Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Task-Specific Assessment and Characterization of Perception Accuracy in People with Multiple Sclerosis: A Comparative Study with Healthy Controls.

Clare-Joyce Fomonyuy Ngoran

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Data Science

SUPERVISOR :
dr. Anna IVANOVA
Prof. dr. Peter FEYS
De heer Gianluca FLORIO

2024
2025

Maastricht University

Faculty of Sciences

School for Information Technology

Master of Statistics and Data Science

Master's thesis

Task-Specific Assessment and Characterization of Perception Accuracy in People with Multiple Sclerosis: A Comparative Study with Healthy Controls.

Clare-Joyce Fomonyuy Ngoran

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Data Science

SUPERVISOR :

dr. Anna IVANOVA

Prof. dr. Peter FEYS

De heer Gianluca FLORIO

Contents

List of Figures

Abstract

Demyelination of nerves in people with multiple sclerosis (PwMS) disrupts communication in the central nervous system, leading to deficits in perception. Perception accuracy in PwMS has been evaluated through self-report retrospective questionnaires and lab-based experiments. While these tools provide useful information, they fail to fully capture the real-time perceptual challenges patients encounter in functional activities. This thesis advances the field by focusing on the task-specific assessment of perception accuracy in PwMS, offering a more objective alternative reflecting daily experiences.

Data collection was conducted during a 6-minute walking test in a 30-meter corridor. Participants were equipped with six APDM wearable sensors to record gait speed and held a pinch sensor to indicate when they perceived a change in walking speed. Using SciPy *find_peaks* method, with parameters informed by prior research, objective speed change points were identified and then time-aligned with the timestamped pinch data. A $\pm$2s tolerance window was applied: pinches within this window of a speed change were classified as true positives, those outside as false positives, and missed events as false negatives. This yielded a binary outcome capturing perception accuracy for subsequent analyses.

Given the binary and repeated nature of the outcome, Generalized Estimating Equations and Generalized Linear Mixed Models were used to model the group effect on the outcome while adjusting for age, sex, and BMI. Linear Mixed Models were also used to model subjective data collected via the MAIA, ISAQ and PAS questionnaires. A post-protocol administration subjective confidence measure was also evaluated for its effect on perception accuracy. Results revealed that subjective measures of perception, captured via questionnaires, did not significantly differentiate PwMS from healthy controls and the confidence did not have a significant effect on the perception accuracy. In contrast, task-specific features derived from walking data provided more nuanced insights, highlighting subtle but meaningful differences in perceptual accuracy.

This work highlights how questionnaire-based assessments, while informative, may not fully capture perception in PwMS, and emphasizes the added value of integrating sensor-derived, task-specific real-time measures into perception assessment in PwMS. By building robust data pipelines from collection through modeling, this thesis lays the groundwork for data-driven rehabilitation strategies. Such approaches hold promise for supporting energy-efficient engagement in exercise, addressing kinesiophobia, and ultimately promoting engagement in exercise and imrpoving quality of life for PwMS.

1 Introduction

1.1 Background

Multiple Sclerosis (MS) is a prevalent and complex autoimmune neuropathology of the central nervous system (CNS) characterized by inflammatory demyelination of neural tissues. MS affects approximately 2.7 million people worldwide and this number is expected to grow exponentially [1, 2]. Although the exact cause of MS remains indefinite, the resulting damage disrupts the intricate communication network between the brain and the rest of the body [1, 3]. This neurological disruption has major detrimental impacts on various bodily functions, including interfering with the body's capacity for allostatic adjustments and leading to a wide array of exhausting symptoms [3, 4]. The findings of several studies report that people with MS (PwMS) have objective and perceived gait impairments compared to healthy controls [5] and about 50% of PwMS suffer emotional problems, particularly depression and anxiety [1, 6, 7]. Among these, impaired balance which results in kinesiophobia emerges as a critical concern for PwMS.

The clinical manifestation of MS is highly heterogeneous [1, 2, 8]. The symptoms fluctuate in severity and presentation, contributing to the unpredictable nature of the disease [1]. Critically, motor and cognitive disabilities are frequently observed and contribute significantly to improper interpretation of sensory stimuli and delayed or inappropriate motor responses which results in reduced functional safety[5, 6, 9]. These safety problems do not only include physical injuries, but also a fear of movement (kinesiophobia), further restricting participation in exercise activities and accelerating a decline in their overall quality of life [6, 1, 10].

Within the spectrum of motor impairments experienced by PwMS, balance deficits, impairment of walking function, and mobility are perceived by people with MS to be the most devastating [9]. Maintaining dynamic postural stability (to sustain an efficient walking pattern) is a complex process that relies on the seamless integration of sensory information (visual, vestibular, and proprioceptive), central processing, and the generation of appropriate motor responses [11]. As a result, it becomes harder for some PwMS to adapt to new experiences, with possible hypoactivity or hyperactivity due to under- or overestimation of true physical capacity [12, 2, 13].

Although the presence of impairments in PwMS is well documented, a deeper understanding of the underlying mechanisms and contributing factors is crucial to developing effective rehabilitation strategies. One critical aspect that warrants further investigation is the accuracy of perception, particularly in relation to task-specific demands [11, 14, 12]. The ability to accurately perceive one's body position and movement in space is fundamental for maintaining balance and executing coordinated movements [15]. However, the extent to which these perceptual inaccuracies are task dependent and how they relate to objective

measures of motor performance remains an area requiring further exploration.

Therefore, this thesis aims to address the critical gap in perception accuracy in daily activities by exploring "task-specific evaluation and characterization of perception accuracy in PwMS compared to healthy controls" during a 6-minute walking test (6MWT). The protocol is designed to evoke perceptual judgments related to movement and perception. Furthermore, the correspondence between measured and subjective dimensions of perceptive precision will be evaluated.

1.2 Problem Statement

Demyelination and neurodegeneration of the CNS in PwMS disrupts communication in the network and affects task-specific perceptual abilities of patients. Historically, the perception accuracy of PwMS has been largely assessed through self-report questionnaires and clinical scales that rely on retrospective reflections. Although these tools provide insightful information, they do not fully capture the real-time perceptive challenges PwMS face in daily activities. Task-specific assessments, which assess perception accuracy during actual functional tasks, such as walking, offer a more objective and ecologically valid measure of perceptual deficits. However, despite their advantages, such task-specific approaches remain underutilized in MS research, particularly for dynamic and prolonged activities such as walking. The challenge is not just to collect this data, but to structure, clean, and analyze it in a way that reveals subtle, yet critical, patterns of neurological impairments. Developing robust data pipelines and analytical frameworks is essential to transform raw sensor data into meaningful insights that can inform new diagnostic and therapeutic strategies for PwMS.

1.3 Study Objectives

- Characterize and compare the accuracy of perceiving changes in walking speed between PwMS and HC during a 6MWT.
- Explore the relationship between perception accuracy and demographic factors.
- Evaluate the the correspondence between subjective and objective perception.

1.4 Research Questions

- Do PwMS and HCs differ in their perceptive accuracy for detecting changes in walking speed during a self-paced experimental walking task at their maximum walking speed?
- How do PwMS and HCs differ in the correspondence between objective measures of perceptive accuracy for walking speed changes during a 6MWT and their subjective perceptions reported via questionnaires?

2 Literature Review

Impaired proprioception and altered gait stability are well-recognized and studied challenges among the many prominent and disabling symptoms in PwMS, [6, 9, 16, 17]. These difficulties can affect walking safety and overall quality of life. However, most of the research has focused on objective gait measurement or lab-based proprioceptive tests, rarely integrating how individuals perceive their own movement during sustained motor tasks (e.g walking).

Prior research has demonstrated that proprioceptive pathways in PwMS are compromised. Mental chronometry tasks, which compare imagined and actual walking times, offer indirect insights into perception accuracy. Wajda et al. (2021) found that PwMS exhibited significantly larger discrepancies between imagined and real walking durations, particularly under complex conditions such as dual-tasking [18]. HCs, in contrast, improved their internal predictions following the task, whereas PwMS failed to update their internal model of performance. Furthermore, in PwMS, greater impairment in motor imagery correlated with lower balance confidence and higher disability scores. This study underscores that PwMS have diminished internal awareness of their walking performance. However, since mental chronometry assesses perception in an offline, imagined context, it does not capture real-time perceptual accuracy during actual movement. This limitation highlights the importance of studies which directly assess real-time perception of speed changes during walking.

To understand walking dynamics outside controlled laboratory settings, researchers have turned to continuous monitoring techniques. Smith et al. (2018) used Microsoft Kinect cameras to monitor walking speed in home environments, calculating summary statistics like median speed and variance to capture walking variability over time [19]. Grobelny et al. (2017) also employed Kinect to assess maximum walking speed in PwMS, demonstrating strong test-retest reliability [20]. Meanwhile, Pau et al. (2016) used accelerometers during a 25-foot walk to quantify gait features and found meaningful correlations with self-reported walking difficulties [21]. These studies effectively capture physiological changes in movement, but they overlook whether individuals are aware of their own walking fluctuations as they occur.

Researchers have begun using more advanced signal-processing techniques to analyze gait dynamics in greater detail. For instance, Quijoux et al. (2023) created a comprehensive "semiogram" using 17 features extracted from inertial sensor data, calculating reliability via intraclass correlation coefficients (ICCs) and using clustering to identify groups based on gait profiles [22]. Similarly, Gholami et al. (2015) applied dynamic time warping to Kinect-based gait data to detect instability patterns [20]. While powerful, these methods still neglect the participant's perception of movement changes, leaving out the importance of self-awareness during walking.

Despite the established understanding of sensory and motor impairments associated with MS, our knowledge of the accuracy with which PwMS perceives dynamic movement parameters such as speed change during functional activities such as walking is limited. Perception is a multifaceted ability that encompasses several distinct dimensions: the accuracy which looks at the degree to which perception corresponds to actual performance, then the attention, which examines the level of focus directed toward perceptive cues, followed by the sensibility that looks at an individual's beliefs about their own perceptive accuracy and attention, and finally the awareness that explores the alignment between objectively measured and subjectively reported perceptual experiences [23].

Traditional statistical methods in gait analysis typically involve correlation coefficients to link gait speed or variability with clinical scales, or agreement measures like Bland–Altman plots and ICCs to validate sensor-based gait measures. Although useful, these approaches do not directly model the accuracy of participants' perception indexed by real-time responses to gait changes.

The current study addresses these gaps by combining several methodological advances. Participants wore 6 sensors to record walking speed, and their real-time perceptual awareness was captured through an easy-to-use hand-held sensor (A quick hand squeeze to indicate perception of speed change) aligned with ongoing walking speed changes. To handle the repeated, nested nature of the data, the study uses Generalized Estimating Equations (GEE) and Generalized Linear Mixed Models (GLMM), which account for correlations within each participant's responses. This is a major improvement over past analyses that often relied on generalized linear models or agreement measures without acknowledging the clustered data structure. Model diagnostics were carried out ensure stable and interpretable estimates, following recommendations from Molenberghs and Verbeke (2005)[24].

This approach closely reflects the daily experiences of PwMS. By evaluating and characterizing perceptive abilities during prolonged functional tasks such as walking, the findings can inform walking rehabilitation strategies, for example, programs designed to optimize energy expenditure and engagement during exercise. This can promote consistency in activity and allow PwMS to experience the benefits of rehabilitation.

3 Materials and Methodology

This section outlines the data source, the data collection approach, and the criteria used to include or exclude participants from the study. It also describes the steps taken to explore, pre-process, and transform the data to ensure suitability for statistical analysis, followed by an overview of the analytical methods and tools employed.

3.1 Data Sources and Description

This study draws on a subset of quantitative and subjective data collected as part of a broader research initiative that is still in progress, with data collection currently ongoing at REVAL, a research group of the Faculty of Rehabilitation Sciences at UHasselt which "focuses on interdisciplinary and technology-supported research within rehabilitation, with attention to clinical applications". The data sources include sensor outputs, standardized questionnaires and self-reported demographics data.

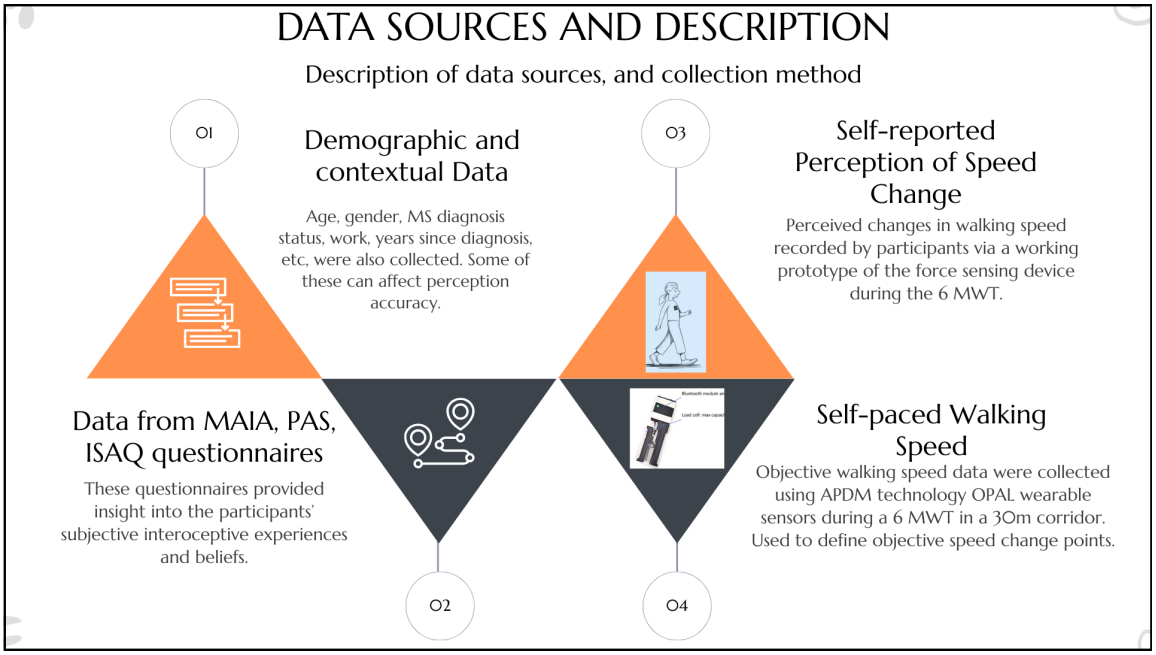


Figure 1: Data sources and descriptions.
(1) The objective self-paced speed data collected with APDM sensors and used to identify actual speed change point against which perception is assessed. (2) Timestamped data of self-reported perception events,collected synchronously with the walking speed allowing for a comparison with the objectively measured speed changes. (3) Standardized questionnaires for measurement of multiple dimensions of interoception by self-report. (4) Demographics data.

Walking Speed Data

Objective walking speed data were collected using APDM technology OPAL wearable sensors strapped to the participant's body. The participants performed a 6-minute walking test (6MWT) in a 30m corridor, walking back and forth as fast and safe as possible, with the strapped sensors continuously measuring and recording timestamped walking speed data. These data were used to identify actual changes in walking speed against which perception accuracy could be assessed.

Perception Input Data (Pinch Sensor)

Perceived changes in walking speed were recorded via a working prototype of the force sensing device (from here on referred to as the pinch sensor). Participants were instructed to pinch the device each time they perceived a change in their walking speed during the 6-minute walk. This produced a timestamped data of self-reported perception events, allowing for a comparison with the objectively measured speed changes.

Questionnaire Data

Several questionnaires (15) were administered as part of the overall protocol. Among which were; MAIA (Multidimensional Assessment of Interoceptive Awareness) and ISAQ (Interoceptive Sensitivity and Attention Questionnaire) [25] that measure self-beliefs over interoceptive abilities and PAS (Postural Awareness Scale) measuring self-beliefs over proprioceptive abilities [26].

Demographic and Contextual Data

Additional data included participant demographic information (age, gender, MS diagnosis status, work, years since diagnosis, etc), as well as self-reported confidence in the accuracy of their pinch reports immediately after the experiment on a VAS Scale (10-100%) levels. These variables were used to explore whether perception accuracy varied with participant characteristics.

3.2 Inclusion and Exclusion Criteria

Participants were recruited through the Flemish MS rehabilitation centers in Melsbroek (NMSC) and Pelt (Noorderhart RMSC), as well as the REVAL research center at UHasselt and its established network. All participants provided informed consent, and no economic compensation was offered.

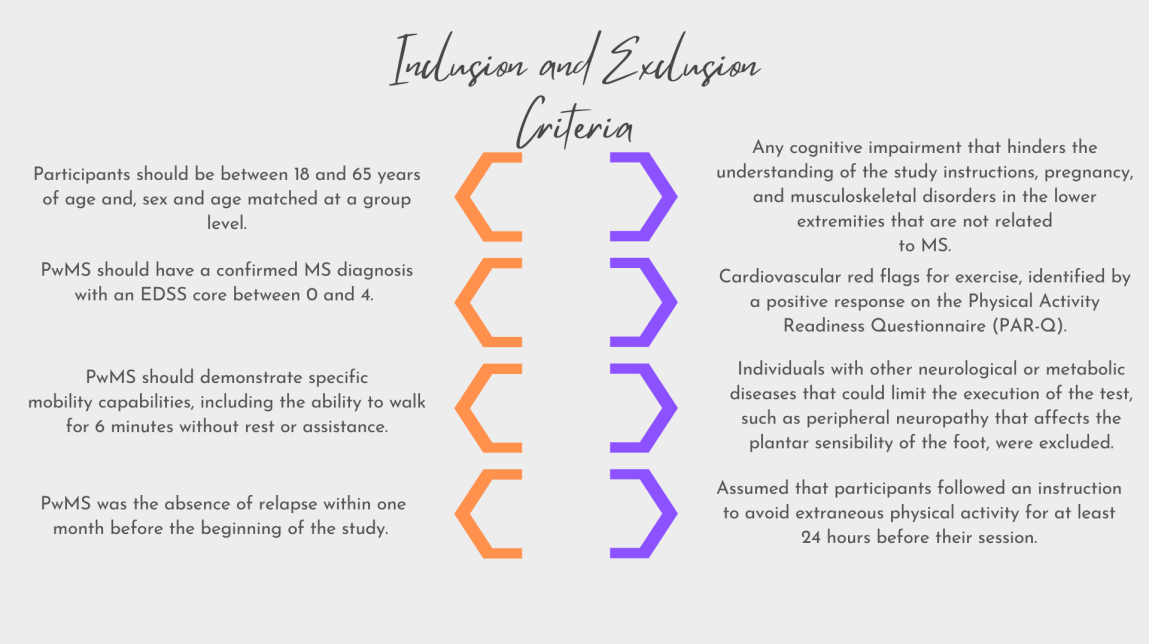
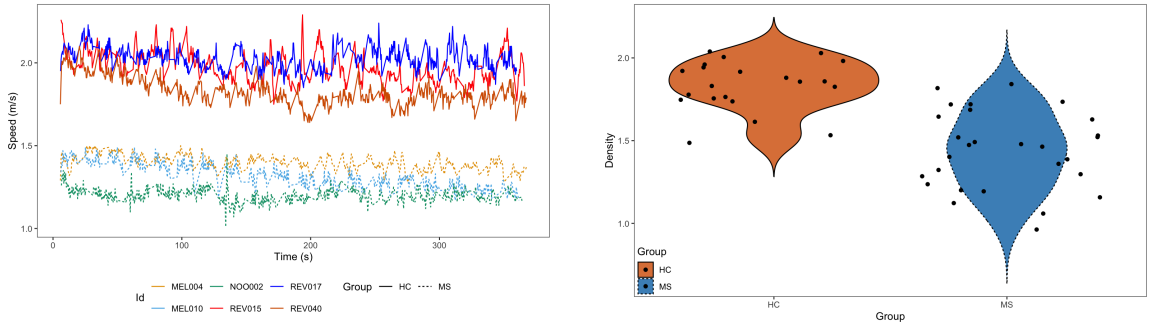


Figure 2: Participant inclusion/exclusion criteria.
EDSS refers to the Expanded Disability Status Scale [27, 28].

3.3 Data Understanding and Exploration

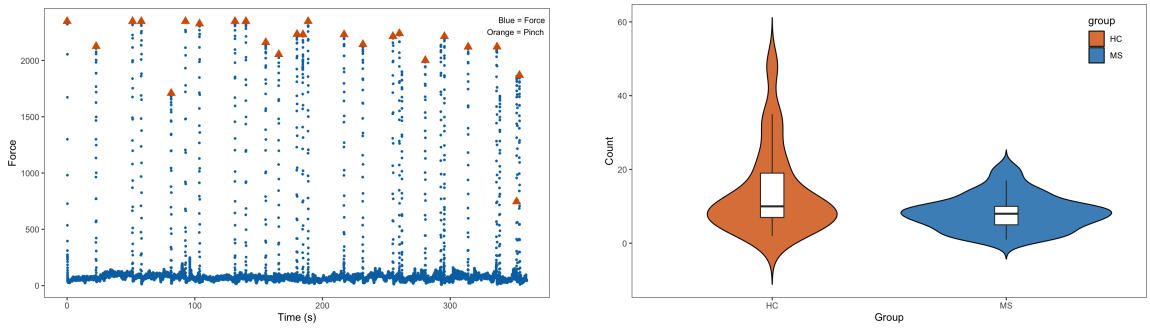
The initial phase of the analysis involved a thorough review of all datasets to understand their structure, distribution, and potential quality issues. The main datasets explored included 6-minutes walking speed data from wearable sensors, timestamped pinch events representing points of perceived changes in walking speed, questionnaire responses, and demographic/contextual variables. This process involved plotting to visualize trends and detect any irregularities and examining the summary statistics to gain an initial understanding of variability in the data.

Figure 3a illustrates consistent group-level differences in walking speed patterns. PwMS participants generally show lower walking speeds compared to healthy controls. Figure 3b as Figure 3a, shows that on average, the MS group walked at lower speeds with greater variability, while the HC group demonstrated more compact and higher speed values. This suggests potential mobility limitations or pacing differences in PwMS during the task. The more variability in PwMS may reflect differences in motor control, fatigue onset, or perception accuracy during the 6-minute walking task.



(a) *Randomly Selected Speed Profiles.* (b) *Distribution of Mean Speed.*

Figure 3: Sample walking speed profiles and distribution of mean speed.
Solid lines represent healthy controls, while dashed lines represent PwMS. The violin plot compares the distribution of average walking speeds between PwMS and HCs.



(a) *Force and Pinch Profile for MEL013.* (b) *Distribution of pinch count by group.*

Figure 4: Sample pinch profile and distribution of pinch counts.
Force output from the pinch sensor (blue dots) and detected pinch events (orange triangles) for a single PwMS participant (MEL008) during the 6-minute walking task. The violin plot compares the distribution of pinch counts between PwMS and HCs.

Figure 4a shows discrete peaks corresponding to self-reported perceived changes in walking speed. The first pinch does not count as it is part of the protocol, it indicates the starting point of the experiment. Figure 4b shows the distribution of the number of pinches from the two groups. The plot suggests high variability in HCs compared to PwMS.

Exploration of demographics data revealed key distinctions between PwMS and HCs across several demographic and health indicators. While age distributions are comparable in terms of variability, the MS group exhibits a slightly higher median BMI, with a lot of participants falling into the overweight and obese categories. Education levels in the MS group are more broadly distributed and peak earlier (around 16 years) compared to the HC group, which

concentrates around 19 years. A notable difference is the sex distribution, with the MS group showing a strong predominance of females. This is expected as several studies have highlighted the prevalence of autoimmune disorders in women [29, 30]. Furthermore, the MS group has a considerably higher proportion of individuals who do not work compared to the HC group. These findings suggest distinct health and socio-economic profiles between the two groups. This section also highlights the key points that will be considered in the statistical analysis.

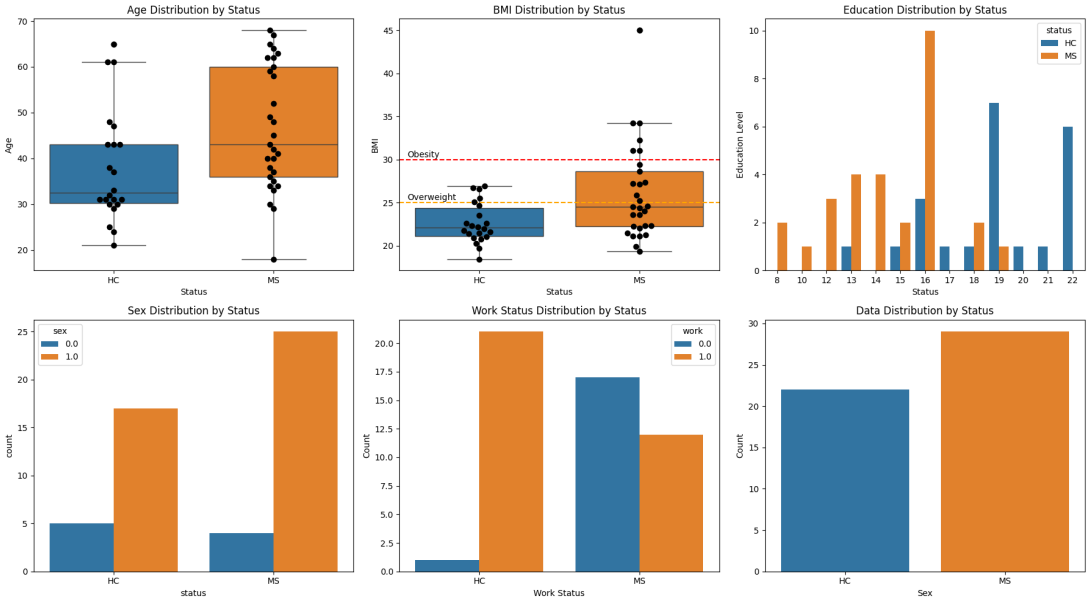
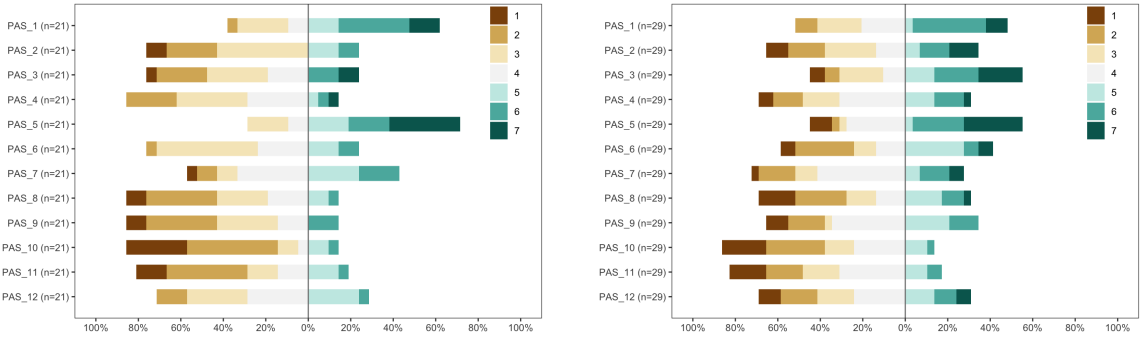


Figure 5: Participants' Demographic Data Exploration.



(a) *PAS responses for HCs*

(b) *PAS responses for PwMS*

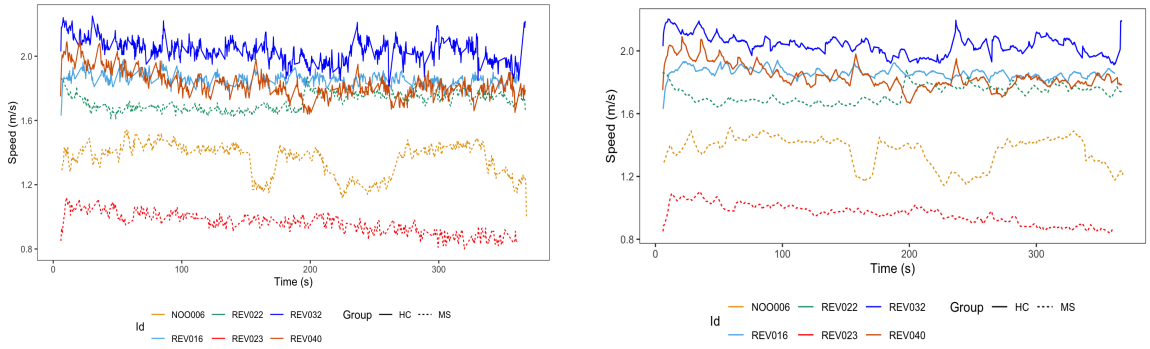
Figure 6: Distribution of the 12-item PAS questionnaire.
Comparing responses between HCs and PwMS. Each item (PAS_1 to PAS_12) shows the percentage distribution of responses across a 7-point Likert scale. The "n" value indicates the sample size for each group (HCs: n=21, PwMS: n=29).

HCs predominantly reported lower PAS scores, reflecting that they do not need to consciously monitor their posture in daily life. In contrast, PwMS showed higher and more variable scores, indicating that they must pay more conscious attention to their bodies to maintain posture. This distinction suggests that postural awareness is more automatic in HCs, whereas it requires active monitoring in PwMS.

3.4 Data Cleaning and Preprocessing

Before conducting statistical analyses, all datasets underwent a series of preprocessing steps to ensure quality, consistency, and compatibility across data sources.

Considering the importance and benefits of signal smoothing, two approaches; the rolling mean and median smoothing, were applied to the speed signal using various window sizes to assess their effectiveness. The rolling mean smoothing technique was used for smoothing the speed signal, and as seen in Figure 7b, it effectively filters out the rapid variations present in the raw speed data. This makes it easier to discern the overall trend and sustained average walking speed for each individual over the duration of the experiment. While the raw data provides instantaneous detail, the smoothed data offers a clearer and more interpretable representation of long-term performance, greatly facilitating visual comparison of the groups and individual trajectories.



(a) *Raw data from the left and right legs*

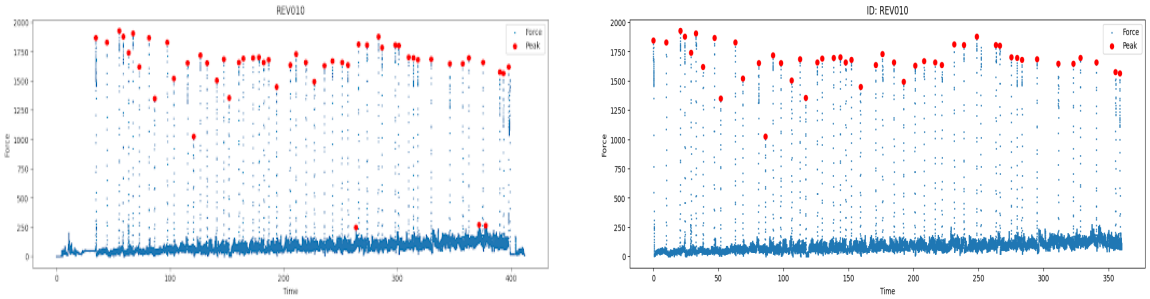
(b) *Smooth signals from both feet*

Figure 7: Sample raw and smoothed speed profiles.
The left (a) plot displays raw walking speed data, while the right (b) plot shows the same data after rolling mean smoothing. HCs (solid lines) consistently exhibit higher walking speeds compared to PwMS (dashed lines), highlighting significant differences in mobility between the groups.

To process the speed change perception data, with expert knowledge guidance, a peak detection algorithm was developed to identify pinch responses from the force signal, with each peak representing a point when a participant reported a perceived change in walking speed. To distinguish intentional pinches from background noise or resting pressure, a

threshold was set at 25% of the participant's maximal hand-grip strength. Additionally, a minimum interval of four seconds between two consecutive pinches was enforced to prevent multiple detections from a single perceived change.

In several cases, participants generated data before initiating the experiment (just holding the pinch sensor) and beyond the intended 360-second duration of the walking task. To address this, a trimming algorithm was applied: the first valid pinch was identified, and only data within the following 360 seconds were retained. Any data recorded before the first pinch or after the 360-second window was discarded.



(a) *Raw force data from the pinch sensor.* (b) *Preprocessed force signal.*

Figure 8: Sample raw and cleaned pinch profile.
Before (a) and after (b) preprocessing. The preprocessing pipeline excluded data outside the valid task window (360 seconds following the first pinch). The cleaned signal retains meaningful peaks representing points of perceived speed variation.

The final output consisted of time-stamped pinch events that were synchronized with walking speed data for subsequent analysis of perception accuracy. Manual inspection was also conducted for participants with unusually low or high numbers of pinch responses to ensure data validity.

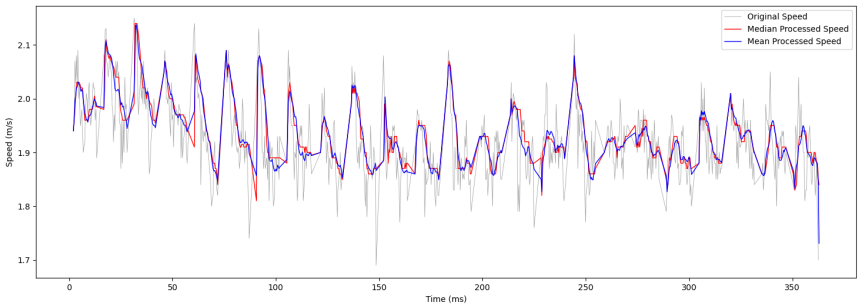
The questionnaire responses also needed to be vetted for clarity and completeness. Responses from the MAIA, ISAQ, and PAS questionnaires were screened for missing values and scoring inconsistencies. Each item was coded as an ordered factor with the appropriate Likert scale range. Total and subscale scores computed per the official scoring instructions for each instrument were included in the datasets. There were no missing values.

As with the other datasets, demographic variables (age, group, work status, BMI, etc) were assessed, validated and encoded as factors and numeric variables as appropriate. Fatigue scores (collected pre- and post-task) and confidence ratings (collected post-task) were screened for completeness.

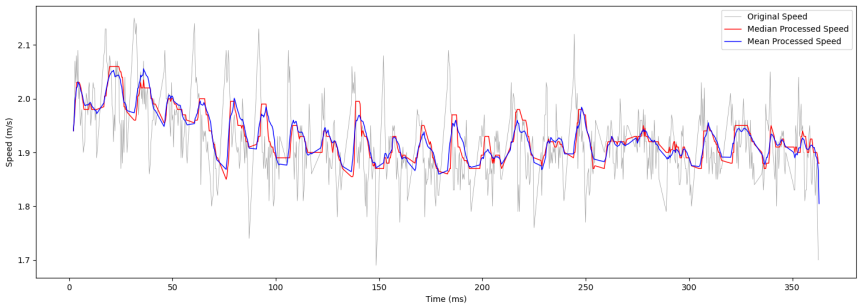
3.5 Feature Engineering: Speed Change Definition

The core feature required to evaluate accuracy of perception was the identification of objective change points in the walking speed signal during the 6MWT. Statistical, derivative-based and threshold-based methods (Figure 10) were explored in order to identify *speed change events* that could be compared against participant-reported pinch responses.

These methods were applied on the 4-second rolling mean smoothed speed signal. After exploring 4-second and 8-second windows, the latter overly smoothens the signal, removing meaningful variation. Also, while median smoothing is robust to outliers, it produces a step-like pattern that is not representative of the true signal dynamics, especially in data with gradual changes.



(a) *Smoothed signal using a 4-second window.*



(b) *Smoothed signal using a 8-second window.*

Figure 9: Mean and median smoothing.
Comparison of smoothing effects using 4-second and 8-second windows with both mean and median approaches. The mean smoothed signal in blue, the median in red and the original signal in gray.

Rolling Slope with Linear Regression

Piecewise linear regression was applied on different sliding windows across the time series speed data and estimated the slope of the local trend. Prominent speed change points were based predefined thresholds (e.g. outside of the interquartile range (IQR), one or

two standard deviations away from the mean slope). Various window sizes were tested to balance sensitivity to gradual versus sharp changes.

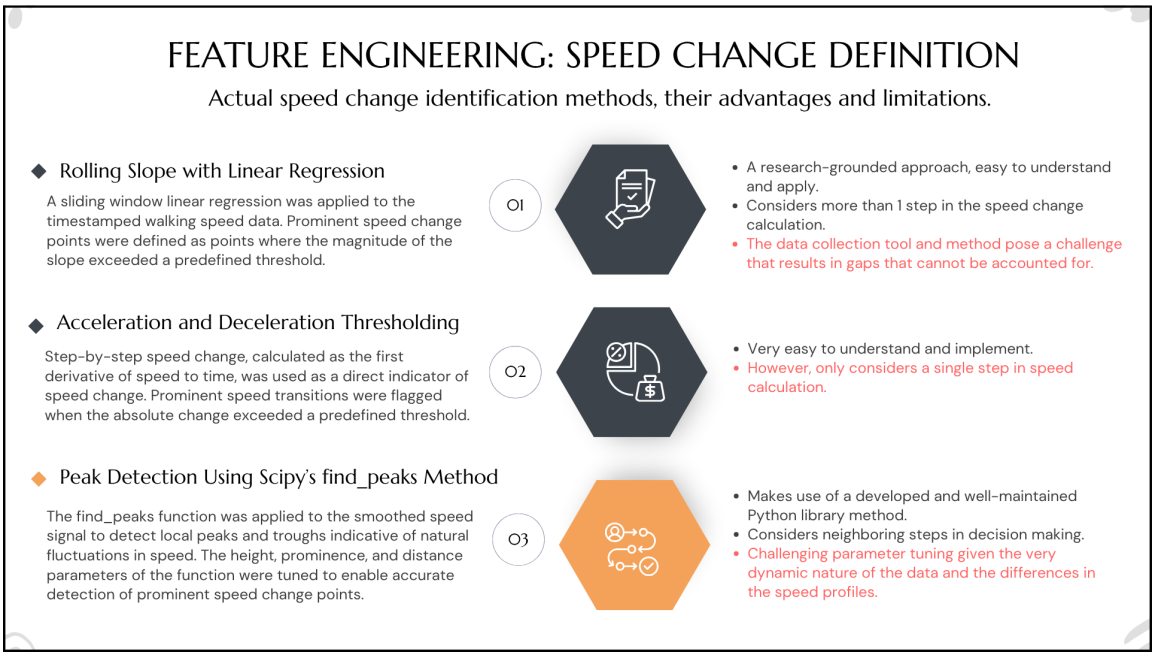


Figure 10: Objective speed change point definition methods.

Discontinuities were observed in the generated rolling slope signal. These discontinuities are not due to data loss, but rather a result of the data collection protocol and tool design. The experiment was conducted in a 30-meter corridor requiring participants to turn at each end. To preserve the integrity of straight-line walking analysis (one of the requirements for the APDM sensor used for data collection), the data collection tool automatically removed three steps before and after each turning point. This truncation helps mimic straight line movement by excluding turns but creates irregular time gaps in the speed data, sometimes exceeding 4s (the sliding window value) between consecutive data points. (See Appendix Figure 18d)

Acceleration and Deceleration Thresholding

Step-by-step acceleration and deceleration, calculated as the first derivative of speed with respect to time, were used as direct indicators of the pace change. Speed transitions were flagged when the absolute acceleration exceeded a predefined threshold. Although this method effectively captured more abrupt changes in walking speed, it did not account for the influence of neighboring points or gradual trends in the signal. Alternative thresholding methods were also explored; each method offers different sensitivity to changes in walking pace. (See Appendix Figure 17)

Peak Detection Using Scipy's *find_peaks* Method

The *find_peaks* function from SciPy, a well-established and used Python library, was applied to the smoothed speed signal to detect local peaks and troughs indicative of natural fluctuations in speed. Parameters of the function such as height, prominence and distance between peaks were tuned to control sensitivity and prevent over-detection [31].

Given a 1D signal,

$$x[0], x[1], \ldots, x[n],$$

With the default *find_peaks* parameters (height=0, distance=1, prominence=0), a sample at index i is considered a peak if:

$$x[i-1] < x[i] \quad \text{and} \quad x[i+1] < x[i].$$

That is, the first derivative changes sign from positive to negative at i. This corresponds to:

$$\Delta x_{i-1} > 0 \quad \text{and} \quad \Delta x_i < 0, \quad \text{where} \quad \Delta x_k = x[k+1] - x[k].$$

With a given height $= h$, the condition becomes:

$$\begin{cases} x[i-1] < x[i] \\ x[i+1] < x[i] \\ x[i] \geq h \end{cases}$$

Given a distance parameter d (# of data points), if two or more peaks are found within distance samples of each other, only the highest is kept. Prominence measures how much a peak stands out due to its height and location. Given a prominence p, p_i of a peak at i is:

$$p_i = x[i] - \max(l, r)$$

where:

$$l = \text{highest "valley" on the left before a higher peak}$$

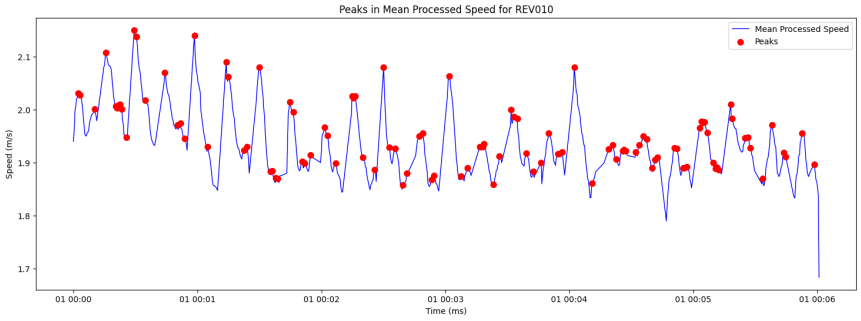
$$r = \text{highest "valley" on the right before a higher peak.}$$

The prominence condition is:

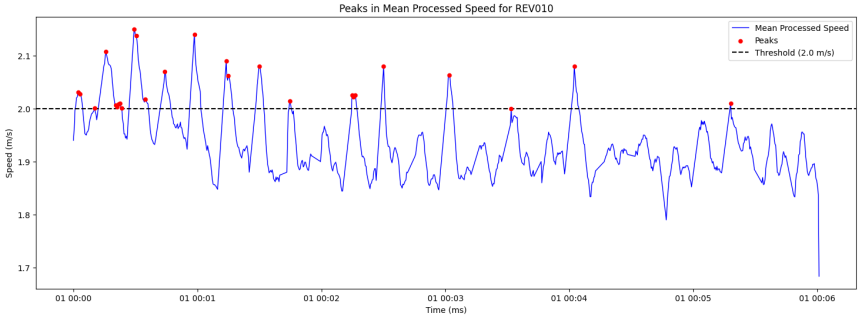
$$p \geq p_i$$

The prominence parameter is used to control the sensitivity of detection by measuring how much a peak stands out relative to its surrounding values. The distance parameter ensures that the peaks are sufficiently separated in time, preventing the identification of multiple peaks within very short intervals because of minor fluctuations.

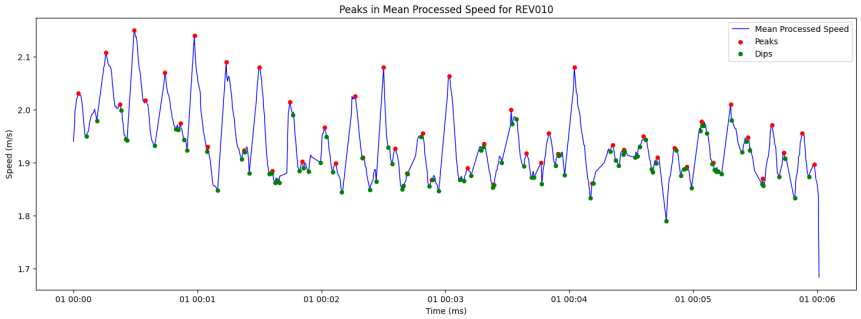
This method allowed for efficient, automated identification of speed change points, both peaks and troughs, especially those characterized by sharp transitions.



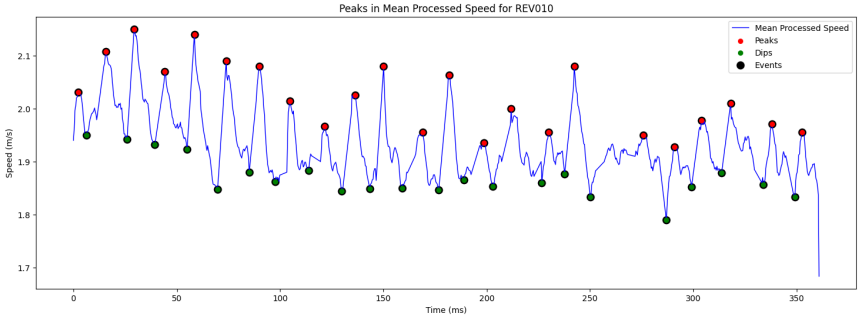
(a) *Using default values of the aforementioned parameters (0, 0, 1)*



(b) *By changing the height alone, we end up with fewer peaks (2, 0, 1).*



(c) *Identify the troughs by passing the negative of the signal to the function.*



(d) *Peaks and troughs with fine tuned parameters (0, 0.0497, 5).*

Figure 11: Overview of signal processing steps using Scipy's find_peaks function. *Various parameter values (height, prominence, distance) were explored to optimize detection.*

The function was applied to the speed signal under different parameter settings (Figure 11 above). Initially, the default parameters were used (11a), which resulted in a large number of detected peaks, many of which reflected minor fluctuations rather than meaningful transitions. In 11b, a height threshold of 2 was introduced to reduce false positives. While this reduced the total number of peaks to 26, some prominent changes were missed, and a few detected peaks occurred too close together, suggesting that a height-only constraint was insufficient to separate distinct events. 11c demonstrates the application of *find_peaks* to the negative speed signal to detect troughs, allowing for the capture of deceleration events.

Finally, in Figure 11d, a combination of tuned parameters used to achieve a more reliable result. This configuration produced a set of peaks that investigators could visually confirm as meaningful, aligning with observable transitions in walking speed and improving consistency with the perceptual data reported by participants. The prominence value of 0.0497m/s [32] is informed by findings from related studies; empirically grounded and physiologically meaningful for detecting meaningful changes in walking pace.

3.6 Feature Engineering: Accuracy Quantification

At this point, perceived speed changes have been defined based on cleaned pinch sensor data that exceed a set threshold, and objective speed changes have also been formally identified. However, this step only sets the foundation, the primary aim of this project is not to define events, but to assess perception accuracy by comparing these subjective responses to actual speed variations

To assess how accurately participants perceived changes in walking speed, a framework was developed to compare objectively detected speed change events against subjective pinch responses recorded during the 6MWT. The following steps were taken to define and quantify perception accuracy:

1. Temporal alignment of events and pinches
2. Numeric outcome definition for the correct and non-correct responses
3. Translation of the correct/non-correct responses to a binary outcome

Temporal Alignment of Events and Pinches

All pinch responses and detected speed change events were time-aligned based on a shared time axis. A tolerance window of ±2 seconds (an expert-informed decision) was applied to determine correspondence: a pinch was considered a true positive (TP) if it occurred within 2 seconds before or after a detected speed change event. Pinches falling outside this window were labeled as false positives (FP), while detected speed changes without a corresponding pinch were labeled as false negatives (FN).

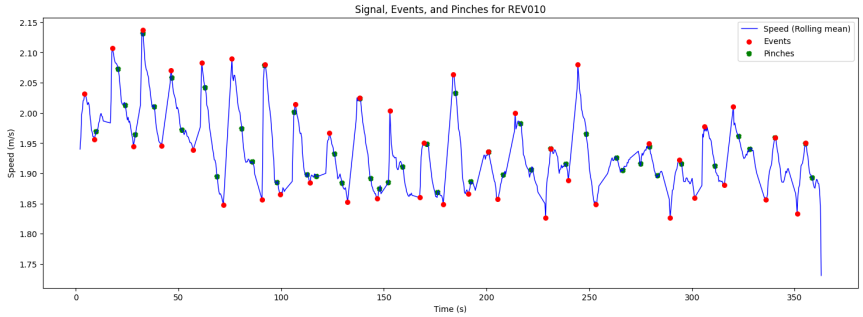


Figure 12: Alignment of Perceived and Objective Speed Changes.
The blue line represents the smoothed walking speed signal. Green points indicate pinch responses (perceived speed changes), while red points mark detected speed change events based on signal analysis. The plot visualizes the temporal correspondence between subjective perception and objectively defined events.

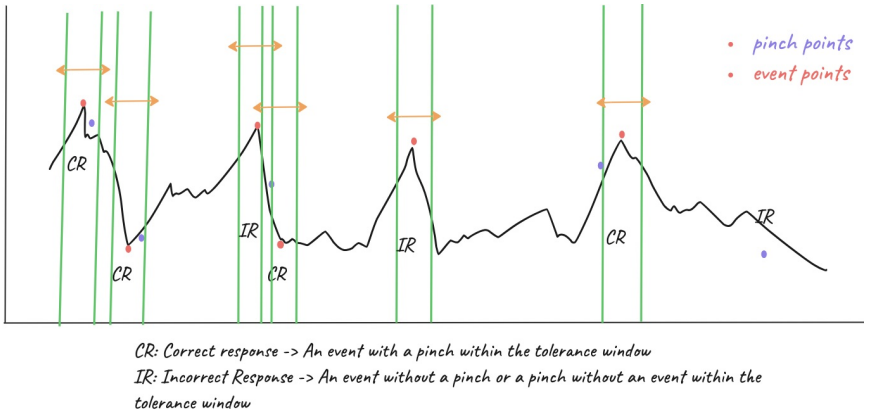


Figure 13: Screenshot Pinch-Event Alignment and Outcome Definition.
This snapshot captures a key discussion phase where criteria for aligning perceived (pinch) and objective (event) speed changes were defined. It illustrates the reasoning behind the binary outcome classification used for modeling perception accuracy.

Binary Outcome

To evaluate perception accuracy, a binary outcome variable was constructed by aligning participant-reported pinch events with objectively detected speed change events. Given TP, FP and FN defined above, the following binary outcome was derived: The TPs corresponding matched pinch-event pairs were labeled as Correct (1). Then the sum of FP + FN corresponding to unmatched pinches (false perceptions) and unmatched speed changes (missed detections) were labeled as Incorrect (0).

To illustrate the construction of the binary outcome variable, consider a participant with 48 pinches, 52 detected speed change events, and 22 matched pairs. Each matched pinch-event

pair was labeled as Correct (1), while unmatched pinches (false perception) and unmatched speed changes (missed speed change) were labeled as Incorrect (0). This resulted in 78 binary outcome entries for the example: 22 correct and 56 incorrect (26 unmatched pinches and 30 unmatched speed changes).

This marks the end of the data processing and feature engineering phase of the project. The subsequent sections detail the statistical methods used to analyze the data.

3.7 Statistical Methods / Modeling Approach

This section outlines the statistical methods and models used to assess and characterize differences between PwMS and healthy controls. GEE and GLMs were used for the binary outcome from the sensor data, and LMMs were used for the questionnaire data analyses.

3.7.1 Generalized Estimating Equations - Binary Outcome

GEE, proposed by Liang and Zeger (1986), extends Generalized Linear Models (GLMs) to handle correlated data structures such as repeated measures and clustered observations. It is a quasi-likelihood approach that provides consistent population-averaged regression parameter estimates even when the working correlation matrix is misspecified. Although GEE provides consistent regression estimates even under a misspecified correlation structure (Liang & Zeger, 1986), choosing a reasonable structure can improve efficiency and the reliability of standard error estimates.

In this study, participants contributed multiple binary outcomes (correct or incorrect pinch responses), leading to clustered responses within individuals. A standard GLM would assume independence of observations, which could result in underestimated standard errors. Given the binomial nature of the outcome, GEE was used with a logit link function. This models the log odds of the probability of accurate perception.

Let:

- Y_{ij} be the binary outcome for subject i at measurement j,
- $\mathbf{x}_{ij}$ the covariate vector (e.g., group, age, sex, BMI),
- $\mu_{ij} = \mathbb{E}[Y_{ij}|\mathbf{x}_{ij}]$,
- $g(\cdot)$ the link function (logit for binary outcomes).

The marginal model is:

$$g(\mu_{ij}) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} \tag{1}$$

Using the logit link:

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \mathbf{x}_{ij}^{\top}\boldsymbol{\beta} \tag{2}$$

The estimate $\hat{\boldsymbol{\beta}}$ is obtained by solving the estimating equation:

$$\sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0 \tag{3}$$

where:

- $\mathbf{Y}_i$ is the outcome vector for subject i,
- $\boldsymbol{\mu}_i$ is the expected value vector,
- $\mathbf{D}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^\top}$ an n_i p matrix,
- $\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R}_i(\alpha) \mathbf{A}_i^{1/2}$ is the working covariance matrix,
- $\mathbf{A}_i$ is a diagonal matrix of variances $\mathrm{Var}(Y_{ij})$,
- $\mathbf{R}_i(\alpha)$ is the working correlation matrix.

To ensure valid inference even if $\mathbf{R}_i(\alpha)$ is misspecified, robust (sandwich) standard errors are computed:

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \mathbf{M}^{-1} \left(\sum_{i=1}^{N} \mathbf{D}_i^\top \mathbf{V}_i^{-1} \mathbf{S}_i \mathbf{S}_i^\top \mathbf{V}_i^{-1} \mathbf{D}_i \right) \mathbf{M}^{-1} \tag{4}$$

where $\mathbf{S}_i = \mathbf{Y}_i - \boldsymbol{\mu}_i$ and $\mathbf{M} = \sum_i \mathbf{D}_i^\top \mathbf{V}_i^{-1} \mathbf{D}_i$.

GEE allows for different working correlation structures [33]:

- **Independence**: assumes no correlation between repeated measurements,
- **Exchangeable**: assumes equal correlation,
- **AR(1)**: assumes decaying correlation with time,
- **Unstructured**: estimates all correlations freely.

The correlation structure will be selected based on how close the model-based and the empirical errors are [24]. The choice for exchangeable working correlation in this analysis stemmed from the fact that the data is more of clustered rather than longitudinal [34]. To assess the adequacy of the assumed working correlation structure, both empirical-based and model-based standard errors were compared. This comparison provided insight into how closely the assumed working correlation matched the true correlation structure.

The GEE model was specified as follows:

$$\mathrm{logit}\left(P(Y_{ij} = 1)\right) = \beta_0 + \beta_1 \cdot \mathrm{status}_{ij} + \beta_2 \cdot \mathrm{age}_{ij} + \beta_3 \cdot \mathrm{sex}_{ij} + \beta_4 \cdot \mathrm{bmi}_{ij} \tag{5}$$

where:

- Y_{ij} is the outcome for the j^{th} measurement of participant i
- β_0 is the log odds of an male HC of average age and average BMI,
- status represents the group (e.g., HC vs. PwMS),
- age is the participant's age,
- sex is the participant's gender,
- bmi is the body mass index,
- $\beta_1, \beta_2, \beta_3, \beta_4$ are fixed effect coefficients for the predictors status, age, sex and bmi.

Hypothesis

$H_0 : \beta_1 = 0$ (Group status has no effect on the odds of an accurate outcome)
$H_A : \beta_1 \neq 0$ (Group status significantly affects the odds of an accurate outcome)

3.7.2 Generalized Linear Mixed Models - Binary Outcome

In addition to using GEE, a GLMM was explored. GLMMs model binary outcomes while explicitly incorporating random effects, in this case, a random intercept for each participant. This has several benefits: Unlike GEE (which provides population-averaged effects), GLMMs estimate individual-level (conditional) effects, which can be more relevant in studies interested in within-subject variation. An argument of interest here is the number of adaptive Gauss-Hermite quadrature points used to approximate the likelihood in the presence of random effects. Increasing the number of quadrature points generally improves the accuracy of the likelihood approximation, leading to more precise parameter estimates at the expense of increased computational cost [24].

With a vector of random effects $\mathbf{b}_i$ for cluster i, and all responses Y_{ij} are independent, the density is given by:

$$f(y_{ij} \mid \theta_{ij}, \phi) = \exp\left\{\phi^{-1}\left[y_{ij}\theta_{ij} - \psi(\theta_{ij})\right] + c(y_{ij}, \phi)\right\}.$$

- The canonical parameter θ_{ij} is modelled as

$$\theta_{ij} = \mathbf{x}_{ij}^{\top}\beta + \mathbf{z}_{ij}^{\top}\mathbf{b}_i,$$

  where $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ are the fixed- and random-effects covariates, respectively.

- The random effects are assumed to follow a multivariate normal distribution:

$$\mathbf{b}_i \sim N(\mathbf{0}, D).$$

- Let $f_{ij}(y_{ij} \mid \mathbf{b}_i, \beta, \phi)$ denote the conditional density of Y_{ij} given $\mathbf{b}_i$. The conditional density of $\mathbf{Y}_i$ equals

$$f_i(\mathbf{y}_i \mid \mathbf{b}_i, \beta, \phi) = \prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \mathbf{b}_i, \beta, \phi).$$

The marginal distribution of $\mathbf{Y}_i$ is

$$f_i(\mathbf{y}_i \mid \beta, D, \phi) = \int f_i(\mathbf{y}_i \mid \mathbf{b}_i, \beta, \phi)\, f(\mathbf{b}_i \mid D)\, d\mathbf{b}_i = \int \left[\prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \mathbf{b}_i, \beta, \phi)\right] f(\mathbf{b}_i \mid D)\, d\mathbf{b}_i,$$

where $f(\mathbf{b}_i \mid D)$ is the density of the $N(\mathbf{0}, D)$ distribution.
The likelihood function for β, D, and ϕ is then

$$L(\beta, D, \phi) = \prod_{i=1}^{N} f_i(\mathbf{y}_i \mid \beta, D, \phi) = \prod_{i=1}^{N} \int \left[\prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \mathbf{b}_i, \beta, \phi)\right] f(\mathbf{b}_i \mid D)\, d\mathbf{b}_i.$$

Log-Likelihood

$$\ell(\beta, D, \phi) = \sum_{i=1}^{N} \log \int \left[\prod_{j=1}^{n_i} f_{ij}(y_{ij} \mid \mathbf{b}_i, \beta, \phi)\right] f(\mathbf{b}_i \mid D)\, d\mathbf{b}_i.$$

Score for β:

$$U_\beta = \frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{N} \int \left[\sum_{j=1}^{n_i} \mathbf{x}_{ij}\,(y_{ij} - \mu_{ij}(\mathbf{b}_i))\right] f(\mathbf{b}_i \mid \mathbf{y}_i, \beta, D, \phi)\, d\mathbf{b}_i,$$

where $\mu_{ij}(\mathbf{b}_i) = \mathbb{E}[Y_{ij} \mid \mathbf{b}_i]$ and $f(\mathbf{b}_i \mid \mathbf{y}_i, \beta, D, \phi)$ is the posterior density of $\mathbf{b}_i$ given the data for cluster i. In practice, these integrals have no closed form and are evaluated using numerical integration techniques such as Laplace approximation or adaptive Gaussian quadrature [24].

The GLMM model fitted was specified as follows:

$$\text{logit}\big(P(Y_{ij} = 1)\big) = \beta_0 + \beta_1 \cdot \text{status}_{ij} + \beta_2 \cdot \text{age}_{ij} + \beta_3 \cdot \text{sex}_{ij} + \beta_4 \cdot \text{bmi}_{ij} + b_i \qquad (6)$$

where

- Y_{ij} is the outcome for the j^{th} measurement of participant i,
- β_0 is the fixed intercept,
- $\beta_1, \beta_2, \beta_3, \beta_4$ are fixed effect coefficients for the predictors status, age, sex, and bmi respectively,
- $b_i \sim \mathcal{N}(0, \sigma_b^2)$ is the participant-specific random intercept capturing variability across participants.

3.7.3 Linear Mixed Models (LMMs) on Raw Questionnaire Responses

Although the outcome variables of the three questionnaires are ordinal, each has 5, 6, or 7 response categories. The literature suggests that ordinal scales with 5 or more categories can be treated as approximately continuous without introducing significant bias. Or, as Geoffrey Norman puts it, "parametric methods can be used without fear of coming to the wrong conclusion" [35]. This approximation allows for the use of LMM, which offer better computational stability, especially in the presence of random effects. Given the complexity and convergence issues often associated with ordinal mixed models, using LMMs provides a practical and theoretically acceptable alternative for modeling these data. We are specifically interested in testing whether the participant's **status** (PwMS vs. HCs group) has a significant effect on their ability to correctly perceive speed changes.

Consider the linear mixed model

$$\mathbf{y}_i = X_i\beta + Z_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i,$$

where

$$\mathbf{b}_i \sim N(\mathbf{0}, D), \quad \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 I_{n_i}), \quad \mathbf{b}_i \text{ and } \boldsymbol{\varepsilon}_i \text{ are independent.}$$

Integrating out $\mathbf{b}_i$, the marginal distribution of $\mathbf{y}_i$ is

$$\mathbf{y}_i \sim N\left(X_i\beta,\ V_i\right), \quad V_i = Z_i D Z_i^\top + \sigma^2 I_{n_i}.$$

The marginal likelihood is

$$L(\beta, D, \sigma^2) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{n_i/2}|V_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - X_i\beta)^\top V_i^{-1}(\mathbf{y}_i - X_i\beta)\right\}.$$

The log-likelihood is

$$\ell(\beta, D, \sigma^2) = -\frac{1}{2}\sum_{i=1}^{N}\left[n_i \log(2\pi) + \log|V_i| + (\mathbf{y}_i - X_i\beta)^\top V_i^{-1}(\mathbf{y}_i - X_i\beta)\right].$$

Score for β:

$$U_\beta = \frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{N} X_i^\top V_i^{-1}(\mathbf{y}_i - X_i\beta).$$

Setting $U_\beta = 0$ gives the generalized least squares (GLS) estimator:

$$\hat{\beta} = \left[\sum_{i=1}^{N} X_i^\top V_i^{-1} X_i\right]^{-1}\left[\sum_{i=1}^{N} X_i^\top V_i^{-1}\mathbf{y}_i\right].$$

Let y_{ijk} be the response of the i-th participant in the j-th group to the k-th questionnaire item. The linear mixed model was specified as follows:

$$Y_{ij} = \beta_0 + \beta_1 \text{status}_i + \beta_2 \text{age}_i + \beta_3 \text{bmi}_i + \beta_4 \text{sex}_i + u_i + \varepsilon_{ij} \tag{7}$$

Where:

- Y_{ij} is the response of participant i to question j.
- β_0 is the expected response for a healthy male of average age and average BMI across all questions.
- $\beta_1, \beta_2, \beta_3, \beta_4$ represent the average effect of status, age, BMI, and sex on the response across all questions, accounting only for subject-level differences.
- $u_i \sim \mathcal{N}(0, \sigma_u^2)$ is the random intercept for participant i.
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is the residual error term.

Hypothesis

$$H_0 : \beta_1 = 0 \quad \text{(No effect of participant status on the questionnaire responses)}$$
$$H_A : \beta_1 \neq 0 \quad \text{(The expected response differs between the two status groups)}$$

3.7.4 Linear Mixed Models (LMMs) on Subscale Data

MAIA and ISAQ have predefined subscale or sub-aggregation of the data. MAIA has eight subscales; however, the validity of these subscales have been questioned in research given that questions in different subscales correlate, and as a result, expert-defined summary scores were created to capture the response patterns of the participants. The outcome was modeled as a repeated-measures factor within participants. A random intercept for participants was included to account for within-subject correlation and individual baseline differences. The effect of the subscore caterogies was also explored.

This modeling strategy allows for simultaneous examination of how participant characteristics influence the subscale scores while properly accounting for the dependence of repeated measures within individuals. By combining expert knowledge with mixed modeling, the analysis gains interpretability and understanding patterns in questionnaire data.

Specifically for MAIA, two composite outcomes were constructed for each participant:

- Negative: Sum of responses to items identified as eliciting negative emotions.
- Positive: Sum of responses to items identified as reflecting positive emotions.

On the other hand, ISAQ has it's predefined subscales:

- F1: Sensitivity to neutral body sensations.
- F2: Attention to unpleasant body sensations.

- F3: Difficulty disengaging from unpleasant body sensations.

The model is specified as follows:

$$Y_{ij} = \beta_0 + \beta_1 \text{status}_i + \beta_2 \text{age}_i + \beta_3 \text{sex}_i + \beta_4 \text{bmi}_i + \beta_5 \text{factor}_j + \beta_6 \text{status}_i * \text{factor}_j + u_i + \varepsilon_{ij} \tag{8}$$

where

- Y_{ij} is the score for participant i and factor j,
- factor_j is a categorical fixed effect indicating the subscales,
- $u_i \sim \mathcal{N}(0, \sigma_u^2)$ is the random intercept for participant i,
- $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is the residual error.

3.8 Software and Tools

For data cleaning, exploration and understanding, Python 3.12.0 was used and for visualization, R version 4.4.0 and Python 3.12.0 were used and for fitting the models, packages from both R and Pytohn were explored. For statistical analysis, R was used.

4 Results

This section presents the results of the analysis, beginning with descriptive statistics and visual summaries of the data, followed by signal processing outcomes, accuracy quantification, and statistical modeling results.

4.1 Results: Exploratory Data Analysis

The dataset included a total of 50 participants, comprising 28 PwMS and 22 HCs. Demographic characteristics were broadly comparable across groups. HC participants were generally younger, clustered below BMI 25, and predominantly employed. PwMS tended to be older, had more cases of BMI > 25 and > 30, and a larger proportion are no employed. The predomininace of females is expected for MS patients as biologically the disease has a predilection for the female gender. The similar distribution in HCs is normal since the experiment was gender-matched.

Table 1: Demographic Characteristics of Participants

Variable	**Total (N = 50)**	**PwMS (n = 28)**	**HC (n = 22)**
Age, years	Mean ± SD: 42.86 ± 13.67	Median: 43.0 Wider range Older on average	Median: 32.5 Tighter range Younger on average
Sex, n (%)	Female: 41 (82%) Male: 9 (18%)	Female: 25 (89.3%) Male: 3 (10.7%)	Female: 17 (77.3%) Male: 5 (22.7%)
BMI	Mean ± SD: 24.60 ± 4.74	Greater variability BMI > 25: 13 (46.4%) BMI > 30: 6 (21.4%)	Tighter clustering BMI > 25: 5 (22.7%) BMI > 30: 0 (0%)
Work, n (%)	Employment status	Working: 11 (60.7%) Not working: 7 (39.3%)	Working: 21 (95.5%) Not working: 1 (4.5%)

4.2 Results: Research Question 1

The evaluation of perception accuracy between the two groups yielded the following results.

4.2.1 GEE Results on Binary Outcome

An exchangeable correlation structure was initially considered. However, comparison of the naïve and robust (sandwich) standard errors under this working correlation structure revealed substantial differences, indicating that the correlation structure might be misspecified (Molenberghs & Verbeke, 2005)[24]. As a result, the independence (having rolled out

the AR-1 and unstructured correlation structures because of the nature of the data) structure was explored. Though more conservative, it produced closer naïve and robust SEs, suggesting a better choice for the working correlation. Choosing a simpler correlation structure is preferable when higher-order structures yield unstable or uninterpretable estimates (Molenberghs & Verbeke, 2005) [24]. Based on this rationale, and the improved alignment between naïve and robust standard errors, the independence structure was selected for the final model.

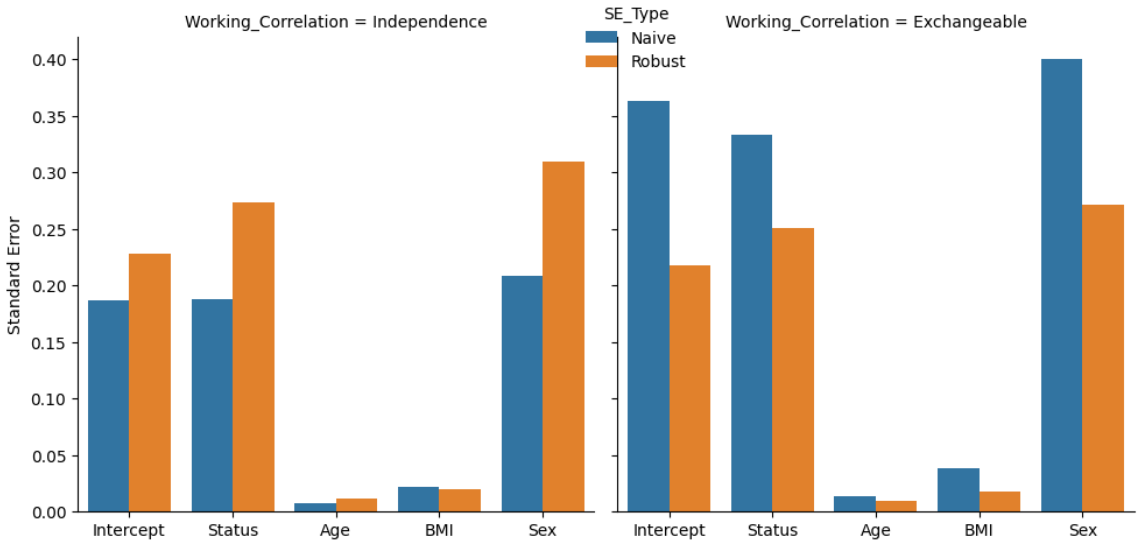


Figure 14: Naive and Robust SEs: Independence and Exchangeable correlation structures. *Notice that with the exchangeable working correlation, the Naive SEs are higher for most variables and the reverse is true for the independence working correlation. The notable SE inflation for Naive estimates under the Exchangeable structure suggests that correlation misspecification can substantially affect standard error magnitude, impacting inference validity.*

Table 2: GEE logistic regression results (Independence structure).

Variable	**Estimate**	**SE (Robust)**	**z**	**p-value**	**OR**	**95% CI**
Intercept	-1.928	0.224	-8.59	< 0.001	0.15	[0.10, 0.22]
Status (MS)	-0.812	0.291	-2.79	0.005	0.44	[0.25, 0.77]
Age (centered)	-0.025	0.012	-1.99	0.047	0.98	[0.95, 1.00]
BMI (centered)	0.008	0.026	0.32	0.747	1.01	[0.96, 1.06]
Sex (Female)	-0.145	0.318	-0.45	0.650	0.87	[0.46, 1.65]

The results indicate that PwMS had significantly lower odds of correctly detecting speed changes compared to healthy controls (OR = 0.44, 95% CI: [0.25, 0.77], p = 0.005). Age

was also a significant predictor, with each additional year associated with a slight decrease in the odds of correct perception (OR = 0.98, 95% CI: [0.95, 1.00], p = 0.047). BMI and sex were not significantly associated with perception accuracy. The intercept (OR = 0.15, 95% CI: [0.10, 0.22], p < 0.001) represents the baseline odds of correct perception for a male healthy control participant of average age and BMI. These findings suggest that both status and age influence perceptual accuracy during the walking task, with PwMS demonstrating measurable deficits compared to controls.

4.2.2 GLMM Results on Binary Outcome

A GLMM with a logit link was fitted to assess whether perception accuracy differed between PwMS and HCs, while accounting for within-subject clustering. The model included a random intercept for each participant and fixed effects for group status, age, sex, and BMI.

An important modeling consideration was the choice of quadrature points in the *glmer()* function, from the *lme4* package in R. The model did not converge with the default setting of 1 point (Laplace approximation), which offers speed at the cost of accuracy. To improve numerical stability and estimation precision, the number of adaptive Gauss-Hermite quadrature points was increased. After testing multiple values, stability in estimates was observed at 12 quadrature points, which was used in the final model. This step was crucial to ensure accurate approximation of the likelihood, particularly for fixed effects, as too few points can lead to biased or unstable estimates due to poor integration over the random effects.

Table 3: Fixed effects estimates from the GLMM Model (Int. Var=0.36)
P-values were computed using the standard normal distribution: $p = 2 \times (1 - \Phi(|z|))$.

Variable	**Estimate**	**SE**	**z**	**p-value**	**OR**	**95% CI**
Intercept	-2.132	0.318	-6.70	< 0.001	0.12	[0.06, 0.20]
Status (MS)	-0.690	0.282	-2.45	0.014	0.50	[0.29, 0.89]
Age (centered)	-0.031	0.012	-2.59	0.009	0.97	[0.95, 0.99]
Sex (Female)	-0.275	0.348	-0.79	0.429	0.76	[0.38, 1.50]
BMI (centered)	0.017	0.032	0.52	0.604	1.02	[0.95, 1.10]

The group effect was statistically significant: PwMS had significantly lower (50% lower) odds of correctly perceiving walking speed changes compared to HCs (β_1 = -0.690, SE = 0.282, p = 0.014). This means that for a specific individual, being in the MS group is associated with a 50% decrease in the odds of the outcome, compared to being in the HC group, adjusting for other covariates. Age was also a significant predictor of perception accuracy (β_2 = -0.031, SE = 0.012, p = 0.0095), suggesting that for a specific person, a one-unit increase in age is associated with a 3% decrease in the odds of the outcome. Sex

and BMI were not significant predictors in the model ($p > 0.4$), implying no detectable influence of these variables on perception accuracy within this sample.

The random intercept variance (0.393) indicates there was meaningful variability in baseline perception accuracy across individuals, justifying the use of a mixed-effects model.
These results reinforce findings from the GEE analysis, confirming that both neurological status and age significantly influence perception accuracy, even after accounting for subject-level variability in baseline performance.

4.3 Results: Research Question 2

The following results were obtained from the LMMs on the raw and aggregated questionnaire outcomes. In addition, a GEE model was used to evaluate the effect of post-experience confidence in pinch accuracy, which helped assess the correspondence between subjective and measured perception accuracy.

4.3.1 LMM Results on Raw Clustered Questionnaire Responses

Table 4: Linear Mixed Model Results for ISAQ Questionnaire (Int. Var = 0.088)

Variable	**Estimate**	**Std. Err.**	**z**	**p-value**	**95% CI**
Intercept	3.298	0.148	22.232	0.000	[3.007, 3.589]
Status (MS)	0.014	0.110	0.126	0.900	[-0.203, 0.230]
Age (centered)	0.001	0.004	0.210	0.834	[-0.007, 0.009]
Sex (Female)	0.137	0.151	0.905	0.366	[-0.159, 0.433]
BMI (centered)	-0.001	0.013	-0.053	0.958	[-0.025, 0.024]

Table 5: Linear Mixed Model Results for MAIA Questionnaire (Int. Var=0.109)

Variable	**Estimate**	**Std. Err.**	**z**	**p-value**	**95% CI**
Intercept	2.987	0.147	20.325	0.000	[2.699, 3.275]
Status (MS)	0.016	0.109	0.147	0.884	[-0.198, 0.230]
Age (centered)	0.004	0.004	0.971	0.332	[-0.004, 0.012]
Sex (Female)	-0.056	0.150	-0.374	0.709	[-0.349, 0.237]
BMI (centered)	0.003	0.012	0.230	0.818	[-0.021, 0.027]

Across all three models, the average predicted response across all questions for a healthy male of average age and BMI is significantly different from zero ($p < 0.001$). However, the

group status showed no significant association with the response (ISAQ: $\beta = 0.014$, p = 0.900, MAIA: $\beta = 0.016$, p = 0.884, PAS: $\beta = 0.002$, p = 0.517). This was the case for the other predictors.

Table 6: Linear Mixed Model Results for PAS Questionnaire (Int. Var=0.005)

Variable	**Estimate**	**Std. Err.**	**z**	**p-value**	**95% CI**
Intercept	0.662	0.032	20.411	0.000	[0.599 0.726]
Status (MS)	0.019	0.024	0.797	0.425	[-0.028 0.067]
Age (centered)	-0.000	0.001	-0.017	0.986	[-0.002 0.002]
Sex (Female)	-0.030	0.033	-0.897	0.370	[-0.094 0.035]
BMI (centered)	-0.001	0.003	-0.347	0.729	[-0.006 0.004]

Collectively, these results indicate that perception of accuracy and interoceptive awareness may not differ substantially between PwMS and healthy individuals, at least as measured by self-report. The consistency of this pattern across three distinct tools suggests a potential disconnect between self-reported interoception and objective performance, as observed in task-based analyses. This underscores the importance of complementing questionnaires with real-time behavioral assessments when evaluating perception accuracy in clinical populations.

4.3.2 LMM Results on Expert Defined Subscaled Data

MAIA and ISAQ as mentioned earlier, have 2 and 3 expert-defined variables, respectively. Below are the results from exploring how group status and other covariates explain the outcome of the defined variables.

Table 7: LMM Results for MAIA Subscale Responses (Int. Var=18.58)

	Estimate	**Std. Err**	**z**	**p_value**	**95% CI**
Intercept	25.544	2.988	8.548	0.000	[19.69, 31.401]
Status (MS)	-0.087	2.580	-0.034	0.973	[-5.142, 4.969]
NP (Positive)	59.333	2.448	24.234	0.000	[54.54, 64.132]
Sex (Female)	-0.966	2.775	-0.348	0.728	[-6.405, 4.474]
Status x NP	0.900	3.192	0.282	0.778	[-5.357, 7.157]
Age (centered)	0.080	0.078	1.019	0.308	[-0.073, 0.233]
BMI (centered)	0.043	0.230	0.188	0.851	[-0.407, 0.493]

In this model, status was not significantly associated with the outcome in PwMS, indicating minimal differences between the MS group and the HCs group after adjusting for other predictors. NP (Positive) was highly significant, but since it is based on many more items (27 vs 10 for NP Negative), its stronger effect may partly reflect the way the scale is constructed rather than a true difference in predictive power. Age, sex, and BMI showed no significant effects. The interaction between Status and NP (Positive) was not significant, indicating that the relationship between NP and the outcome was similar for people with MS and healthy controls.

Similar to the results from the MAIA subscale model, status showed no significant association with the outcome. Only the Sensitivity subscale (F1) was a strong and significant predictor of ISAQ responses. Moreover, the non-significant interaction terms indicate that Sensitivity and Attention did not differ between PwMS and healthy controls.

Table 8: LMM Results for ISAQ Subscale Responses (Int. Var=0.583)

	Estimate	**Std. Err**	**z**	**p_value**	**95% CI**
Intercept	12.576	0.806	15.603	0.000	[10.99, 14.16]
Status (MS)	-0.331	1.037	-0.319	0.749	[-2.365, 1.702]
Sex (Female)	-0.775	0.856	-0.905	0.366	[-2.453, 0.904]
Age (Centered)	0.005	0.024	0.210	0.834	[-0.042, 0.053]
BMI (Centered)	-0.004	0.071	-0.053	0.958	[-0.143, 0.136]
F1 (Sensitivity)	18.857	1.091	17.282	0.000	[16.72, 20.996]
F2 (Attention)	1.810	1.091	1.658	0.097	[-0.329, 3.948]
Status × F1	0.660	1.433	0.461	0.645	[-2.148, 3.468]
Status × F2	0.570	1.433	0.398	0.691	[-2.238, 3.378]

4.3.3 Confidence Ratings and Perception Accuracy Analysis Results

Confidence in pinch accuracy was included as a predictor to examine whether individuals' perceived certainty aligned with their actual perception accuracy, especially relevant in the MS context, where impaired self-awareness can affect mobility decisions. Interestingly, boxplots showed that PwMS reported higher confidence levels on average than healthy controls. For this (sensitivity) analysis, only the GEE model was explored.

The GEE results indicated that confidence was not a significant predictor of accuracy (Estimate = -0.009, p = 0.125), and the interaction with group status also did not reach significance (p = 0.154). This disconnect suggests that despite feeling more confident, PwMS were not necessarily more accurate, highlighting a potential mismatch between subjective

and objective performance.

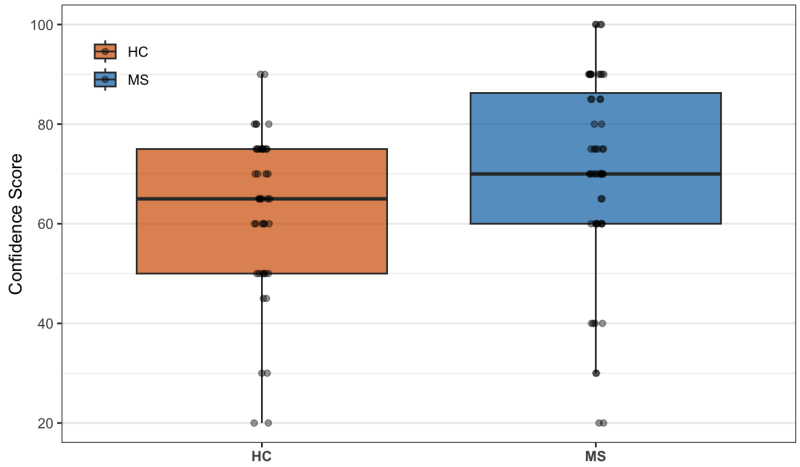


Figure 15: Boxplot comparing self-reported confidence levels between PwMS and HCs. *PwMS reported higher confidence on average, despite not demonstrating significantly greater perception accuracy in the objective task. This suggests a potential mismatch between subjective confidence and actual performance.*

Table 9: Self-reported confidence post protocol impact on perception accuracy

Variable	**Estimate**	**NSE**	**Nz**	**RSE**	**Rz**	**p-value**	**95% CI**
Intercept	-1.705	0.243	-7.025	0.300	-5.687	< 0.001	[-2.292, -1.118]
StatusMS	-1.083	0.309	-3.509	0.326	-3.326	0.0009	[-1.721, -0.445]
Confidence	-0.009	0.006	-1.479	0.006	-1.534	0.1250	[-0.021, 0.003]
Age (Centered)	-0.022	0.008	-2.810	0.012	-1.780	0.0750	[-0.047, 0.003]
BMI (Centered)	0.023	0.022	1.047	0.021	1.101	0.2710	[-0.018, 0.065]
Sex (Female)	-0.120	0.211	-0.571	0.318	-0.379	0.7040	[-0.942, 0.702]
StatusMS:Conf	0.014	0.011	1.264	0.010	1.425	0.1540	[-0.005, 0.033]

5 Discussion and Conclusion

This study aimed to explore the accuracy of perception of speed variation during a 6MWT in PwMS compared to HCs, and compare the correspondence between objective measures of walking speed changes from the 6MWT and their subjective perceptions reported via questionnaires. The goal was to move beyond general self-reported measures of perception and proprioception, and instead quantify perception accuracy using objective, time-aligned events during movement, a task reflective of the daily experiences of PwMS. The findings provide key insights into the complex interplay between subjective awareness and actual motor behavior in PwMS.

Summary of Findings

Using GEE and GLMM, perception accuracy was modeled as a function of group (PwMS vs. HC) and relevant covariates. The results indicated a statistically significant difference between the two groups, with PwMS showing reduced accuracy in detecting walking speed changes. This supports earlier research suggesting impaired proprioception and sensorimotor integration in PwMS, but extends it by offering a task-specific framework to measure these deficits. The use of GEE and GLMM was appropriate given the repeated nature of the binary outcome and the need to model within-subject correlation.

When questionnaire-based scores (MAIA, PAS, ISAQ) were analyzed separately using LMMs, no significant group differences emerged. Interestingly, when exploring confidence in pinch accuracy during the 6MWT, the subjective post-protocol measure did not have a significant effect on the perception accuracy. And the interaction between this measurement and the status was not significant. This further reinforces the limitations of relying solely on subjective self-report instruments for assessing perceptual accuracy. Also, This suggests that while these instruments may be useful for capturing general interoceptive and sensorimotor tendencies, they may not reflect the moment-to-moment accuracy of perception during real-world tasks.

Relationship between GEE and GLMM

GEE provides a population-average or marginal interpretation, meaning its parameters represent the average effect of a predictor across the entire population. It focuses on the big picture, treating the correlation as a nuisance to be adjusted for. In contrast, GLMM provides a subject-specific or conditional interpretation. Its parameters represent the effect of a predictor for a specific individual, explicitly accounting for individual-level variation through the inclusion of random effects.

The parameters from these two models are not directly comparable, especially in non-linear models (like logistic regression). Under the assumption of normality of the random effects, the marginal parameters from a GEE are typically "shrunken" or closer to zero compared to

the conditional parameters from a GLMM. This is because the GEE averages out individual differences, while the GLMM isolates the effect within each subject. The link between the two estimates is approximated as follows for logistic regression;

$$\beta_{marginal} \approx \left[\frac{16\sqrt{3}}{15\pi}V + 1\right]^{-1/2} \beta_{conditional}, \text{ V is the variance of the random intercept}$$

In this case, the GEE estimate for Status (MS) (-0.812) was more extreme than the GLMM estimate (-0.690), highlighting that the data may not perfectly fit the assumptions of the simple approximation. This doesn't mean either model is wrong, but rather that they might be making different assumptions about the underlying data structure.

Strengths and Limitations

A key strength of this study lies in the development of a signal processing and event-matching pipeline that allowed subjective pinch responses to be aligned with objectively defined speed change events during a 6-minute walk. Multiple methods for detecting speed changes including slope-based thresholding, acceleration thresholds, and peak detection were explored to define meaningful changes in walking speed. Similarly, perceived changes were cleaned using peak detection and threshold-based force filtering. Temporal alignment of the speed change points and pinch points enabled the construction of a binary outcome variable that captured whether a pinch was correctly or incorrectly timed relative to actual speed changes, thus operationalizing perception accuracy.

While this study provides valuable insights into task-specific perception accuracy among PwMS compared to healthy controls, several limitations should be acknowledged. The sample size may restrict how broadly the findings apply, and participants might not fully represent the diverse MS population. The study relies partly on self-reported questionnaires, which can be influenced by bias or varied interpretation. The walking task, although simulating natural movement, was performed in a controlled environment that may not capture real-world complexities. The chosen threshold for correct perception (± 2 seconds) is somewhat arbitrary and might oversimplify actual perceptual sensitivity. The accuracy of sensors used was assumed but not independently verified, possibly affecting data precision. Additionally, factors like disease duration, MS subtype, medication, and cognitive fatigue were not considered, which could have influenced the results.

Future work

Future studies could leverage the already developed speed change and accuracy definitions algorithms to conduct more comprehensive studies on perception accuracy. Future studies should aim to address the subjectivity inherent in the perception reports used in this research. Although validated questionnaires were employed, the reliance on self-reported data

introduces potential biases that may affect accuracy. To improve objectivity, subsequent work could incorporate more direct or physiological measures of perception.

Additionally, the current study utilized a self-paced walking task, which, while reflective of natural movement, may have introduced variability in speed changes that complicate interpretation. Implementing a controlled walking protocol, such as a treadmill with administrator-regulated speed adjustments, could provide a more standardized stimulus. This controlled input would allow for more precise alignment between speed changes and perceptual responses, potentially yielding clearer insights into perception accuracy. Incorporating these methodological refinements will strengthen the robustness of future investigations and enhance the generalizability of findings related to perception accuracy in PwMS.

In this study, a fixed prominence threshold and distance parameter were applied in the peak detection process to define speed change points. While this provides consistency across participants, it may not fully account for the variability inherent in human walking behavior, particularly among PwMS who often present with heterogeneous gait patterns. Future studies could explore more adaptive or individualized approaches to parameter selection such as tailoring prominence and distance thresholds to each participant's baseline gait variability or leveraging data-driven techniques to dynamically adjust these parameters. Such approaches would allow the detection algorithm to better capture subtle and subject-specific changes in walking speed, potentially improving the accuracy and ecological validity of perception assessments.

Conclusion

In conclusion, this study demonstrates a novel, task-specific method for quantifying perception accuracy in PwMS. It combines objective signal processing with robust statistical modeling, offering an improved alternative to existing self-report measures. These findings could inform future perception assessment technique in PwMS. It also has the potential to inform rehabilitation protocols that aim to target energy management during exercise in PwMS.

6 Ethical Thinking, Societal Relevance and Stakeholder Awareness

6.1 Ethical Considerations

The data used in this study is from participants who provided informed consent prior to participation. They were fully informed about the nature and purpose of the study, the procedures involved, and their right to withdraw at any point without any consequences. The study involved a low-risk physical task, a self-paced 6MWT. All procedures were designed to avoid physical or emotional distress, particularly for PwMS, for whom extra care and monitoring were in place during data collection. Also, the data was fully anonymized prior to shairnf via secure channels. Therefore, there was no infringement on participants' privacy.

6.2 Societal Relevance

MS is a chronic, progressive neurological condition that significantly impair quality of life. Fatigue, reduced mobility, and depression are among the most disabling symptoms, directly affecting daily independence and social participation. This project marks the beginning of a series of studies aimed at developing data-driven rehabilitation strategies for people with Multiple Sclerosis (PwMS). By combining quantitative analysis with clinical expertise, our goal is to design interventions that not only address physical limitations but also target the psychological barriers to movement, such as fatigue and kinesiophobia (fear of movement). For example, tailored exercise programs can be guided by data to manage effort and energy levels, keeping PwMS engaged in physical activity without overexertion. Sustaining this engagement is crucial, as regular exercise has been shown to improve both physical function and quality of life in PwMS. With a data-driven approach, rehabilitation can be optimized to meet each patient's needs, enhancing motivation, and safety.

6.3 Stakeholder Awareness

This research directly benefits and aligns with the interests of several key stakeholders. REVAL (Uhasselt), as a leading rehabilitation research group, focuses on translating scientific insights into practical improvements for patient care. The findings of this project provide measurable indicators that can inform ongoing rehabilitation protocols and research within the group. By integrating finding from this study alongside traditional functional assessments, MS Rehabilitation Centers in Belgium can better design therapy interventions that address both physical and psychological needs. Also, the insights gained can empower patients and advocacy groups by encouraging broader discussion about patient-reported outcomes in rehabilitation planning.

References

[1] Zeinab Abdallah, Awatef Mohammed, Salwa Rabee, and Adel Abd Elwahab. Assessment of health related quality of life among patients with multiple sclerosis at minia university hospital. 011(1):48–55. ISSN 2785-9797. doi: 10.21608/msnj.2022.128947.1021. URL https://msnj.journals.ekb.eg/article_227936.html.

[2] Agata Beczek, S. G. Roikjær, C. Simonÿ, E. M. Landt, L. Storr, M. Dahl, and M. Beck. Experiences and challenges of people with multiple sclerosis and low attendance to clinical follow-up: a qualitative study. 25(1):107. ISSN 1471-2377. doi: 10.1186/s12883-025-04106-7. URL https://doi.org/10.1186/s12883-025-04106-7.

[3] Paula C. Salamone, Sol Esteves, Vladimiro J. Sinay, Indira García-Cordero, Sofía Abrevaya, Blas Couto, Federico Adolfi, Miguel Martorell, Agustín Petroni, Adrián Yoris, Kathya Torquati, Florencia Alifano, Agustina Legaz, Fátima P. Cassará, Diana Bruno, Andrew H. Kemp, Eduar Herrera, Adolfo M. García, Agustín Ibáñez, and Lucas Sedeño. Altered neural signatures of interoception in multiple sclerosis. 39 (12):4743–4754. ISSN 1065-9471, 1097-0193. doi: 10.1002/hbm.24319. URL https://onlinelibrary.wiley.com/doi/10.1002/hbm.24319.

[4] Akram Jamali, Ebrahim Sadeghi-Demneh, Niloufar Fereshtenajad, and Susan Hillier. Somatosensory impairment and its association with balance limitation in people with multiple sclerosis. 57:224–229. ISSN 09666362. doi: 10.1016/j.gaitpost.2017.06.020. URL https://linkinghub.elsevier.com/retrieve/pii/S0966636217302436.

[5] Hatice Ayan, Özge Ertekin, Turhan Kahraman, and Serkan Özakbaş. Balance and gait impairment in persons with multiple sclerosis with the absence of clinical disability. 26 (3):224–229. ISSN 1301062X. doi: 10.4274/tnd.2020.36036. URL https://tjn.org.tr/doi/10.4274/tnd.2020.36036.

[6] Dagmara Wasiuk-Zowada, Anna Brzek, Ewa Krzystanek, and Andrzej Knapik. Kinesiophobia in people with multiple sclerosis and its relationship with physical activity, pain and acceptance of disease. 58(3):414. ISSN 1648-9144. doi: 10.3390/medicina58030414. URL https://www.mdpi.com/1648-9144/58/3/414. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[7] Derval McCormack, Dr Fiadhnait O'Keeffe, Christina Seery, and Dr Fiona Eccles. The association between body image and psychological outcomes in multiple sclerosis. a systematic review. 93:106226. ISSN 22110348. doi: 10.1016/j.msard.2024.106226. URL https://linkinghub.elsevier.com/retrieve/pii/S2211034824008022.

[8] Chadia Ed-driouch, Florent Chéneau, Françoise Simon, Guillaume Pasquier, Benoit Combès, Anne Kerbrat, Emmanuelle Le Page, Sophie Limou, Nicolas Vince, David-Axel Laplaud, Franck Mars, Cédric Dumas, Gilles Edan, and Pierre-Antoine Gourraud.

Multiple sclerosis clinical decision support system based on projection to reference datasets. 9(12):1863–1873. ISSN 2328-9503, 2328-9503. doi: 10.1002/acn3.51649. URL https://onlinelibrary.wiley.com/doi/10.1002/acn3.51649.

[9] Björn Zörner, Pascal Hostettler, Christian Meyer, Tim Killeen, Pauline Gut, Michael Linnebank, Michael Weller, Dominik Straumann, and Linard Filli. Prognosis of walking function in multiple sclerosis supported by gait pattern analysis. 63. ISSN 2211-0348, 2211-0356. doi: 10.1016/j.msard.2022.103802. URL https://www.msard-journal.com/article/S2211-0348(22)00314-5/fulltext. Publisher: Elsevier.

[10] Jane Desborough, Crystal Brunoro, Anne Parkinson, Katrina Chisholm, Mark Elisha, Janet Drew, Vanessa Fanning, Christian Lueck, Anne Bruestle, Matthew Cook, Hanna Suominen, Antonio Tricoli, Adam Henschke, and Christine Phillips. 'it struck at the heart of who i thought i was': A meta-synthesis of the qualitative literature examining the experiences of people with multiple sclerosis. 23(5):1007–1027. ISSN 1369-7625. doi: 10.1111/hex.13093.

[11] (PDF) laboratory review: The role of gait analysis in seniors' mobility and fall prevention. . doi: 10.1159/000322194. URL https://www.researchgate.net/publication/47566721_Laboratory_Review_The_Role_of_Gait_Analysis_in_Seniors'_Mobility_and_Fall_Prevention.

[12] Study explores whether MS patients accurately perceive their risk of falling, . URL https://www.yu.edu/news/katz/study-explores-whether-ms-patients-accurately-perceive-their-risk-falling.

[13] Tobia Zanotto, Danya Pradeep Kumar, Daniel Golan, Jeffrey Wilken, Glen M. Doniger, Myassar Zarif, Barbara Bumstead, Marijean Buhse, Joanna Weller, Sarah A. Morrow, Iris-Katharina Penner, Laura Hancock, Thomas J. Covey, Edward Ofori, Daniel S. Peterson, Robert W. Motl, Hans Bogaardt, Marissa Barrera, Riley Bove, Herbert Karpatkin, Jacob J. Sosnoff, and Mark Gudesblatt. Does cognitive performance explain the gap between physiological and perceived fall-risk in people with multiple sclerosis? 95. ISSN 2211-0348, 2211-0356. doi: 10.1016/j.msard.2025.106322. URL https://www.msard-journal.com/article/S2211-0348(25)00065-3/fulltext. Publisher: Elsevier.

[14] Amie Wallman-Jones, Pandelis Perakakis, Manos Tsakiris, and Mirko Schmidt. Physical activity and interoceptive processing: Theoretical considerations for future research. 166:38–49. ISSN 01678760. doi: 10.1016/j.ijpsycho.2021.05.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0167876021001604.

[15] Randolph Blake and Maggie Shiffrar. Perception of human motion. 58(1):47–73. ISSN 0066-4308, 1545-2085. doi: 10.1146/annurev.psych.57.102904.190152. URL https://www.annualreviews.org/doi/10.1146/annurev.psych.57.102904.190152.

[16] Camille J. Shanahan, Frederique M. C. Boonstra, L. Eduardo Cofré Lizama, Myrte Strik, Bradford A. Moffat, Fary Khan, Trevor J. Kilpatrick, Anneke van der Walt, Mary P. Galea, and Scott C. Kolbe. Technologies for advanced gait and balance assessments in people with multiple sclerosis. 8. ISSN 1664-2295. doi: 10.3389/fneur.2017.00708. URL https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2017.00708/full. Publisher: Frontiers.

[17] Andrew S. Monaghan, Jessie M. Huisinga, and Daniel S. Peterson. The application of principal component analysis to characterize gait and its association with falls in multiple sclerosis. 11:12811. ISSN 2045-2322. doi: 10.1038/s41598-021-92353-2. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8211858/.

[18] Douglas A. Wajda, Tobia Zanotto, and Jacob J. Sosnoff. Motor imagery of walking in people living with and without multiple sclerosis: A cross-sectional comparison of mental chronometry. 11(9):1131. ISSN 2076-3425. doi: 10.3390/brainsci11091131. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8466525/.

[19] Victoria MJ Smith, Jonathan S Varsanik, Rachel A Walker, Andrew W Russo, Kevin R Patel, Wendy Gabel, Glenn A Phillips, Zebadiah M Kimmel, and Eric C Klawiter. Movement measurements at home for multiple sclerosis: walking speed measured by a novel ambient measurement system. 4(1):2055217317753465. ISSN 2055-2173. doi: 10.1177/2055217317753465. URL https://doi.org/10.1177/2055217317753465. Publisher: SAGE Publications Ltd.

[20] Farnood Gholami, Daria A. Trojan, Jozsef K ovecses, Wassim M. Haddad, and Behnood Gholami. Gait assessment for multiple sclerosis patients using microsoft kinect. URL http://arxiv.org/abs/1508.02405.

[21] Massimiliano Pau, Silvia Caggiari, Alessandro Mura, Federica Corona, Bruno Leban, Giancarlo Coghe, Lorena Lorefice, Maria Giovanna Marrosu, and Eleonora Cocco. Clinical assessment of gait in individuals with multiple sclerosis using wearable inertial sensors: Comparison with patient-based measure. 10:187–191. ISSN 2211-0348, 2211-0356. doi: 10.1016/j.msard.2016.10.007. URL https://www.msard-journal.com/article/S2211-0348(16)30191-2/abstract. Publisher: Elsevier.

[22] Cyril Voisard, Nicolas de l'Escalopier, Aliénor Vienne-Jumeau, Albane Moreau, Flavien Quijoux, Flavie Bompaire, Magali Sallansonnet, Marie-Laure Brechemier, Irina Taifas, Camille Tafani, Eve Drouard, Nicolas Vayatis, Damien Ricard, and Laurent Oudre. Innovative multidimensional gait evaluation using IMU in multiple sclerosis: introducing the semiogram. 14. ISSN 1664-2295. doi: 10.3389/fneur.2023.1237162. URL https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2023.1237162/full. Publisher: Frontiers.

[23] Jennifer Murphy, Caroline Catmur, and Geoffrey Bird. Classifying individual differences in interoception: Implications for the measurement of interoceptive awareness. 26(5):1467–1471. ISSN 1531-5320. doi: 10.3758/s13423-019-01632-7.

[24] Geert Verbeke Geert Molenberghs. *Models for Discrete Longitudinal Data*. Springer Series in Statistics. Springer-Verlag. ISBN 978-0-387-25144-8. doi: 10.1007/0-387-28980-1. URL http://link.springer.com/10.1007/0-387-28980-1.

[25] Katleen Bogaerts, Marta Walentynowicz, Maaike Van Den Houte, Elena Constantinou, and Omer Van Den Bergh. The interoceptive sensitivity and attention questionnaire: Evaluating aspects of self-reported interoception in patients with persistent somatic symptoms, stress-related syndromes, and healthy controls. 84(2):251–260. ISSN 1534-7796, 0033-3174. doi: 10.1097/psy.0000000000001038. URL https://journals.lww.com/10.1097/PSY.0000000000001038. Publisher: Ovid Technologies (Wolters Kluwer Health).

[26] Holger Cramer, Wolf E. Mehling, Felix J. Saha, Gustav Dobos, and Romy Lauche. Postural awareness and its relation to pain: validation of an innovative instrument measuring awareness of body posture in patients with chronic pain. 19(1):109. ISSN 1471-2474. doi: 10.1186/s12891-018-2031-9. URL https://doi.org/10.1186/s12891-018-2031-9.

[27] M. S. Trust. Expanded disability status scale (EDSS) | MS trust. URL https://mstrust.org.uk/a-z/expanded-disability-status-scale-edss.

[28] The McDonald criteria for MS diagnosis | MS society, . URL https://www.mssociety.org.uk/about-ms/diagnosis/the-tests-for-ms/mcdonald-criteria.

[29] Suhaylah Ingar. Autoimmune disease and its prevalence in women. URL https://blog.mdpi.com/2024/02/05/autoimmune-disease-in-women/.

[30] Fariha Angum, Tahir Khan, Jasndeep Kaler, Lena Siddiqui, and Azhar Hussain. The prevalence of autoimmune disorders in women: A narrative review. 12(5):e8094. ISSN 2168-8184. doi: 10.7759/cureus.8094. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7292717/.

[31] scipy. scipy/scipy/signal/_peak_finding.py at v1.16.1 · scipy/scipy. URL https://github.com/scipy/scipy/blob/v1.16.1/scipy/signal/_peak_finding.py.

[32] Characterizing human perception of speed differences in walking: Insights from a drift diffusion model | eNeuro, . URL https://www.eneuro.org/content/12/5/ENEURO.0343-23.2025.

[33] Stano Pekár and Marek Brabec. Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. 124 (2):86–93. ISSN 0179-1613, 1439-0310. doi: 10.1111/eth.12713. URL https://onlinelibrary.wiley.com/doi/10.1111/eth.12713.

[34] Garrett M. Fitzmaurice, Nan M. Laird, and James H. Ware. *Applied Longitudinal Analysis*. John Wiley & Sons. ISBN 978-0-471-21487-8. Google-Books-ID: gCoTIFejMgYC.

[35] Geoffrey Norman. LIkert scales, levels of measurement adn the "laws" of statistics. 15:625–32. doi: 10.1007/s10459-010-9222-y.

Appendix

Introduction

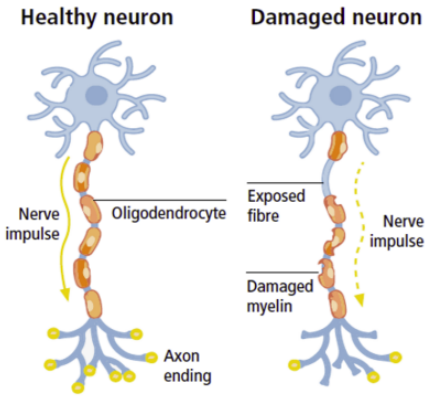
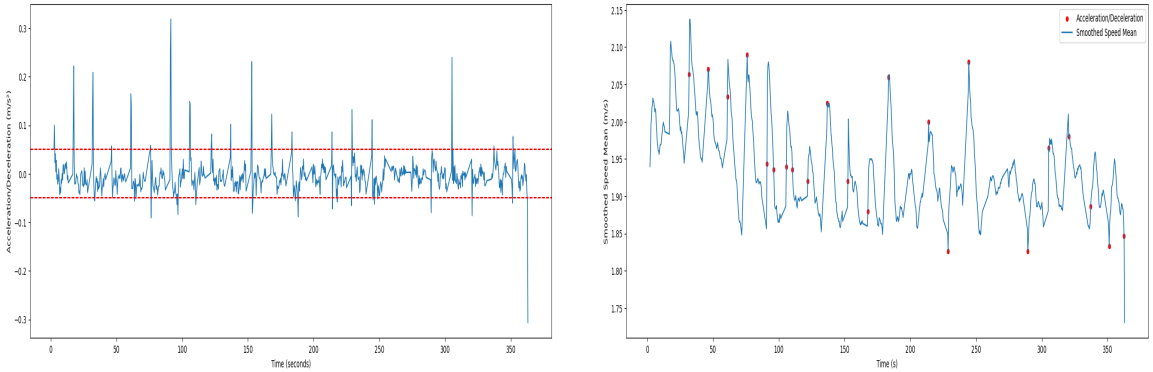


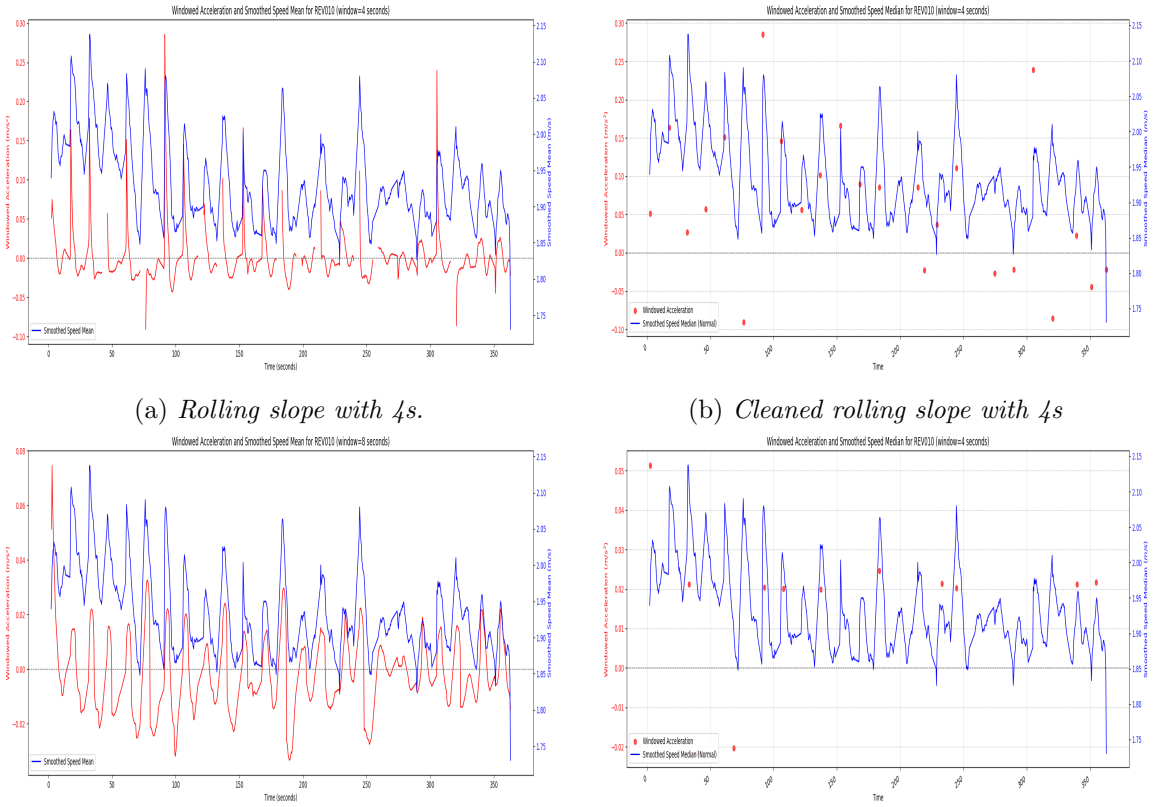
Figure 16: Comparison of a healthy and a damaged neuron

Feature Engineering: Speed Change Definition



(a) *Acceleration profile for one participant* (b) *Acceleration overlaid on the smoothed signal*

Figure 17: Acceleration-based detection of walking speed changes
Different thresholds explored; (1) Fixed threshold (0.0497) in (a), (2) IQR-based outlier detection, (3) 75th percentile, and (4) ±2 standard deviations from the mean. Candidate events were identified as peaks above the respective thresholds.



(a) *Rolling slope with 4s.*

(b) *Cleaned rolling slope with 4s*

(c) *Rolling slope with 8s*

(d) *Cleaned rolling slope with 8s*

Figure 18: Identification of speed change points using rolling slope analysis.
The first row plots shows the smoothed speed profile with overlaid rolling slope estimates using a rolling window of 4-second and the second row correspond to a rolling window of 8-second. Points where the slope exceeded a predefined threshold (IQR) were initially flagged as candidate speed changes. To avoid over-detection, consecutive points occurring within 4 seconds of each other were merged into a single event. The second column shows the final set of detected speed change events aligned with the smoothed speed signal, illustrating clear transitions in walking pace consistent with the algorithm's intent.

Initially, a 4-second window was used. However, due to the irregular sampling created by the trimming, this led to unintended fragmentation of speed change detection in some cases. An 8-second window was then tested to bridge these gaps. While this approach reduced discontinuities, it was found to over-merge distinct speed change events, thereby masking obvious accelerations and decelerations in walking speed that participants likely perceived. Notice the fewer event points detected with the 8-second rolling slope window.

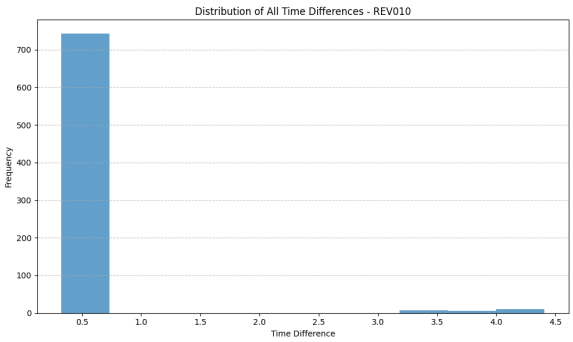


Figure 19: Distribution of Time Intervals Between Successive Data Points.
The histogram illustrates the time differences between consecutive data points in the speed signal, providing insight into the sampling consistency of the data collection process. Interval widths of less than 1 second are expected, however, values beyond 3 as in the plot indicate truncated turning points.

Modeling

Work status was initially considered as a covariate. However, due to its high collinearity with participant group (PwMS vs HC), potential for complete separation, and the lack of clear theoretical justification for a direct effect of employment status on perception accuracy, work status was excluded from the final GEE model.

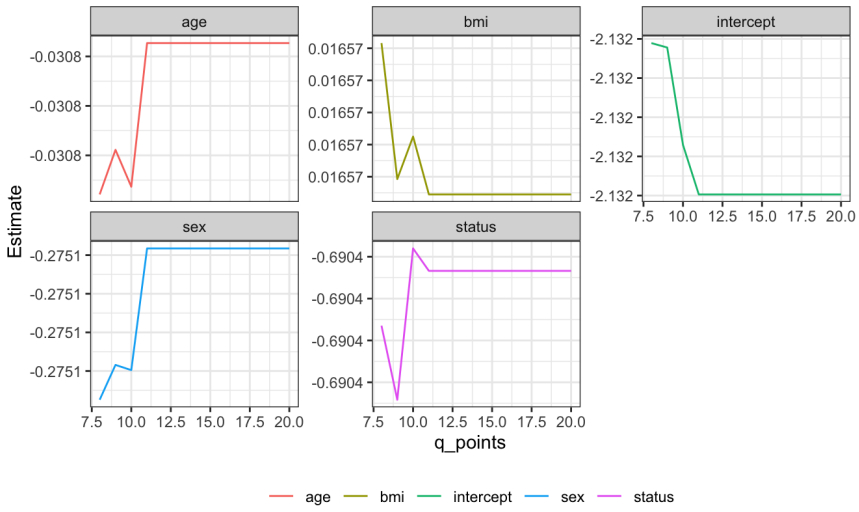


Figure 20: Stability of fixed effect estimates in the GLMM.
With across varying numbers of adaptive Gauss-Hermite quadrature points. Estimates stabilized around 12 points.

Table 10: Comparison of GEE estimates under independence and exchangeable correlation structures

Variable	**Estimate**	**Naïve SE**	**z**	**Robust SE**	**z**	**p (Robust)**
Independence						
Intercept	-1.928	0.187	-10.33	0.224	-8.59	< 0.001
statusMS	-0.812	0.193	-4.22	0.291	-2.79	0.005
age_centered	-0.025	0.008	-3.10	0.012	-1.99	0.047
bmi_centered	0.008	0.023	0.35	0.026	0.32	0.747
sex (Female)	-0.145	0.210	-0.69	0.318	-0.45	0.650
Exchangeable						
Intercept	-2.026	0.385	-5.26	0.216	-9.38	< 0.001
statusMS	-0.704	0.360	-1.95	0.268	-2.63	0.009
age_centered	-0.0259	0.015	-1.71	0.010	-2.56	0.010
bmi_centered	0.0115	0.043	0.27	0.024	0.49	0.624
sex (Female)	-0.2095	0.425	-0.49	0.279	-0.75	0.452

QQ-plots were used to assess the normality of residuals assumption in linear mixed-effects models. Despite slight deviations, particularly in MAIA and ISAQ, linear mixed models are fairly robust to mild non-normality, especially when analyzing repeated measures or aggregated item scores.

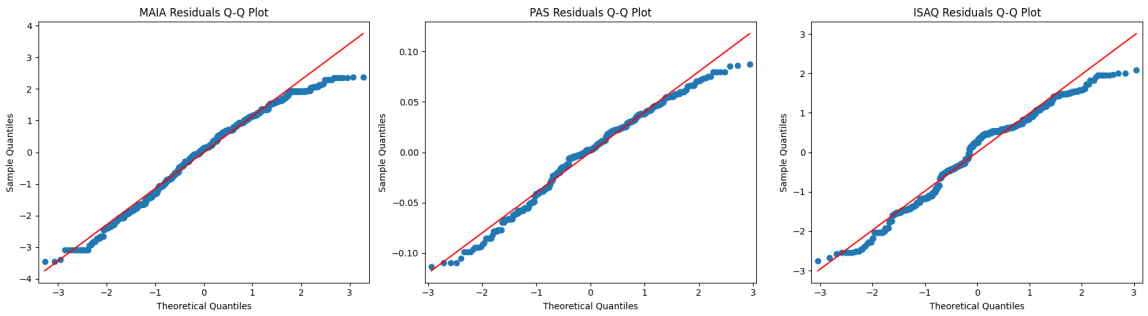


Figure 21: Q-Q plots of residuals

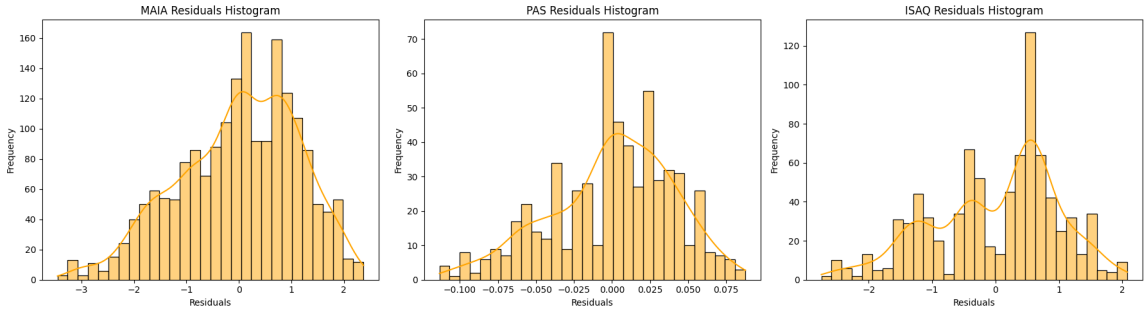


Figure 22: Distribution of Residuals

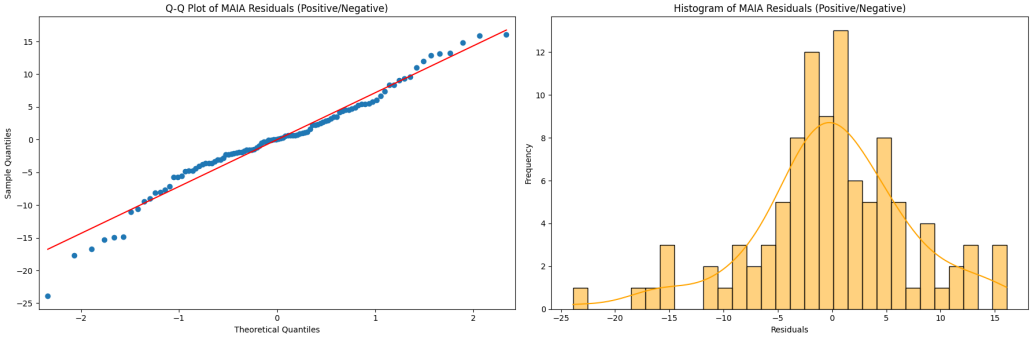


Figure 23: MAIA subscale model residuals

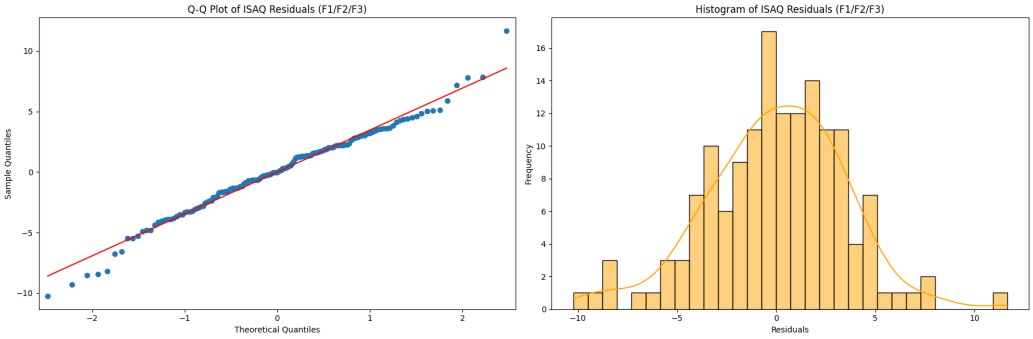


Figure 24: ISAQ subscale model residuals

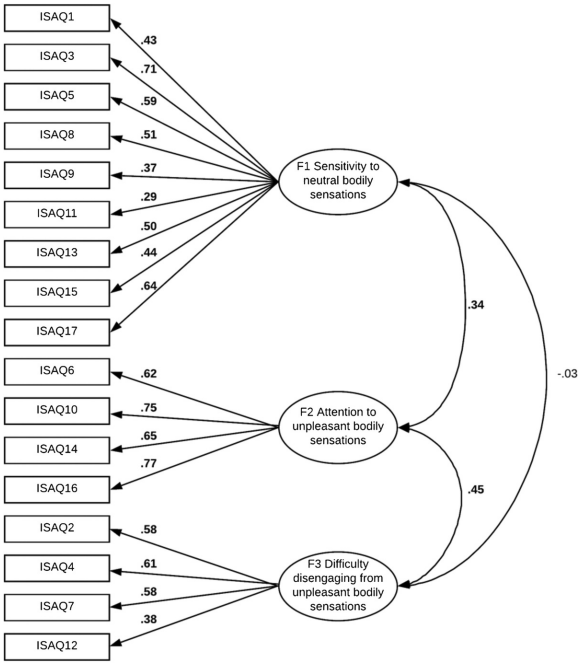


Figure 25: ISAQ Subscale Construct.

Sample code

This section contains a sample of the Python and R code used in the processing and analysis conducted in this project. More can be found on Github.

```python
def find_dips_and_shoots_by_id(df: pd.DataFrame,
                               col_value: str,
                               id_col: str,
                               height=None,
                               prominence: float = 0.0497,
                               distance: int = 400) -> pd.DataFrame:
    """
    Find dips and shoots in a signal for each unique ID in the DataFrame.

    Args:
        df (pd.DataFrame): The input DataFrame containing the signal data.
        col_value (str): The name of the column containing the signal values.
        id_col (str): The name of the column containing the IDs.
        height (float, optional): Required height of peaks. Defaults to None.
        prominence (float, optional): Required prominence of peaks. Defaults
        ↪  to 0.0497.
```

```
        distance (int, optional): Required distance between peaks. Defaults to 10.

    Returns:
        pd.DataFrame: A DataFrame containing the dips and shoots for each ID.
    """
    df_events = []

    for unique_id in df[id_col].unique():
        df_sub = df[df[id_col] == unique_id].reset_index(drop=True)
        signal = df_sub[col_value].values

        dips, _ = fp(-signal, height=height, prominence=prominence, distance=distance)
        shoots, _ = fp(signal, height=height, prominence=prominence, distance=distance)

        events = sorted(set(dips.tolist() + shoots.tolist()))

        df_events_id = pd.DataFrame({
            'id': unique_id,
            'time_modified': df_sub['time_modified'].iloc[events].values,
            'smoothed_speed_mean': df_sub['smoothed_speed_mean'].iloc[events].values
        })

        df_events.append(df_events_id)
        # print(len(df_events_id), "events found for ID:", unique_id)

    return pd.concat(df_events, ignore_index=True)
```

```
gee_model_i = gee(outcome ~ status + age_centered + bmi_centered + sex,
                  id=id, family=binomial, data=df,
                  corstr = "independence", Mv = 1)
gee_model_e = gee(outcome ~ status + age_centered + bmi_centered + sex,
                  id=id, family=binomial, data=df,
                  corstr = "exchangeable", Mv = 1)
```