A 30-s Chair-Stand Test as a Measure of Lower Body Strength in Community-Residing Older Adults," *Res. Q. Exerc. Sport*, vol. 70, no. 2, pp. 113–119, Jun. 1999, doi: 10.1080/02701367.1999.10608028.
[92] J. Graham, "Foot drop: Explaining the causes, characteristics and treatment," *Br. J. Neurosci. Nurs.*, vol. 6, no. 4, pp. 168–172, Apr. 2010, doi: 10.12968/bjnn.2010.6.4.47792.

[93] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, “LSTM Networks for Mobile Human Activity Recognition,” presented at the 2016 International Conference on Artificial Intelligence: Technologies and Applications, Atlantis Press, Jan. 2016, pp. 50–53. doi: 10.2991/icaita-16.2016.13.
[94] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural Netw.*, vol. 18, no. 5, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.
[95] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.
[96] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *ArXiv14126980 Cs*, Jan. 2017, Accessed: Feb. 04, 2022. [Online]. Available: http://arxiv.org/abs/1412.6980
[97] R. D. Gurchiek *et al.*, “Open-Source Remote Gait Analysis: A Post-Surgery Patient Monitoring Application,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, Nov. 2019, doi: 10.1038/s41598-019-54399-1.
[98] R. D. Gurchiek, C. P. Garabed, and R. S. McGinnis, “Gait event detection using a thigh-worn accelerometer,” *Gait Posture*, vol. 80, pp. 214–216, Jul. 2020, doi: 10.1016/j.gaitpost.2020.06.004.
[99] A. Supratak, G. Datta, A. R. Gafson, R. Nicholas, Y. Guo, and P. M. Matthews, “Remote Monitoring in the Home Validates Clinical Gait Measures for Multiple Sclerosis,” *Front. Neurol.*, vol. 9, p. 561, Jul. 2018, doi: 10.3389/fneur.2018.00561.
[100] J. M. Huisinga, M. Mancini, R. J. St. George, and F. B. Horak, “Accelerometry Reveals Differences in Gait Variability Between Patients with Multiple Sclerosis and Healthy Controls,” *Ann. Biomed. Eng.*, vol. 41, no. 8, pp. 1670–1679, Aug. 2013, doi: 10.1007/s10439-012-0697-y.
[101] J. J. Craig, A. P. Bruetsch, S. G. Lynch, and J. M. Huisinga, “The relationship between trunk and foot acceleration variability during walking shows minor changes in persons with multiple sclerosis,” *Clin. Biomech.*, vol. 49, pp. 16–21, Nov. 2017, doi: 10.1016/j.clinbiomech.2017.07.011.
[102] M. Pau *et al.*, “Clinical assessment of gait in individuals with multiple sclerosis using wearable inertial sensors: Comparison with patient-based measure,” *Mult. Scler. Relat. Disord.*, vol. 10, pp. 187–191, Nov. 2016, doi: 10.1016/j.msard.2016.10.007.
[103] U. Givon, G. Zeilig, and A. Achiron, “Gait analysis in multiple sclerosis: Characterization of temporal–spatial parameters using GAITRite functional ambulation system,” *Gait Posture*, vol. 29, no. 1, pp. 138–142, Jan. 2009, doi: 10.1016/j.gaitpost.2008.07.011.
[104] Y. Moon, D. A. Wajda, R. W. Motl, and J. J. Sosnoff, “Stride-Time Variability and Fall Risk in Persons with Multiple Sclerosis,” *Mult. Scler. Int.*, vol. 2015, 2015, doi: 10.1155/2015/964790.
[105] V. V. Shah *et al.*, “Effect of Bout Length on Gait Measures in People with and without Parkinson’s Disease during Daily Life,” *Sensors*, vol. 20, no. 20, Art. no. 20, Jan. 2020, doi: 10.3390/s20205769.
[106] D. W. Hosmer, *Applied logistic regression.*, Third edition / David W. Hosmer, Jr., Stanley Lemeshow, Rodney X. Sturdivant.. Hoboken, New Jersey : Wiley, 2013.

[107] Aurélien Géron, *Understanding support vector machines*. O'Reilly Media, Inc, 2017.
[108] Zsolt Nagy, *Artificial Intelligence and Machine Learning Fundamentals*. Packt Publishing, 2018.
[109] W. Lee, *Python® Machine Learning*. Indianapolis, Indiana: Indianapolis, Indiana: John Wiley & Sons, Inc. doi: 10.1002/9781119557500.
[110] J. E. T. Akinsola, "Supervised Machine Learning Algorithms: Classification and Comparison," *Int. J. Comput. Trends Technol. IJCTT*, vol. 48, pp. 128–138, Jun. 2017, doi: 10.14445/22312803/IJCTT-V48P126.
[111] R. Z. U. Rehman *et al.*, "Gait Analysis with Wearables Can Accurately Classify Fallers from Non-Fallers: A Step toward Better Management of Neurological Disorders," *Sensors*, vol. 20, no. 23, Dec. 2020, doi: 10.3390/s20236992.
[112] P. Bet, P. C. Castro, and M. A. Ponti, "Fall detection and fall risk assessment in older person using wearable sensors: A systematic review," *Int. J. Med. Inf.*, vol. 130, p. 103946, Oct. 2019, doi: 10.1016/j.ijmedinf.2019.08.006.
[113] V. V. Shah *et al.*, "Laboratory versus daily life gait characteristics in patients with multiple sclerosis, Parkinson's disease, and matched controls," *J. NeuroEngineering Rehabil.*, vol. 17, no. 1, p. 159, Dec. 2020, doi: 10.1186/s12984-020-00781-4.
[114] L. Weed, C. Little, S. L. Kasser, and R. S. McGinnis, "A Preliminary Investigation of the Effects of Obstacle Negotiation and Turning on Gait Variability in Adults with Multiple Sclerosis," *Sensors*, vol. 21, no. 17, Art. no. 17, Jan. 2021, doi: 10.3390/s21175806.
[115] V. V. Shah *et al.*, "Does gait bout definition influence the ability to discriminate gait quality between people with and without multiple sclerosis during daily life?," *Gait Posture*, vol. 84, pp. 108–113, Feb. 2021, doi: 10.1016/j.gaitpost.2020.11.024.
[116] G. Bateman, J. Lechner-Scott, A. Bateman, J. Attia, and R. Lea, "Multiple sclerosis 2020," vol. 2, Aug. 2020, doi: 10.17632/mytp5z2zdd.2.
[117] Ž. Lesjak *et al.*, "A Novel Public MR Image Dataset of Multiple Sclerosis Patients With Lesion Segmentations Based on Multi-rater Consensus," *Neuroinformatics*, vol. 16, no. 1, pp. 51–63, Jan. 2018, doi: 10.1007/s12021-017-9348-7.
[118] "NITRC: Longitudinal Multiple Sclerosis Lesion Imaging Archive: Tool/Resource Info." https://www.nitrc.org/projects/longitudinal_ms/ (accessed Dec. 07, 2021).
[119] "Full dataset of relapsing-remitting MS patients (N = 145)." PLOS ONE, Jan. 24, 2019. doi: 10.1371/journal.pone.0211120.s001.
[120] C. Mosquera-Lopez *et al.*, "Automated Detection of Real-World Falls: Modeled From People With Multiple Sclerosis," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 6, pp. 1975–1984, Jun. 2021, doi: 10.1109/JBHI.2020.3041035.
[121] R. D. Gurchiek *et al.*, "Open-Source Remote Gait Analysis: A Post-Surgery Patient Monitoring Application," *Sci. Rep.*, vol. 9, no. 1, Art. no. 1, Nov. 2019, doi: 10.1038/s41598-019-54399-1.
[122] 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.

[123] 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. PP, Apr. 2022, doi: 10.1109/TNSRE.2022.3169962.

[124] R. S. McGinnis *et al.*, “A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis,” *PLOS ONE*, vol. 12, no. 6, p. e0178366, Jun. 2017, doi: 10.1371/journal.pone.0178366.

[125] R. D. Gurchiek *et al.*, “Sprint Assessment Using Machine Learning and a Wearable Accelerometer,” *J. Appl. Biomech.*, vol. 35, no. 2, pp. 164–169, Apr. 2019, doi: 10.1123/jab.2018-0107.

[126] 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.

[127] F. P. Bernhard *et al.*, “Wearables for gait and balance assessment in the neurological ward - study design and first results of a prospective cross-sectional feasibility study with 384 inpatients,” *BMC Neurol. Lond.*, vol. 18, 2018, doi: http://dx.doi.org/10.1186/s12883-018-1111-7.

[128] R. S. McGinnis *et al.*, “Rapid detection of internalizing diagnosis in young children enabled by wearable sensors and machine learning,” *PLOS ONE*, vol. 14, no. 1, p. e0210267, Jan. 2019, doi: 10.1371/journal.pone.0210267.

[129] C. R. Bellenger, D. J. Miller, S. L. Halson, G. D. Roach, and C. Sargent, “Wrist-Based Photoplethysmography Assessment of Heart Rate and Heart Rate Variability: Validation of WHOOP,” *Sensors*, vol. 21, no. 10, Art. no. 10, Jan. 2021, doi: 10.3390/s21103571.

[130] W. Ahmed, J. Capodilupo, A. Nicolae, R. Aruh, C. J. Liden, and B. T. Tom, “Wearable continuous physiological monitoring device,” USD792597S1, Jul. 18, 2017 Accessed: May 20, 2022. [Online]. Available: https://patents.google.com/patent/USD792597S1/en?assignee=whoop%2c+inc&oq=whoop%2c+inc

[131] C. B. Brumback, D. W. Knight, J. Park, A. C. Axley, and S. G. J. Yuen, “Biometric monitoring device with contextually-or environmentally-dependent display,” US8784271B2, Jul. 22, 2014 Accessed: May 20, 2022. [Online]. Available: https://patents.google.com/patent/US8784271B2/en?q=fitbit&oq=fitbit

[132] M. Patel, A. Pavic, and V. A. Goodwin, “Wearable inertial sensors to measure gait and posture characteristic differences in older adult fallers and non-fallers: A scoping review,” *Gait Posture*, vol. 76, pp. 110–121, Feb. 2020, doi: 10.1016/j.gaitpost.2019.10.039.

[133] N. Baker, C. Gough, and S. J. Gordon, “Inertial Sensor Reliability and Validity for Static and Dynamic Balance in Healthy Adults: A Systematic Review,” *Sensors*, vol. 21, no. 15, p. 5167, Jul. 2021, doi: 10.3390/s21155167.

[134] M. Ghislieri, L. Gastaldi, S. Pastorelli, S. Tadano, and V. Agostini, “Wearable Inertial Sensors to Assess Standing Balance: A Systematic Review,” *Sensors*, vol. 19, no. 19, Art. no. 19, Sep. 2019, doi: 10.3390/s19194075.

[135] J. M. Huisinga, M. Mancini, R. J. St. George, and F. B. Horak, "Accelerometry Reveals Differences in Gait Variability Between Patients with Multiple Sclerosis and Healthy Controls," *Ann. Biomed. Eng.*, vol. 41, no. 8, Art. no. 8, Aug. 2013, doi: 10.1007/s10439-012-0697-y.
[136] G. Allali *et al.*, "Gait variability in multiple sclerosis: a better falls predictor than EDSS in patients with low disability," *J. Neural Transm.*, vol. 123, no. 4, pp. 447–450, Apr. 2016, doi: 10.1007/s00702-016-1511-z.
[137] 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.
[138] M. D. Czech *et al.*, "Age and environment-related differences in gait in healthy adults using wearables," *Npj Digit. Med.*, vol. 3, no. 1, Art. no. 1, Sep. 2020, doi: 10.1038/s41746-020-00334-y.
[139] F. E. Godkin *et al.*, "Feasibility of a continuous, multi-sensor remote health monitoring approach in persons living with neurodegenerative disease," *J. Neurol.*, Oct. 2021, doi: 10.1007/s00415-021-10831-z.
[140] R. S. McGinnis and E. W. McGinnis, "Advancing Digital Medicine with Wearables in the Wild," *Sensors*, vol. 22, no. 12, Art. no. 12, Jan. 2022, doi: 10.3390/s22124576.
[141] S. L. Kasser, A. Goldstein, P. K. Wood, and J. Sibold, "Symptom variability, affect and physical activity in ambulatory persons with multiple sclerosis: Understanding patterns and time-bound relationships," *Disabil. Health J.*, vol. 10, no. 2, Art. no. 2, Apr. 2017, doi: 10.1016/j.dhjo.2016.10.006.
[142] A. Mueller *et al.*, "Continuous Digital Monitoring of Walking Speed in Frail Elderly Patients: Noninterventional Validation Study and Longitudinal Clinical Trial," *JMIR MHealth UHealth*, vol. 7, no. 11, p. e15191, Nov. 2019, doi: 10.2196/15191.
[143] 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, Art. no. 22, Jan. 2020, doi: 10.3390/s20226618.
[144] T. L. Hart, A. M. Swartz, S. E. Cashin, and S. J. Strath, "How many days of monitoring predict physical activity and sedentary behaviour in older adults?," *Int. J. Behav. Nutr. Phys. Act.*, vol. 8, p. 62, Jun. 2011, doi: 10.1186/1479-5868-8-62.
[145] C. E. Matthews, B. E. Ainsworth, R. W. Thompson, and D. R. Bassett, "Sources of variance in daily physical activity levels as measured by an accelerometer," *Med. Sci. Sports Exerc.*, vol. 34, no. 8, pp. 1376–1381, Aug. 2002, doi: 10.1097/00005768-200208000-00021.
[146] R. J. Gretebeck and H. J. Montoye, "Variability of some objective measures of physical activity," *Med. Sci. Sports Exerc.*, vol. 24, no. 10, pp. 1167–1172, Oct. 1992.
[147] K. S. van Schooten, S. M. Rispens, P. J. Elders, P. Lips, J. H. van Dieën, and M. Pijnappels, "Assessing physical activity in older adults: required days of trunk accelerometer measurements for reliable estimation," *J. Aging Phys. Act.*, vol. 23, no. 1, pp. 9–17, Jan. 2015, doi: 10.1123/japa.2013-0103.
[148] M. Kang, K. Bjornson, T. V. Barreira, B. G. Ragan, and K. Song, "The minimum number of days required to establish reliable physical activity estimates in

children aged 2-15 years," *Physiol. Meas.*, vol. 35, no. 11, pp. 2229–2237, Nov. 2014, doi: 10.1088/0967-3334/35/11/2229.
[149] S. Levin, D. R. Jacobs, B. E. Ainsworth, M. T. Richardson, and A. S. Leon, "Intra-individual variation and estimates of usual physical activity," *Ann. Epidemiol.*, vol. 9, no. 8, pp. 481–488, Nov. 1999, doi: 10.1016/s1047-2797(99)00022-8.
[150] A. Supratak, G. Datta, A. R. Gafson, R. Nicholas, Y. Guo, and P. M. Matthews, "Remote Monitoring in the Home Validates Clinical Gait Measures for Multiple Sclerosis," *Front. Neurol.*, vol. 9, Jul. 2018, doi: 10.3389/fneur.2018.00561.
[151] 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, Art. no. 1, Dec. 2019, doi: 10.1038/s41598-019-52697-2.
[152] E. Jortberg *et al.*, "A novel adhesive biosensor system for detecting respiration, cardiac, and limb movement signals during sleep: validation with polysomnography," *Nat. Sci. Sleep*, vol. 10, pp. 397–408, Nov. 2018, doi: 10.2147/NSS.S179588.
[153] 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.
[154] J. F. Kurtzke, "Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS)," *Neurology*, vol. 33, no. 11, pp. 1444–1444, Nov. 1983, doi: 10.1212/WNL.33.11.1444.
[155] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM Networks for Mobile Human Activity Recognition," Atlantis Press, Jan. 2016, pp. 50–53. doi: 10.2991/icaita-16.2016.13.
[156] J. J. Craig, A. P. Bruetsch, S. G. Lynch, and J. M. Huisinga, "The relationship between trunk and foot acceleration variability during walking shows minor changes in persons with multiple sclerosis," *Clin. Biomech.*, vol. 49, pp. 16–21, Nov. 2017, doi: 10.1016/j.clinbiomech.2017.07.011.
[157] J. C. Goldsack *et al.*, "Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs)," *NPJ Digit. Med.*, vol. 3, p. 55, 2020, doi: 10.1038/s41746-020-0260-4.
[158] Y. Nilsagård, A. Carling, and A. Forsberg, "Activities-Specific Balance Confidence in People with Multiple Sclerosis," *Mult. Scler. Int.*, vol. 2012, p. e613925, Aug. 2012, doi: 10.1155/2012/613925.
[159] S. Tajali *et al.*, "Predicting falls among patients with multiple sclerosis: Comparison of patient-reported outcomes and performance-based measures of lower extremity functions," *Mult. Scler. Relat. Disord.*, vol. 17, pp. 69–74, Oct. 2017, doi: 10.1016/j.msard.2017.06.014.
[160] N. Téllez, J. Río, M. Tintoré, C. Nos, I. Galán, and X. Montalban, "Does the Modified Fatigue Impact Scale offer a more comprehensive assessment of fatigue in MS?," *Mult. Scler. J.*, vol. 11, no. 2, pp. 198–202, Apr. 2005, doi: 10.1191/1352458505ms1148oa.

[161] L. B. Strober *et al.*, "Tired of not knowing what that fatigue score means? Normative data of the Modified Fatigue Impact Scale (MFIS)," *Mult. Scler. Relat. Disord.*, vol. 46, p. 102576, Nov. 2020, doi: 10.1016/j.msard.2020.102576.
[162] K. Kleinman and S. S. Huang, "Calculating Power by Bootstrap, with an Application to Cluster-Randomized Trials," *eGEMs*, vol. 4, no. 1, p. 1202, Feb. 2017, doi: 10.13063/2327-9214.1202.
[163] K. J. Loughran *et al.*, "Balance impairment in individuals with COPD: a systematic review with meta-analysis," *Thorax*, vol. 75, no. 7, pp. 539–546, Jul. 2020, doi: 10.1136/thoraxjnl-2019-213608.
[164] S. W. Muir, K. Berg, B. Chesworth, N. Klar, and M. Speechley, "Quantifying the magnitude of risk for balance impairment on falls in community-dwelling older adults: a systematic review and meta-analysis," *J. Clin. Epidemiol.*, vol. 63, no. 4, pp. 389–406, Apr. 2010, doi: 10.1016/j.jclinepi.2009.06.010.
[165] C. L. Martin *et al.*, "Gait and balance impairment in early multiple sclerosis in the absence of clinical disability," *Mult. Scler. J.*, vol. 12, no. 5, pp. 620–628, Sep. 2006, doi: 10.1177/1352458506070658.
[166] F. Khan and M. F. Chevidikunnan, "Prevalence of Balance Impairment and Factors Associated with Balance among Patients with Stroke. A Cross Sectional Retrospective Case Control Study," *Healthcare*, vol. 9, no. 3, Art. no. 3, Mar. 2021, doi: 10.3390/healthcare9030320.
[167] I. Schwartz, L. Kandel, A. Sajina, D. Litinezki, A. Herman, and Y. Mattan, "Balance is an important predictive factor for quality of life and function after primary total knee replacement," *J. Bone Joint Surg. Br.*, vol. 94, no. 6, pp. 782–786, Jun. 2012, doi: 10.1302/0301-620X.94B6.27874.
[168] 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.
[169] G. Quinn, L. Comber, R. Galvin, and S. Coote, "The ability of clinical balance measures to identify falls risk in multiple sclerosis: a systematic review and meta-analysis," *Clin. Rehabil.*, vol. 32, no. 5, pp. 571–582, May 2018, doi: 10.1177/0269215517748714.
[170] J. J. Sosnoff *et al.*, "Mobility, Balance and Falls in Persons with Multiple Sclerosis," *PLOS ONE*, vol. 6, no. 11, p. e28021, Nov. 2011, doi: 10.1371/journal.pone.0028021.
[171] S. L. Kasser, J. V. Jacobs, M. Ford, and T. W. Tourville, "Effects of balance-specific exercises on balance, physical activity and quality of life in adults with multiple sclerosis: a pilot investigation," *Disabil. Rehabil.*, vol. 37, no. 24, pp. 2238–2249, Nov. 2015, doi: 10.3109/09638288.2015.1019008.
[172] 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, pp. 247–255, 2020, doi: 10.7224/1537-2073.2020-014.
[173] S. M. Rao, *Neurobehavioral Aspects of Multiple Sclerosis*. Oxford University Press, 1990.
[174] 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, pp. 433–440, Aug. 2010, doi: 10.1586/erp.10.34.

[175] "Berg Balance Scale," *Shirley Ryan AbilityLab*. https://www.sralab.org/rehabilitation-measures/berg-balance-scale (accessed Dec. 13, 2022).
[176] 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.
[177] 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.
[178] 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.
[179] 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.
[180] 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.
[181] 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.
[182] 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.
[183] 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.
[184] E. Sen-Gupta *et al.*, "A Pivotal Study to Validate the Performance of a Novel Wearable Sensor and System for Biometric Monitoring in Clinical and Remote Environments," *Digit. Biomark.*, vol. 3, no. 1, pp. 1–13, Apr. 2019, doi: 10.1159/000493642.
[185] M. Hilty *et al.*, "Continuous monitoring with wearables in multiple sclerosis reveals an association of cardiac autonomic dysfunction with disease severity," *Mult. Scler. J. - Exp. Transl. Clin.*, vol. 8, no. 2, p. 20552173221103436, 2022, doi: 10.1177/20552173221103436.
[186] G. Prigent, S. Apte, A. Paraschiv-Ionescu, C. Besson, V. Gremeaux, and K. Aminian, "Concurrent Evolution of Biomechanical and Physiological Parameters With Running-Induced Acute Fatigue," *Front. Physiol.*, vol. 13, p. 814172, 2022, doi: 10.3389/fphys.2022.814172.
[187] 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.

[188] 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.
[189] 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, p. e0000120, Oct. 2022, doi: 10.1371/journal.pdig.0000120.
[190] H. B. Mann and D. R. Whitney, “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other,” *Ann. Math. Stat.*, vol. 18, no. 1, pp. 50–60, Mar. 1947, doi: 10.1214/aoms/1177730491.
[191] D. Lakens, “Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs,” *Front. Psychol.*, vol. 4, p. 863, Nov. 2013, doi: 10.3389/fpsyg.2013.00863.
[192] 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.
[193] M. H. G. Gerards, C. McCrum, A. Mansfield, and K. Meijer, “Perturbation-based balance training for falls reduction among older adults: Current evidence and implications for clinical practice,” *Geriatr. Gerontol. Int.*, vol. 17, no. 12, pp. 2294–2303, 2017, doi: 10.1111/ggi.13082.
[194] D. L. Sturnieks *et al.*, “Force-Controlled Balance Perturbations Associated with Falls in Older People: A Prospective Cohort Study,” *PLOS ONE*, vol. 8, no. 8, p. e70981, Aug. 2013, doi: 10.1371/journal.pone.0070981.
[195] 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, p. e0196463, 2018, doi: 10.1371/journal.pone.0196463.
[196] 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.
[197] 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.
[198] S. M. Rao, *Neurobehavioral Aspects of Multiple Sclerosis*. Oxford University Press, 1990.
[199] 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.
[200] 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.
[201] 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.
[202] 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.
[203] 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.
[204] 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.
[205] 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.
[206] 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.
[207] 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.
[208] "Modified Fatigue Impact Scale," *Shirley Ryan AbilityLab*, Jun. 16, 2020. https://www.sralab.org/rehabilitation-measures/modified-fatigue-impact-scale (accessed Jun. 16, 2020).
[209] 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.
[210] 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.
[211] 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.
[212] 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.
[213] 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.

Appendix A

Strides	AGG	Short Model	Short AUC	Medium Model	Medium AUC	Long Model	Long AUC	All Model	All AUC
22	None	N/A	N/A	N/A	N/A	LSTM 2 ABC	0.68	LSTM 3 ABC	0.64
	Median	N/A	N/A	N/A	N/A	LSTM 3 ABC	0.69	LSTM 2 ABC	***0.76***
21	None	N/A	N/A	N/A	N/A	LSTM 2 ABC	0.70	LSTM 3 ABC	0.60
	Median	N/A	N/A	N/A	N/A	LSTM 2 ABC	0.68	LSTM 3 ABC	0.69
20	None	N/A	N/A	N/A	N/A	LSTM 2 ABC	0.65	LSTM 2 ABC	0.64
	Median	N/A	N/A	N/A	N/A	LSTM 3 ABC	0.67	LSTM 2 ABC	0.67
19	None	N/A	N/A	N/A	N/A	LSTM 3 ABC	0.68	LSTM 3 ABC	0.67
	Median	N/A	N/A	N/A	N/A	LSTM 3 ABC	0.64	LSTM 2 ABC	0.63
18	None	N/A	N/A	N/A	N/A	LSTM 2 ABC	0.65	LSTM 2 ABC	0.66
	Median	N/A	N/A	N/A	N/A	LSTM 2 ABC	0.68	LSTM 3 ABC	0.64
17	None	N/A	N/A	N/A	N/A	LSTM 2 ABC	0.61	LSTM 2 ABC	0.63
	Median	N/A	N/A	N/A	N/A	LSTM 3 ABC	0.62	LSTM 2 ABC	0.62
16	None	N/A	N/A	LSTM 2	0.59	LSTM 2 ABC	0.66	LSTM 3 ABC	0.63
	Median	N/A	N/A	LSTM 2 ABC	0.61	LSTM 2 ABC	0.64	LSTM 2 ABC	0.64
15	None	N/A	N/A	LSTM 3	0.60	LSTM 2 ABC	0.67	LSTM 3 ABC	0.69
	Median	N/A	N/A	LSTM 3 ABC	0.63	LSTM 2 ABC	0.66	LSTM 3 ABC	0.72
14	None	N/A	N/A	LSTM 3 ABC	0.63	LSTM 2 ABC	0.68	LSTM 3 ABC	0.63
	Median	N/A	N/A	LSTM 2 ABC	0.64	LSTM 2 ABC	0.69	LSTM 2 ABC	0.65
13	None	N/A	N/A	LSTM 3 ABC	0.57	LSTM 3 ABC	0.69	LSTM 3 ABC	0.66
	Median	N/A	N/A	LSTM 3 ABC	0.62	LSTM 3 ABC	0.67	LSTM 2 ABC	0.69
12	None	N/A	N/A	LSTM 2 ABC	0.61	LSTM 2 ABC	0.67	LSTM 3 ABC	0.65
	Median	N/A	N/A	LSTM 3 ABC	0.66	LSTM 2 ABC	0.69	LSTM 2 ABC	0.69
11	None	N/A	N/A	LSTM 3 ABC	0.60	LSTM 2 ABC	0.68	LSTM 3 ABC	0.69
	Median	N/A	N/A	LSTM 2 ABC	0.66	LSTM 2 ABC	0.64	LSTM 3 ABC	0.68
10	None	N/A	N/A	LSTM 2 ABC	0.62	LSTM 3 ABC	0.66	LSTM 2 ABC	0.70
	Median	N/A	N/A	LSTM 2 ABC	0.68	LSTM 3 ABC	0.63	LSTM 2 ABC	0.74
9	None	N/A	N/A	LSTM 3 ABC	0.63	LSTM 2 ABC	0.68	LSTM 2 ABC	0.62
	Median	N/A	N/A	LSTM 2 ABC	0.67	LSTM 2 ABC	0.63	LSTM 2 ABC	0.64
8	None	N/A	N/A	LSTM 2 ABC	0.64	LSTM 3 ABC	0.68	LSTM 3 ABC	0.68
	Median	N/A	N/A	LSTM 2 ABC	0.71	LSTM 3 ABC	0.69	LSTM 3 ABC	0.59
7	None	N/A	N/A	LSTM 3 ABC	0.64	LSTM 2 ABC	0.66	LSTM 2 ABC	***0.71***

	Median	N/A	N/A	LSTM 2 ABC	0.65	LSTM 2 ABC	0.63	LSTM 2 ABC	0.63
6	None	N/A	N/A	LSTM 3 ABC	0.65	LSTM 2 ABC	0.66	LSTM 2 ABC	0.69
	Median	N/A	N/A	LSTM 3 ABC	0.75	LSTM 2 ABC	0.64	LSTM 2 ABC	0.64
5	None	LSTM 2 ABC	0.64	LSTM 3 ABC	0.62	LSTM 3 ABC	0.68	LSTM 3 ABC	0.68
	Median	LSTM 3 ABC	0.66	LSTM 2 ABC	0.70	LSTM 3 ABC	0.61	LSTM 2 ABC	0.58
4	None	LSTM 3 ABC	0.60	LSTM 2 ABC	0.61	LSTM 2 ABC	0.64	LSTM 2 ABC	0.66
	Median	LSTM 3 ABC	0.57	LSTM 3 ABC	0.71	LSTM 2 ABC	0.59	LSTM 2 ABC	0.62
3	None	LSTM 3 ABC	0.66	LSTM 3 ABC	0.65	LSTM 2 ABC	0.65	LSTM 3 ABC	***0.71***
	Median	LSTM 3 ABC	0.65	LSTM 3 ABC	0.74	LSTM 3 ABC	0.60	LSTM 3 ABC	0.73
2	None	LSTM 3 ABC	0.66	LSTM 2 ABC	0.60	LSTM 3 ABC	0.66	LSTM 2 ABC	***0.71***
	Median	LSTM 3 ABC	0.66	LSTM 2 ABC	0.73	LSTM 3 ABC	0.64	LSTM 2 ABC	0.73
1	None	LSTM 2 ABC	0.65	LSTM 2 ABC	0.61	LSTM 2 ABC	0.69	LSTM 3 ABC	0.69
	Median	LSTM 2 ABC	0.66	LSTM 3 ABC	0.67	LSTM 2 ABC	0.61	LSMT 3 ABC	0.73

LSTM: Long-Short Term Memory Neural Network; LSTM 2: Model with one LSTM layer and one BilSTM layer; LSTM 3: Model with LSTM Layers; AGG: Aggregation technique (none or median of all remote stride observations); AUC: Area Under the Receiver Operating Characteristic Curve; ABC: Activity Specific Balance Confidence added as input feature; N/A: Not enough data available to extract specified number of strides from each subject.

Appendix B

Difference Testing Results

Table B1: Gait Feature Ranksum Difference Testing in Persons with MS

Gait Feature (14 total)	**1D v 1W**	**2D v 1W**	**3D v 1W**	**WE v WD**	**1W v 2W**	**1W v 3W**	**1W v 4W**	**1W v 5W**	**1W v 6W**
Acceleration Asymmetry									
Correlation Asymmetry									
Double Support Duration	CV								
Duty Factor									
Duty Factor Asymmetry									
Entropy Ratio									
Entropy Ratio Asymmetry									
Frequency Dispersion ML									
Lyapunov Exponent AP									
Lyapunov Exponent ML									
RMS AP									
Stance Duration									
Stride Duration									
Swing Duration	CV								
Number of Significant Differences	2-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P

[B1] Difference testing of gait features for PwMS. A significant difference of the feature at the timeframe is denoted by CV, M, or 95th P for Coefficient of Variation, Median, and 95th Percentile, respectively. The timeframes are abbreviated as follows: 1D = 1 Day, 2D = 2Days, 1W = 1 Week, WE = Weekend, WD = Weekday, etc. If the box is empty, no significant difference was found. The significance threshold was 0.05.

Table B2: Sway Feature Ranksum Difference Testing in Persons with MS

Sway Feature (13 total)	**1D v 1W**	**2D v 1W**	**3D v 1W**	**WE v WD**	**1W v 2W**	**1W v 3W**	**1W v 4W**	**1W v 5W**	**1W v 6W**
Area	CV								
Centroidal Frequency									
Distance									
50th Percentile Frequency									
95th Percentile Frequency									
Frequency Dispersion									
Jerk									
Mean Period	CV								
Mean Velocity									
Path									
Power	CV								
Range									
RMS									
Number of Significant Differences	3-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P	0-CV 0-M 0-95th P

[B2] Difference testing of sway features for PwMS. Timeframes are abbreviated as follows: 1D = 1 Day, 2D = 2Days, 1W = 1 Week, WE = Weekend, WD = Weekday, etc. A significant difference of the feature at the timeframe is denoted by CV, M, or 95th P for Coefficient of Variation, Median, and 95th Percentile, respectively. If the box is empty, no significant difference was found. The significance threshold was 0.05.

ICC Results

Table B3: Intra-class Correlation Wear Duration Gait Analysis in Persons with MS

Gait Feature (14 total)	**1D v 1W**	**2D v 1W**	**3D v 1W**	**WE v WD**	**1W v 2W**	**1W v 3W**	**1W v 4W**	**1W v 5W**	**1W v 6W**
Acceleration Asymmetry	0.92	0.95	0.98	0.96	0.99	0.97	0.96	0.95	0.98
	0.89	0.97	0.98	0.94	0.98	0.97	0.94	0.93	0.98
	0.80	0.90	0.91	0.92	0.98	0.90	0.91	0.89	0.94
Correlation Asymmetry	0.96	0.98	0.99	0.98	1.00	0.99	0.98	0.97	0.96
	0.75	0.90	0.95	0.91	0.98	0.96	0.95	0.96	0.95
	0.94	0.96	0.98	0.95	0.98	0.98	0.96	0.93	0.96
Double Support Duration	0.95	0.96	0.98	0.93	0.99	0.98	0.87	0.89	0.96
	0.84	0.83	0.91	***0.38***	0.93	0.88	0.74	***0.48***	0.72
	0.62	0.84	0.98	***0.60***	0.96	0.92	0.87	0.78	0.79
Duty Factor	0.96	0.98	0.99	0.98	0.98	0.97	0.91	0.95	0.93
	0.70	0.94	0.97	0.96	0.99	0.96	0.95	0.86	0.89
	0.80	0.91	0.92	0.87	0.97	0.93	0.84	0.76	0.83
Duty Factor Asymmetry	0.96	0.97	0.99	0.98	0.99	0.99	0.97	0.98	0.97
	0.79	0.97	0.96	0.96	0.99	0.99	0.95	0.97	0.96
	0.89	0.94	0.97	0.96	0.97	0.97	0.95	0.94	0.96
Entropy Ratio	-	0.84	0.85	0.71	0.98	0.95	***0.62***	-	-
	-	0.91	0.96	0.91	0.98	0.97	0.81	-	-
	-	0.82	0.91	0.90	0.87	0.90	***0.53***	-	-
Entropy Ratio Asymmetry	-	0.97	0.96	0.88	0.95	0.94	0.92	-	-
	-	0.83	0.85	***0.60***	0.86	0.87	0.78	-	-
	-	***0.45***	0.72	***0.32***	0.91	0.85	***0.39***	-	-
Frequency Dispersion ML	0.96	0.95	0.97	0.94	0.99	0.99	0.93	0.89	0.79
	0.76	0.93	0.95	0.93	0.99	0.99	0.96	0.89	0.96
	0.87	0.94	0.98	0.87	0.98	0.98	0.91	0.84	0.94
Lyapunov Exponent AP	-	-	-	***0.57***	0.99	0.97	0.83	***0.66***	0.89
	-	-	-	0.71	0.97	0.97	0.73	***0.44***	0.93
	-	-	-	***0.02***	***0.00***	***0.00***	***0.08***	***0.19***	***0.00***
Lyapunov Exponent ML	-	-	-	0.79	0.84	0.88	0.83	0.84	0.92
	-	-	-	***0.57***	0.98	0.97	0.80	***0.60***	0.82
	-	-	-	***0.15***	***0.57***	***0.07***	***0.53***	***0.00***	***0.00***
RMS AP	0.83	0.95	0.96	0.94	0.98	0.97	0.88	0.94	0.95
	0.68	0.93	0.95	0.87	0.97	0.92	0.86	0.87	0.90
	0.83	0.88	0.94	0.86	0.95	0.95	***0.64***	0.83	0.91
Stance Duration	0.89	0.95	0.98	0.89	0.99	0.98	0.93	0.94	0.98
	0.53	0.96	0.96	0.90	0.99	0.98	0.94	0.88	0.92
	0.57	0.94	0.96	0.85	0.97	0.96	0.78	0.85	0.88
Stride Duration	0.92	0.95	0.99	0.90	0.99	0.99	0.96	0.98	0.99
	0.84	0.92	0.95	0.88	0.98	0.97	0.95	0.88	0.96
	0.45	0.90	0.96	0.81	0.97	0.95	0.84	0.90	0.94
Swing Duration	0.97	0.98	0.99	0.96	0.99	0.98	0.93	0.96	0.97
	0.82	0.91	0.92	0.92	0.96	0.93	0.82	0.80	0.87
	0.38	***0.65***	0.78	0.72	0.94	0.86	***0.58***	***0.59***	0.72
Number of Strong Correlations	***10***	12	12	13	14	14	13	11	12
	8	12	12	11	14	14	14	9	12
	6	10	12	10	12	12	8	9	10

[B3] Intra-class correlation of gait features between differing timeframes in PwMS. Non-shaded cells are the correlations between feature medians, grey shaded cells are the correlation between feature 95th percentiles and blue shaded cells are the correlation between feature coefficient of variation (CV) values. Features and/or timeframes are abbreviated as follows: AP: Anterior-posterior; ML: Medio-lateral; RMS: Root Mean Square; 1D = 1 Day, 2D = 2Days, 1W = 1 Week, WE = Weekend, WD = Weekday, etc. Correlations are considered strong if the reported value is greater than or equal to 0.70, values below are bolded and italicized.

Table B4: Intra-class Correlation Wear Duration Sway Analysis in Persons with MS

Sway Feature (13 total)	**1D v 1W**	**2D v 1W**	**3D v 1W**	**WE v WD**	**1W v 2W**	**1W v 3W**	**1W v 4W**	**1W v 5W**	**1W v 6W**

Area	***0.49***	0.86	0.94	0.93	0.96	0.94	0.84	0.87	0.93
	0.40	***0.38***	0.95	0.91	0.98	0.95	0.92	0.81	0.96
	0.46	0.81	0.89	0.85	0.92	0.90	0.88	0.86	0.91
Centroidal Frequency	0.84	0.96	0.99	0.95	1.00	0.99	0.97	0.98	0.99
	0.73	0.90	0.97	0.88	0.99	0.99	0.96	0.97	0.98
	0.49	0.90	0.95	0.91	0.96	0.96	0.93	0.92	0.95
Distance	0.75	0.94	0.97	0.94	0.97	0.97	0.83	0.89	0.94
	0.46	0.86	0.95	0.76	0.96	0.92	0.89	0.88	0.92
	0.47	0.94	0.96	0.93	0.96	0.96	0.77	0.86	0.94
50th Percentile Frequency	0.80	0.96	0.99	0.96	0.98	0.97	0.94	0.95	0.96
	0.74	0.94	0.98	0.94	0.98	0.98	0.94	0.96	0.97
	0.52	0.84	0.88	0.78	0.96	0.93	0.90	0.83	0.91
95th Percentile Frequency	0.74	0.97	0.99	0.97	0.98	0.96	0.95	0.94	0.96
	0.64	0.94	0.98	0.94	0.98	0.98	0.95	0.97	0.98
	0.32	0.71	0.90	0.79	0.95	0.91	***0.56***	0.88	0.89
Frequency Dispersion	***0.64***	0.95	0.96	0.95	0.98	0.98	0.91	0.93	0.97
	0.66	0.91	0.96	0.89	0.97	0.95	0.93	0.92	0.94
	0.46	0.87	0.93	0.93	0.97	0.96	0.86	0.92	0.95
Jerk	0.85	0.88	0.96	0.92	0.98	0.95	0.71	0.84	0.92
	0.68	0.87	0.96	0.87	0.96	0.93	0.76	0.83	0.90
	0.27	0.79	0.87	0.90	0.93	0.91	0.87	0.83	0.90
Mean Period	***0.01***	0.97	0.97	0.97	0.90	0.90	0.84	0.83	0.94
	0.43	0.78	0.96	0.90	0.94	0.96	0.95	0.72	0.96
	0.56	***0.67***	0.87	***0.49***	0.84	0.83	***0.40***	***0.41***	0.75
Mean Velocity	***0.56***	0.97	0.96	0.96	0.97	0.96	0.95	0.94	0.96
	0.85	0.80	0.93	0.94	0.95	0.97	0.94	0.86	0.97
	0.72	0.94	0.97	0.90	0.92	0.90	0.76	0.93	0.88
Path	***0.51***	0.97	0.96	0.98	0.97	0.96	0.95	0.93	0.96
	0.84	0.79	0.94	0.94	0.95	0.97	0.94	0.86	0.97
	0.62	0.91	0.96	0.94	0.94	0.93	0.83	0.93	0.91
Power	***0.24***	0.96	0.95	0.98	0.96	0.97	0.97	0.92	0.97
	0.82	0.72	0.94	0.94	0.95	0.97	0.92	0.82	0.97
	0.77	0.91	0.96	0.93	0.94	0.92	0.76	0.92	0.88
Range	***0.65***	0.91	0.94	0.94	0.97	0.97	0.82	0.88	0.96
	0.48	0.94	0.97	0.95	0.98	0.96	0.88	0.91	0.94
	0.39	0.77	0.87	***0.68***	0.91	0.87	0.85	***0.58***	0.83
RMS	***0.49***	0.97	0.96	0.97	0.97	0.97	0.95	0.93	0.97
	0.83	0.79	0.95	0.94	0.95	0.97	0.93	0.86	0.98
	0.61	0.92	0.96	0.95	0.95	0.94	0.93	0.93	0.92
Number of Strong Correlations	5	13	13	13	13	13	13	13	13
	6	12	13	13	13	13	13	13	13
	1	12	13	11	13	13	11	11	13

[B4] Intra-class correlation of sway features between differing timeframes in PwMS. Non-shaded cells are the correlations between feature medians, grey shaded cells are the correlation between feature 95th percentiles and blue shaded cells are the correlation between feature coefficient of variation (CV) values. Timeframes are abbreviated as follows: 1D = 1 Day, 2D = 2Days, 1W = 1 Week, WE = Weekend, WD = Weekday, etc. Correlations are considered strong if the reported value is greater than or equal to 0.70, values below are bolded and italicized.

Appendix C

TABLE C1
PRM AND POSTURAL SWAY FEATURES CORRELATIONS FOR CHEST SENSOR LOCATION USING SO METHOD

	Age			EDSS			ABC			MFIS			MSWS		
Feature	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS
Jerk	-	-	0.41	*0.28*	0.43	0.48	-	-0.40	-0.39	-	-	*0.28*	-	0.39	0.31
Dist	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
RMS	-	-	-	-	-	-	-	-	-0.32	-	-	-	-	-	0.38
Path	-	-	-	-	-	-	-	-	-0.32	-	-	-	-	-	0.36
Range	-	-	-	-	0.32	*0.27*	-	-0.32	*-0.26*	-	-	*0.26*	-	0.36	*0.28*
MV	-	-	-	-	-	-	-	-	-0.31	-	-	-	-	-	0.36
MF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Area	-	-	-	-	-	-	-	-	-0.35	-	-	-	-	*0.29*	0.43
Pwr	-	-	-	-	-	-	-	-	-0.32	-	-	-	-	-	0.38
F50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F95	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CF	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
ApEn	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LyExp	-	-	-	-	-	-	-	-	-	-	-	-	*-0.30*	-	-

Significant and approaching significant correlations of postural sway features with PRMs for the Single Observation (SO) method. Approaching significant correlations are italicized. Eyes-Open (EO); Eyes-Closed (EC); Tandem Standing (TS).

TABLE C2
PRM AND POSTURAL SWAY FEATURES CORRELATIONS FOR SACRUM SENSOR LOCATION USING SO METHOD

	Age			EDSS			ABC			MFIS			MSWS		
Feature	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS
Jerk	-	-	0.37	0.42	0.50	0.49	-0.34	-0.50	-0.37	-	0.32	0.36	-	0.48	*0.30*
Dist	0.33	-	-	-	*0.27*	0.33	-	-	-	-	-	-	-	-	-
RMS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Path	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Range	0.38	0.31	0.34	0.36	0.36	0.47	-0.38	-0.38	-0.42	-	0.33	*0.27*	0.33	0.51	0.40
MV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MF	*-0.29*	-	-	-	-	-0.34	-	-	-	-	-	-	-	-	-
Area	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pwr	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F50	-	-	-	*0.27*	-	*0.29*	-	-	-	-	-	-	-	-	*0.27*
F95	-	-	-	-	-	*0.27*	-	-	-	-	-	-	-	-	-
CF	-	-	-	-	-	-	-	-	-	-	-	-0.32	-	-	-
FD	-	*-0.27*	-	-	-	-	-	-	-	-	-	-	-	-	-
ApEn	-	-0.38	-	-	-	-	-	-	-	-	-	-	-	-	-
LyExp	-	-	-	-	-	0.44	-	-	-	-	-	-	-	-	*0.27*

Significant and approaching significant correlations of postural sway features with PRMs for the Single Observation (SO) method. Approaching significant correlations are italicized. Eyes-Open (EO); Eyes-Closed (EC); Tandem Standing (TS).

TABLE C3
PRM AND POSTURAL SWAY FEATURES CORRELATIONS FOR CHEST SENSOR LOCATION USING ID METHOD

Feature	Summary Statistic	Age			EDSS			ABC			MFIS			MSWS		
		EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS
Jerk	p5	-	*0.30*	0.33	-	0.46	0.54	-	-0.38	-0.47	-	-	0.37	-	-	0.38
	p25	-	*0.30*	0.32	*0.27*	0.45	0.49	-	-0.36	-0.43	-	-	0.33	-	-	0.33
	Med	-	-	0.33	-	0.45	0.49	-	-0.36	-0.44	-	-	0.35	-	-	0.34
	p75	-	-	0.31	-	0.45	0.43	-	-0.37	-0.38	-	-	0.31	-	-	*0.30*
	p95	-	-	0.34	*0.28*	0.46	0.43	-	-0.36	-0.37	-	-	0.31	-	-	*0.30*
	STD	-	-	*0.27*	-	0.34	-	-	-0.32	-	-	-	-	-	-	-
Dist	p5	-	-	-	-	-	-	-	0.32	-	-	-0.43	-	-	-0.43	-
	p25	-	-	-	-	-	-	-	0.28	*-0.31*	*-0.28*	-0.48	-	-	-0.48	*0.27*
	Med	-	-	-	-	-	-	-	-	*-0.27*	-0.38	-0.47	-	*-0.30*	-0.47	-
	p75	-	-	-	-	-	-	-	-	-	-0.37	-0.45	-	*-0.31*	-0.45	*0.28*
	p95	-	-	-	-	-	-	-	-	-	-0.35	-0.32	-	-	-0.32	-
	STD	-	-	-	-	-	-	-	-	-	-0.33	*-0.30*	-	-	*-0.30*	*0.29*
RMS	p5	-	-	-	-	-	-	-	-	-0.34	-	-	-	-	-	0.35
	p25	-	-	-	-	-	-	-	-	-0.36	-	-	-	-	-	0.37
	Med	-	-	-	-	-	-	-	-	-0.35	-	-	-	-	-	0.36
	p75	-	-	-	-	-	-	-	-	-0.34	-	-	-	-	-	0.35
	p95	-	-	-	-	-	-	-	-	-0.35	-	-	-	-	-	0.36
	STD	-	-	-	-	-	-	-	-	-	*-0.27*	-	-	-	-	-
Path	p5	-	-	-	-	-	-	-	-	-0.31	-	-	-	-	-	0.33
	p25	-	-	-	-	-	-	-	-	-0.33	-	-	-	-	-	0.33
	Med	-	-	-	-	-	-	-	-	-0.34	-	-	-	-	-	0.34
	p75	-	-	-	-	-	-	-	-	-0.33	-	-	-	-	-	0.34
	p95	-	-	-	-	-	-	-	-	-0.35	-	-	-	-	-	0.36
	STD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*0.27*
Range	p5	-	-	-	-	*0.28*	-	-	-0.27	-0.32	-	-	-	-	-	*0.26*
	p25	-	-	-	-	-	0.31	-	-	-0.36	-	-	-	-	-	*0.31*
	Med	-	-	-	-	-	*0.27*	-	-	-0.32	-	-	-	-	-	*0.28*
	p75	-	-	-	-	-	-	-	-	*-0.29*	-	-	-	-	-	-
	p95	-	-	-	-	-	-	-	-	-0.31	-	-	-	-	-	0.32
	STD	-	-	0.35	-	-	-	-	-	-	-	-	-	-	-	-
MV	p5	-	-	-	-	-	-	-	-	-	-	-	-	*0.28*	-	*0.27*
	p25	-	-	-	-	-	-	-	-	*-0.27*	-	-	-	-	-	*0.29*
	Med	-	-	-	-	-	-	-	-	*-0.28*	-	-	-	-	-	*0.29*
	p75	-	-	-	-	-	-	-	-	*-0.29*	-	-	-	-	-	*0.30*
	p95	-	-	-	-	-	-	-	-	-0.32	-	-	-	-	-	0.34
	STD	-	-	-	-	-	-	-	-	-0.32	-0.36	-	-	-	-	0.34
MF	p5	-	-	-	-	-	-	-	-	-	0.32	0.37	-	0.33	0.37	-
	p25	-	-	-	-	-	-	-	-0.25	-	*0.28*	0.45	-	0.35	0.45	-
	Med	-	-	-	-	-	-	-	-0.26	-	-	0.43	-	*0.31*	0.43	-
	p75	-	-	-	-	-	-	-	-0.33	-	-	0.40	-	-	0.40	-
	p95	-	-	-	-	*0.27*	-	-	-0.37	-	-	0.40	-	-	0.40	-
	STD	-	-	-	-	-	-	-	-0.32	-	-	0.36	-	-	0.36	-
Area	p5	-	-	-	-	-	-	-	-	-0.32	-	-	-	-	-	0.38
	p25	-	-	-	-	-	-	-	-	-0.33	-	-	-	-	-	0.37
	Med	-	-	-	-	-	-	-	-	-0.33	-	-	-	-	-	0.39
	p75	-	-	-	-	-	-	-	-	-0.33	-	-	-	-	-	0.39
	p95	-	-	-	-	-	-	-	-	-0.32	-	-	-	-	-	0.39
	STD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.33
Pwr	p5	-	-	-	-	-	-	-	-	-0.34	-	-	-	-	-	0.35
	p25	-	-	-	-	-	-	-	-	-0.36	-	-	-	-	-	0.37
	Med	-	-	-	-	-	-	-	-	-0.35	-	-	-	-	-	0.36

	p75	-	-	-	-	-	-	-	-	-0.34	-	-	-	-	-	0.35
	p95	-	-	-	-	-	-	-	-	-0.35	-	-	-	-	-	0.36
	STD	-	-	-	-	-	-	-	-	*-0.29*	-	-	-	-	-	0.31

TABLE C3 (CONT.)
PRM AND POSTURAL SWAY FEATURES CORRELATIONS FOR CHEST SENSOR LOCATION USING ID METHOD

		Age			EDSS			ABC			MFIS			MSWS		
Feature	Summary Statistic	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS
F50	STD	-	-	-	-	-	*0.26*	-	-	-	-	-	-	*-0.29*	-	-
F95	Med	-	-	-	-	-	-	-	-	-	-	-	*0.27*	-	-	-
CF	N/A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FD	STD	-	-	-	-	-	-	-	-	-	-	-	0.31	-	-	-
ApEn	p5	-	-	-	-	-	-	-	-	-	*0.27*	-	-	-	-	-
	p25	-	-	-	-	*0.26*	-	-	-	-	*0.28*	0.34	-	-	0.34	-
	Med	-	-	-	-	-	-	-	-	-	*0.29*	*0.27*	-	-	*0.27*	-
	p75	-	-	-	-	-	-	-	-	-	-	*0.26*	-	-	*0.26*	-
LyExp	p25	-	-	-	-	-	-	-	0.26	-	-	-	-	-	-	-
	Med	-	-	-	-	-	-	-	0.28	-	-	-	-	-	-	-
	STD	0.33	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Significant and approaching significant correlations of postural sway features with PRMs for the Individualized Distributions (ID) method. Approaching significant correlations are italicized. Eyes-Open (EO); Eyes-Closed (EC); Tandem Standing (TS).

TABLE C4
PRM AND POSTURAL SWAY FEATURES CORRELATIONS FOR SACRUM SENSOR LOCATION USING ID METHOD

		Age			EDSS			ABC			MFIS			MSWS		
Feature	Summary Statistic	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS
Jerk	p5	-	-	0.34	0.33	0.51	0.56	-0.32	-0.41	-0.45	-	-	0.41	-	0.39	0.37
	p25	-	-	0.35	0.33	0.50	0.56	-0.32	-0.41	-0.45	-	-	0.42	-	0.40	0.36
	Med	-	-	0.32	0.33	0.50	0.52	-0.33	-0.47	-0.42	-	-	0.40	-	0.44	0.32
	p75	-	-	*0.31*	*0.31*	0.53	0.50	-0.32	-0.49	-0.40	-	*0.28*	0.40	-	0.47	*0.30*
	p95	-	-	0.32	*0.31*	0.51	0.49	*-0.31*	-0.48	-0.39	-	*0.29*	0.40	-	0.47	*0.31*
	STD	-	-	-	*0.31*	0.41	-	*-0.30*	-0.42	-	-	0.44	*0.28*	-	0.46	-
Dist	p5	-	-	-	-	-	0.36	-	-	*-0.27*	-	-	-	-	-	-
	p25	-	-	-	-	-	0.36	-	-	-0.36	-	-	-	-	-	-
	Med	-	-	-	-	-	-	-	-	-	-	-	-	*0.28*	-	-
	p75	-	-	-	-	-	-	*-0.32*	-	-	-	-	-	*0.29*	-	-
	p95	-	-	-	-	-	-	*-0.27*	-	-	-	-	-	-	-	-
	STD	-	-	-	-	-	-	-0.32	-	-	-	-	-	-	-	-
RMS	N/A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Path	N/A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Range	p5	-	0.37	-	-	0.42	0.44	-0.37	-0.42	-0.52	-	-	0.37	0.33	0.49	0.46
	p25	-	0.33	-	-	0.35	0.48	-0.41	-0.37	-0.48	-	-	*0.29*	0.35	0.45	0.37
	Med	-	0.34	-	-	0.34	0.46	-0.37	-0.35	-0.44	-	-	-	0.32	0.42	0.37
	p75	-	0.37	-	-	0.35	0.42	*-0.30*	-0.32	-0.40	-	-	-	*0.29*	0.43	0.33
	p95	-	0.35	*0.30*	-	*0.31*	0.46	*-0.27*	*-0.30*	-0.46	-	-	*0.31*	-	0.41	0.44
	STD	-	-	-	-	-	0.35	-	-	*-0.27*	-	-	-	-	-	0.34
MV	STD	-	-	-	-	-	-	-	-	*-0.28*	-0.33	*-0.29*	-	-	-	-
MF	p75	-	-	-	-	-	-0.34	-	-	0.31	-	-	-	-	-	-
	p95	-	-	-	-	-	-0.31	-	-	-	-	-	-	-	-	-
Area	p25	-	-	-	-	-	-	-0.28	-	-	-	-	-	-	-	-
	Med	-	-	-	-	-	-	*-0.29*	-	-	-	-	-	-	-	-
	p75	-	-	-	-	-	-	*-0.28*	-	-	-	-	-	-	-	-
	p95	-	-	-	-	-	-	*-0.30*	-	-	-	-	-	-	-	-
	STD	-	-	-	-	-	-	-0.37	-	-	-	-	-	*0.29*	-	-
Pwr	N/A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F50	p25	-	-	-	-	-	0.31	-	-	-0.27	-	-	-	-	-	-
	Med	-	-	-	-	-	*0.31*	-	-	*-0.28*	-	-	*0.27*	-	-	-
	p75	-	-	-	-	-	*0.31*	-	-	*-0.28*	-	-	-	-	-	*0.27*
	p95	-	-	-	-	-	*0.29*	-	-	-	-	-	-	-	-	*0.27*
	STD	-	-	-	-	-	*0.29*	-	-	*-0.27*	-	-	-	-	-	*0.29*
F95	p5	-	-	-	-	-	*0.29*	-	-	-	-	-	*0.28*	-	-	-
	p25	-	-	-	-	-	*0.29*	-	-	*-0.27*	-	-	*0.28*	-	-	-
	Med	-	-	-	-	-	*0.28*	-	-	*-0.27*	-	-	*0.27*	-	-	-
	p75	-	-	-	-	-	*0.29*	-	-	-	-	-	-	-	-	-
	p95	-	-	-	-	-	-	-	-	-	-	-	-	-	-	*0.27*
	STD	-	-	-	-	-	*0.30*	-	-	*-0.31*	-	-	-	-	-	0.33
CF	p5	-	-	-	-	-	-	-	-	*0.27*	-	-	*-0.29*	-	-	-
	p75	-	-	-	-	-	-	-	-	-	-	-	*-0.27*	-	-	-
	STD	-	-	-	-	-	*0.29*	-	-	-0.31	-	-	-	-	-	0.31
FD	p5	-	*-0.31*	-	-	-	-	-	-	-	-	-	-	-0.35	-	-
	p25	-	*-0.31*	-	-	-	-	-	-	-	-	-	-	-	-	-
	Med	-	*-0.27*	-	-	-	-	-	-	-	-	-	-	-	-	-
	p95	-	-	-	-	-	*0.27*	-	-	-	-	-	*0.29*	-	-	-
ApEn	p5	-	*-0.30*	-	-	-	-	-	-	-	-	-	-	-	-	-
	p25	-	-0.32	-	-	-	-	-	-	-	-	-	-	-	-	-
	Med	-	-0.34	-	-	-	-	-	-	-	-	-	-	*-0.27*	-	-
	p75	-	-0.36	-	-	-	-	-	-	-	-	-	-	*-0.31*	-	-
	p95	-	-0.37	-	-	-	-	-	-	-	-	-	-	-0.33	-	-
	STD	-	-	-	-0.44	-	-	0.39	-	-	-0.45	-0.34	-	-0.38	-0.34	-

TABLE C4 (CONT.)
PRM AND POSTURAL SWAY FEATURES CORRELATIONS FOR SACRUM SENSOR LOCATION USING ID METHOD

		Age			EDSS			ABC			MFIS			MSWS		
Feature	Summary Statistic	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS	EO	EC	TS
LyExp	p5	-	-	-	-	-	*0.29*	-	-	*-0.31*	-	-	-	-	-	0.33
	p25	-	-	-	-	-	0.43	-	-	-0.36	-	-	0.33	-	-	0.31
	Med	-	-	-	-	-	0.40	-	-	-0.42	-	-	*0.30*	-	-	0.37
	p75	-	-	-	-	-	*0.31*	-	-	*-0.29*	-	-	-	-	-	*0.30*
	p95	-	-	-	-	-	0.34	-	-	*-0.28*	-	-	-	-	-	*0.28*

Significant and approaching significant correlations of postural sway features with PRMs for the Individualized Distributions (ID) method. Approaching significant correlations are italicized. Eyes-Open (EO); Eyes-Closed (EC); Tandem Standing (TS).

TABLE C5
FALLER NON-FALLER COMPARISONS FOR CHEST AND SACRUM DERIVED POSTURAL SWAY FEATURES

		Chest				Sacrum			
Feature	Statistic	F Med	NF Med	p	Cohen's d	F Med	NF Med	p	Cohen's d
Eyes-Closed Single Observation									
Range	N/A	-	-	-	-	0.05	0.03	0.08	0.42
Eyes-Open Subject Specific Distribution									
MV	STD	0.09	0.14	0.06	-0.35	-	-	-	-
Eyes-Closed Subject Specific Distribution									
Jerk	STD	-	-	-	-	0.00	0.00	0.08	0.48
Range	p95	-	-	-	-	0.06	0.04	0.08	0.47
	STD	-	-	-	-	0.01	0.00	0.05	0.36
Tandem Standing Subject Specific Distribution									
Jerk	p75	0.10	0.08	0.06	0.36	-	-	-	-
	p95	0.11	0.08	0.06	0.41	-	-	-	-
	STD	0.01	0.00	0.10	0.48	-	-	-	-
Range	p5	-	-	-	-	0.06	0.04	0.05	0.59
MV	STD	0.19	0.09	0.05	0.60	-	-	-	-
Area	p5	0.06	0.03	0.08	0.33	-	-	-	-
	p25	0.07	0.04	0.07	0.34	-	-	-	-
	Med	0.09	0.04	0.07	0.29	-	-	-	-
	p75	0.09	0.04	0.07	0.32	-	-	-	-
CF	Med	0.70	0.71	0.05	-0.56	-	-	-	-
	p75	0.70	0.71	0.05	-0.51	-	-	-	-
FD	p5	1.77	1.71	0.06	0.36	-	-	-	-
	p25	1.82	1.76	0.05	0.39	-	-	-	-
	Med	1.85	1.78	0.04	0.43	-	-	-	-
	p75	1.90	1.81	0.05	0.46	-	-	-	-
	p95	1.96	1.85	0.04	0.48	-	-	-	-

Significant and nearing significant differences between fallers (F) and non-fallers (NF) for features of postural sway. Median (Med), p values(p), and Cohen's d reported for each test. Results provided for features computed from chest and sacrum.