D6.1 Statistical description report on digital mobility outcomes, health outcomes and their relationships

Mobilise-D

Connecting digital mobility assessment to clinical outcomes for regulatory and clinical endorsement

Grant Agreement No. 820820

[WP6 - Statistical analysis, evaluation of results and data availability]

Lead contributor	Saverio Ranciati (UniBO)
Other contributors	Aida Arselan Aydemir (EMD Serono)
	Judith Garcia–Aymerich (ISGlobal)
	Luca Palmerini (UniBO)
Quality reviewer	Mike Jackson (IXS)

Due date (project month)	M21
Actual delivery date	31/12/2020
Document version	V1.0
Deliverable type	R
Dissemination level	PU

Document History

Version	Date	Description	Contributors
V1.0	30/12/2020	Final version for submission	Saverio Ranciati (UniBO)

List of abbreviation terms

Abbreviation	Definition
DMO	Digital mobility outcome
PD	Parkinson's Disease
MS	Multiple Sclerosis
PFF	Proximal Femur Fracture
COPD	Chronic Obstructive Pulmonary Disease
WB	Walking bout
LLFDI	Late-Life Function and Disability Instrument
CAT	COPD assessment test
EDSS	Extended Disability Status Scale
MSWS-12	Multiple Sclerosis Walking Scale with 12 items
MSIS-29	Multiple Sclerosis Impact Scale with 29 items
UPDRS	Unified Parkinson's Rating Scale
s-FESI	Short form of fall efficacy scale international
H&Y	Hoehn and Yahr scale
CI	Confidence interval
Avg., avg.	Average
SD, sd, std	Standard deviation
Asym., asym.	Asymmetry
Min, min.	Minimum
Max, max.	Maximum
N, n	Refers to sample size and/or number of data points used in computation

Table of Contents

1 Abstract 4
2 Introduction 5
2.1 Objectives of WP6.1 Task and Datasets 5
2.2 Digital Mobility Outcomes (DMOs) and clinical endpoints 8
3 Methods 12
3.1 Marginal distribution of DMO and clinical outcomes (objective 1 and objective 2) 12
3.2 Characteristics of relationships between DMOs (objective 3) 13
3.3 Characteristics of DMOs conditional on disease characteristics or confounders (objective 4) 13
3.4 Relationships between DMOs and clinical endpoints (objective 5) 13
4 Results 15
4.1 Marginal distribution of DMO (Objective 1) 15
4.1.1 Parkinson's Disease 15
4.1.2 Multiple Sclerosis 17
4.1.3 Proximal Femur Fracture 19
4.1.4 Chronic Obstructive Pulmonary Disease 21
4.2 Marginal distribution of clinical outcomes (Objective 2) 23
4.2.1 Parkinson's Disease 23
4.2.2 Multiple Sclerosis 25
4.2.3 Proximal Femur Fracture 27
4.2.4 Chronic Obstructive Pulmonary Disease 28
4.3 Characteristics of relationships between DMOs (Objective 3) 30
4.3.1 Parkinson's Disease 30
4.3.2 Multiple Sclerosis 30
4.3.3 Proximal Femur Fracture 32
4.3.4 Chronic Obstructive Pulmonary Disease 33
4.4 Characteristics of DMOs conditional on disease characteristics (Objective 4) 34
4.4.1 Parkinson's Disease 34
4.4.2 Multiple Sclerosis 36
4.4.3 Proximal Femur Fracture 37
4.4.4 Chronic Obstructive Pulmonary Disease 38
4.5 Relationships between DMOs and clinical outcomes (Objective 5) 39
4.5.1 Parkinson's Disease 39
4.5.2 Multiple Sclerosis 40
4.5.3 Proximal Femur Fracture 41
4.5.4 Chronic Obstructive Pulmonary Disease 42
5 Conclusions 43
Appendix A. *Technical details on datasets and systems used to obtain DMOs.* 46

1 Abstract

The report summarizes findings coming from the statistical analysis conducted on four pre-existing datasets provided by Mobilise-D partners, covering the primary diseases addressed in the Mobilise-D Project: Parkinson's disease, Multiple Sclerosis, Proximal Femur Fracture, and Chronic Obstructive Pulmonary Disease. The purpose of this document is to provide insights into the statistical characteristics of the distributions of Digital Mobility Outcomes (DMOs), clinical endpoints, and their relationship. The general aim of the report is described along its five objectives which organize the information to be extracted from the available data: (1) describe the marginal distribution of each DMO; (2) describe the marginal distribution of each clinical endpoint; (3) describe the characteristics of relationships between DMOs; (4) describe the characteristics of distribution of DMOs conditional on disease characteristics or confounders; (5) understand and explore relationships between DMOs and clinical endpoints. Details on each of the four datasets, such as study design and sample sizes, are provided, together with a brief list of definitions of which quantities and variables are used in the statistical analyses. Results are displayed and organized according to these five objectives, in the form of tables, figures, and highlights covering salient aspects of the analysed data. Finally, a discussion on the implications of the main findings and the impact of the report on future project activities are left as closing remarks of the document.

2 Introduction

The main objective of WP6 Task 6.1 is to provide a quantitative, statistical characterization of Digital Mobility Outcomes (DMOs) and clinical outcomes distributions in the studied cohorts/disease, as well as to describe the behavior of their relationship. The content of this report – as part of the delivery from Task 6.1 - consists of a summary of statistical analyses performed on datasets relating to the four primary conditions addressed in Mobilise-D: Parkinson's Disease (PD), Multiple Sclerosis (MS), Proximal Femur Fracture (PFF), and Chronic Obstructive Pulmonary Disease (COPD).

Data used to perform the statistical analyses are collected before the clinical validation study and were provided by partners of Mobilise-D. To properly investigate the statistical properties of DMOs and clinical outcomes, feedback from WP2 is incorporated into the process of DMOs selection and their level of definition/granularity. Moreover, results reported here are expected to be integrated into the systematic review lead by WP4, with further implications on tasks WP6.2-6.5. These include sample size (re)calculations, variable selection, and all tasks related to testing construct validity, the predictive capacity of DMOs, and responsiveness/ability to detect clinical outcomes changes.

The report is structured as follows: first, in Section 2, five objectives are presented to explore the overall goal of WP6.1 in a more detailed and systematic way, as well as preliminary definitions of terminology and a description of the data used to produce the deliverable; in Section 3, statistical methods and technical details of the procedures to compute the summaries are reported; in Section 4, results are presented, with an overview of the main findings provided both in textual and tabular format, organized by disease and dataset; Section 5 is devoted to discussing the results and their associated implications; finally, Section 6 hosts closing remarks about the report and its impact.

2.1 Objectives of WP6.1 Task and Datasets

Following the Mobilise-D project plan, the primary purpose of WP6.1 is "understanding distributions of DMOs, clinical outcomes, and the shape of their relationships". The study of DMOs and clinical outcomes – also called 'clinical endpoints' – is carried out by performing statistical analyses on pre-existing datasets, collecting information from previous studies shared by partners of Mobilise-D (to the whole consortium or under bilateral agreements), to be used specifically for the production of this deliverable and without permission to post publicly raw data or any confidential information. A list of these pre-existing datasets is reported in Table 1, together with details about their study design. Each dataset can refer to measurements collected in a laboratory environment, in free-living, or both. The final set of variables considered for the analysis is built jointly by combining clinical variables and DMOs measurements via unique anonymous identifiers assigned to each participant. The reported sample sizes refer to the clinical variables, and sometimes available information for the mobility measurements could be smaller.

The overall goal of WP6.1 can be broken down into five objectives: these sub-tasks give structure to the statistical analysis workflow applied to all datasets. A full description of these five objectives is presented in Table 2, along with the input needed in terms of DMOs and clinical endpoints, and which cohort they refer to among the four disease populations considered in Mobilise-D. All five objectives are of interest for the four diseases, and their practical implementation differs among cohorts only because of disease-specific clinical

endpoints.

Table 1. List of analysed datasets, ordered by cohort/disease, design study, and additional notes; see Appendix A for available references.

Dataset/ Study	Partner/ Cohort	Design	Notes
ICICLE	UNEW/ PD	• Observational; longitudinal; case-control (N=297, 113 cases, 184 controls); • study duration=6 years, observations are available every 18 months.	• in laboratory.
MS Project	USFD/ MS	• Observational; case-control (N=93, 24 controls, 69 cases);	• in laboratory.
Step-by-Step	RBK/ PFF	• Interventional; longitudinal (N=115 total, 58 usual care, 57 intervention) • study duration=3 months, observations are available at t0 (baseline), t1 (2-3 weeks after t0), and follow-up at 3 months.	• free-living; • "No Intervention" group received standard rehabilitation after PFF; • "Intervention" group received standard rehabilitation with additional exercises and enhanced program after PFF.
Urban Training	ISG/ COPD	• Interventional; longitudinal; (N=407 total, 205 usual care, 202 UT, at baseline) • study duration=1 year, observations available at t0 (baseline), and follow-up after 12 months (t1).	• free-living; • Intervention groups are: (i) COPD patients following usual care; (ii) COPD patients following Urban Training protocol; • data are additionally labelled as: (i) Intention-To-Treatment population, individuals allocated at the baseline into "Usual Care" or "Urban Training"; (ii) Per-Protocol population, subjects who truly adhered to their corresponding intervention.

Table 2. List of objectives related to WP6 – Task 6.1 and corresponding cohorts, digital mobility outcomes, and clinical endpoints.

Objective	Outcome involved (Digital Mobility, Clinical, both)	Disease
1) Describe the marginal distribution of each DMO	**Primary DMOs:** *(lab)* Walking speed, Cadence, Stride duration, Stride length; (*free-living)* Walking speed, number of walking bouts, total and average duration of walking bouts, number of Strides, Stride duration, Stride length, Cadence. **Secondary DMOs:** *(lab)* Step duration, Stance duration, Swing duration.	All
2) Describe the marginal distribution of each clinical endpoint ***endpoints to be collected in the clinical validation study but not available in the analyzed datasets**	Late-Life Function & Disability Index (LLFDI)	All
	Primary: Fall frequency*****. **Secondary**: Unified Parkinson's Disease Rating Scale (UPDRS), Hoehn and Yahr (H&Y) scale.	PD
	Primary: Fall frequency*****. **Secondary**: Expanded Disability Status Scale (EDSS), Multiple Sclerosis Walking Scale with 12 items (MSWS-12), Multiple Sclerosis Impact Scale with 29 items (MSIS-29).	MS
	Primary: Admission to a care home. **Secondary**: short Falls Efficacy Scale International (s-FESI), Short Physical Performance Battery (SPPB).	PFF
	Primary: Occurrence of moderate-to-severe exacerbations. **Secondary**: COPD assessment test (CAT), modified Medical Research Council (mMRC) Dyspnea Scale.	COPD
3) Describe the characteristics of relationships between DMOs	**Primary DMOs:** *(lab)* Walking speed, Cadence, Stride duration, Stride length; (*free-living)* Walking speed, number of walking bouts, total and average duration of walking bouts, number of Strides, Stride duration, Stride length, Cadence. **Secondary DMOs:** *(lab)* Step duration, Stance duration, Swing duration.	All
4) Describe the characteristics of distribution of DMOs conditional on disease characteristics or confounders	**DMOs:** Primary and Secondary DMOs. **Disease characteristics:** built on relevant subgroups. **Confounders:** demographics (i.e., sex, age), group (case/control/intervention); see Table 4 for more details.	All
5) Understand and explore relationships between DMOs and clinical endpoints	**DMOs:** Primary and Secondary DMOs. **Clinical endpoints:** disease-specific Primary or	All

2.2 Digital Mobility Outcomes (DMOs) and clinical endpoints

The DMOs previously listed are labelled as either Primary or Secondary. This categorization reflects their importance to the broader purposes of the Mobilise-D Project, and the terminology is coherent with outcomes of WP2 about prioritization for statistical analyses. Table 3 shows the cross-reference between each primary DMO and its availability in the datasets considered. Although the same name/DMOs is considered across datasets, the device used to retrieve the raw signals and the algorithms employed to compute the final measurement could differ among studies and differ from the specifics to be followed in the clinical validation study. In particular, the Primary DMOs recorded in a laboratory environment are processed with the same algorithms considered up to now in Mobilise-D, but the type of device providing DMOs could differ (i.e., DMOs may be calculated by two sensors on each shank or by a sterophotogrammetric system).
Some DMOs from existing datasets were provided a at a step-level, i.e., ICICLE dataset (for secondary DMOs), Step-by-Step, and Urban Training. To provide a more homogeneous analysis, DMOs at step-level are mapped to the stride-level, and the original step-level measurements considered as Secondary DMOs. In terms of walking bout, a broad definition is “a continuous portion of walking”, and for two datasets (Step-by-Step, Urban Training) the same DMOs are analyzed considering either all the walking bouts or just walking bouts longer than 10 seconds.
Additional technical details are provided in Appendix A.

Clinical outcomes or endpoints(these terms a used interchangeably in the report) of interest for Mobilise-D are:

- *Late-Life Functional Disability Index* (LLFDI), a measure in scale 0 to 100, comprised of a functional component (32 items) and a disability component (16 items), which is to be considered a primary endpoint for all cohorts; the higher the value, the higher the impact on both aspects of daily life;
- *Falls*, and in particular frequency of the episodes, considered the primary endpoint for both PD and MS cohorts;
- *Occurrence of moderate-to-severe exacerbation* of COPD-affected subjects is the primary endpoint for COPD cohort, together with *hospital admissions*;
- *Admissions to a care home*, the primary endpoint for PFF cohort, together with *mortality*.

Additionally, dataset- and disease-specific secondary endpoints are considered:

- Part III of the Unified Parkison’s Disease Rating scale (UPDRS-III), score ranging from 0 to 100 with increasing values corresponding to worse condition for the subject affected by PD;
- Hoehn & Yahr (H&Y) scale, a five-levels score ranging between 1-5 with 1.0 increments; also reported as a category variable with: (i) “Mild” level, containing scores 1 and 2 on the original scale; (ii) “Moderate” level, containing score 3 on the original scale;
- levodopa equivalent daily dosage (LEDD), reported in milligrams;
- 12-item Multiple Sclerosis Walking Scale (MSWS-12), a self-reported measure of the impact of MS on walking ability, scaled to the 0-100 range;

- 29-item Multiple Sclerosis Impact Scale (MSIS-29), a self-reported measure of the impact of MS on walking ability, scaled to the 0-100 range;
- Expanded Disability Status Scale (EDSS), with scores ranging from 0 to 10 with 0.5 increments;
- Short Falls Efficacy Scale International (sFES-I), measuring the "concern of falling" of the subjects on a scale from 7 to 28, with higher values associated with an increased fear of experiencing a fall;
- Short Physical Performance Battery (SPPB), scoring from 0 to 12 through a series of physical tests of lower extremities functioning; lower values of the variable corresponding to limited mobility;
- modified Medical Research Council (mRC) Dyspnea scale, a self-rating tool on a range 0-4 with higher values associated with impairment on mobility due to breathlessness;
- COPD assessment test, a self-administered questionnaire that measures health-related quality of life for COPD patients; ranges in the 0-40 interval with higher values indicating a more severe impact of COPD on daily life.

The clinical endpoints reported in Table 2 are already filtered according to whether they are present in the four datasets collected or not.

Finally, Table 4 reconciles information on the presence of outcomes, either digital mobility or clinical, in each dataset, with further details about demographics or clinically relevant quantities to be considered as potential confounders for the statistical analyses to follow.

Table 3. List of Primary Digital Mobility Outcomes and their availability in the four analysed datasets in the report.

Environment	DMOs	Dataset			
		ICICLE (Lab)	MS Project	Step-by-Step	Urban Training
Laboratory	Walking speed (m/s)	•	•	•	
	Cadence (steps per min)	•	•		
	Average Stride duration (in seconds)	•	•		
	Average Stride length (m)	•	•		
Free-living	Daily walking speed (m/s)				
	Daily number of WBs			•	•
	Daily total duration of WBs (s)			•	•
	Daily average duration of WBs (s)			•	•
	Daily number of Strides			•	•
	Daily Stride duration (s)			•	•
	Daily Stride length (m)				
	Daily Cadence (m/s)			•	•

(s): seconds; (min): minutes; (m): meters.

Table 4. Summary of availability of digital mobility outcomes, clinical outcomes, and confounders by dataset.

Dataset/ Study	Partner /Cohort	DMOs available	Clinical Outcomes available	Confounders and sub-groups
ICICLE	UNEW/ PD	**Primary:** walking speed, cadence, avg. stride duration, avg. stride length; **Secondary:** avg. step duration, avg. stance duration, avg. swing duration.	**Primary outcomes**: Falls; **Secondary outcome**: UPDRS, H&Y.	• sex, age, group, time point of measurement; • H&Y scale, retrospective fallers, levodopa equivalent daily dosage (LEDD), freezing of Gait, PD duration.
MS Project	USFD/MS	**Primary:** walking speed, cadence, avg. stride duration, avg. stride length; **Secondary:** /	**Primary outcomes: /** **Secondary outcomes**: MSWS-12, MSIS-29, EDSS.	• sex, age; • use of walking aids, EDSS.
Step-by-Step	RBK/ PFF	**Primary:** number of WBs, total duration of WBs, avg. duration of WBs, number of strides, stride duration, cadence; **Secondary:** /	**Primary outcomes: /** **Secondary outcomes:** s-FESI, SPPB	• sex, age, group, time point of measurement; • number of falls prior to the hip/pelvic fracture.
Urban Training	ISG/ COPD	**Primary:** number of WBs, total duration of WBs, avg. duration of WBs, number of strides, stride duration, cadence; **Secondary:** /	**Primary outcomes:** Exacerbation; **Secondary outcomes:** mMRC Dyspnea scale, CAT.	• sex, age, group, time point of measurement; • mMRC, CAT, population (Per-Protocol or Intention-to-Treatment).

3 Methods

Before adopting any statistical procedure, an aggregation level is set for each DMO considered. The granularity of measurements is specified to be at a walking bout level, as per Mobilise-D definition of walking bout: if multiple observations are available for an individual, they are first averaged for the first layer of aggregation. A further layer of aggregation could be adopted for a dataset with daily measurement by averaging the previous means, completing the information to be one observation per subject. When multiple time points are considered in a dataset, summaries are reported conditional on each fixed point in time. To compute the averages, the following definitions are used.

For each DMO considered, each subject "i" at different time points "t" has an average computed as

$$DMO_{i,t} = \frac{\sum_{j=1}^{J_i} DMO_{i,t,j}}{J_i}$$

where "j" is the index running through all the measurements available for that subject "i" at time point "t" – each subject may have different strides and walking bouts.
When daily data are available, the previous formula is modified to be:

$$DMO_{i,t} = \frac{\sum_{d=1}^{D} DailyDMO_{i,d,t}}{D} \quad \text{with} \quad DailyDMO_{i,d,t} = \frac{\sum_{j=1}^{J_{i,d}} DMO_{i,j,d,t}}{J_{i,d}}$$

where the daily average is first computed, starting from each measurement belonging to subject "i" on a set day and time point, and then the measurements are further aggregated by taking the average across all days considered.
If only one time point is available, the previous formulas do not use "t" as a subscript; moreover, the definition is readily adaptable to find standard deviations of DMOs.

Methodologies and techniques employed to produce the results reported in this deliverable are descriptive and exploratory in nature. Due to the limited sample size and unreliable theoretical assumptions for some of the statistical models considered, results should not be read as inferential snapshots of the analyzed phenomena. Each data set is analyzed independently to better characterize each disease features accordingly, nonetheless following the same workflow. Due to constraints on the size of the report and for the sake of readability, the full statistical analyses are condensed into tables and highlights of the main findings.

Statistical analyses were performed using R software 3.6.0 and the RStudio GUI (graphical user interface, version 1.3.1). Additional packages employed for visualization and data manipulation were: ggplot2, ggridges, tidyverse, reshape2. Density estimation was performed with the default setting provided by the corresponding function 'density()' and 'density_ridges()' – Gaussian kernel, data-driven bandwidth selection. Due to the implementation of the algorithm for computing and visualizing the density, in some plots, the tails of the distributions are trimmed to the highest/lowest observed value. For some variables, visualization might produce density curves with tails having close to zero heights outside the range of existence of the variable (i.e., negative values for standard deviations). These are just visualization artifacts from the algorithm producing the plot. For quantities computed/derived from other DMO(s), density plots are used for visualization, even if the variable is discrete in nature. More details specific to each objective, in terms of statistical analysis considered, are provided in the following sub-sections.

3.1 Marginal distribution of DMO and clinical outcomes (objective 1 and objective 2)

Marginal distributions of numerical variables are summarized in terms of mean, median, standard deviation, minimum, maximum, kurtosis, asymmetry index, sample size, conditional on time points (if more than one available) and subjects groups. Categorical variables are instead summarized in terms of their frequency distribution, absolute and relative, with missing values reported as separate categories. Percentages for categorical variables are computed by looking at the total of non-missing values; percentages of missing values are computed with respect to the sample size.

Numerical continuous variables are represented graphically with density-over-boxplot figures, a nonparametric estimation of the density function for the variable with the corresponding boxplot visualized below it, reported by time points and groups.

For distributions of DMOs, if multiple time points are available, relative changes through time are computed with associated error bars. For each subject, the change in each DMO measurement is measured as the difference in the values between the considered time point and the base time point, divided then by these reference values, where the base time point is the first one available in the data. Changes are summarized by their average and a 95% confidence interval; values are treated as percentages and considered for the summaries, even if not reported in the report.

3.2 Characteristics of relationships between DMOs (objective 3)

Pearson and spearman correlations between DMOs were estimated within each data set and disease. Pearson's type and Spearman's type correlations are computed between all the average measurements of DMOs – this is relevant if a DMO is recorded in terms of average and standard deviation as per daily aggregation level. If multiple measurements are available across time, they are computed at each time point. Correlations are ranked in descending order according to Pearson's correlation absolute value in the first available time point in each table but reported with the original sign for both Pearson and Spearman type of correlations. Pair of correlations reported are the top five ones with an absolute value greater than 0.3 for at least the first time point if multiple are available when considering the correlation values computed in the group of subjects affected the specific disease. Correlations values equal or very close to 1 are expected for functionally dependent couples of DMOs and due to rounding to the select number of decimals.

3.3 Characteristics of DMOs conditional on disease characteristics or confounders (objective 4)

Conditional distributions are reported and computed separately for each combination of the level of the categorical variable(s), used to create clinically relevant sub-groups, and the groups identified by the design study; see Table 4 for a list of confounders and sub-grouping variables. For conditional distributions, only the median value is reported to give insights on the location of the distribution – and the average level – while being robust with respect to the potential effect of outliers. Numerical clinical variables are categorized into two levels: below or equal to the median, and above the median, to have the same framework of comparison throughout all datasets, comparing median values conditional on sub-groups.

3.4 Relationships between DMOs and clinical endpoints (objective 5)

For continuous clinical outcomes, univariate linear regression models are estimated by using each DMO one at a time as a predictor in the model with intercept. Estimated coefficients are reported both for standardized and non-standardized predictors: standardized predictors are obtained by rescaling (dividing by the standard deviation) and centering (subtracting the mean value) the original variable to have a quantity centered in zero and with unitary standard deviation. The intercepts in the fitted models are not reported to ease the table reading. Results are also summarized in terms of 95% confidence interval of the regression coefficients, associated p-values, and R^2 value for the model (as overall goodness of fit).

For binary clinical outcomes, univariate logistic regression models are estimated instead. In this case, coefficients are reported on the odds scale, equivalent to the odds ratio (OR). For interpretation purposes: a predictor with no dependency with the dependent variable has an OR equal to 1, whereas an OR greater (lower) than 1 corresponds to a predictor having a positive (negative) dependency with the dependent binary variable. By subtracting 1 to the estimated coefficients in a logistic regression, one can also compute the increase/decrease in the ratio of probabilities for the two events coded by the binary dependent variable.

4 Results

An overview of the main findings is reported below, organized by objective, disease and dataset considered. This section is devoted to showcasing an extract of the full statistical analysis; a discussion about the potential implications of the findings from the analysis is deferred to Section 5.

4.1 Marginal distribution of DMO (Objective 1)

4.1.1 Parkinson's Disease

I – ICICLE dataset

Table 5. Overview of DMOs' distributions characteristics for ICICLE dataset.

	Apparent differences between control/PD	Distribution shape for cases (PD)	Distribution shape for controls	Larger variability	Apparent changes over time in PD/ controls	Skewness	Potential outliers
Primary DMOs							
Avg. Stride duration	N	bi-modal	normal	PD	N/N	right tail for PD	Y
Avg. Stride length	Y	multi-modal	normal	PD	Y/N	left tail for PD	Y
Cadence	N	skewed	skewed	PD	N/N	right tail for both	N
Walking speed	Y	almost symmetrical	normal	PD	Y/N	-	Y
Secondary DMOs							
Avg. Step duration	N	bi-modal	normal	PD	Y/N	N	Y
Avg. Stance duration	N	almost symmetrical	normal	PD	Y/N	right tail for PD	Y
Avg. Swing duration	N	almost symmetrical	normal	same	Y/N	right tail for PD	Y
Avg. Step Length	Y	skewed	almost symmetrical	PD	Y/N	left tail for PD	Y

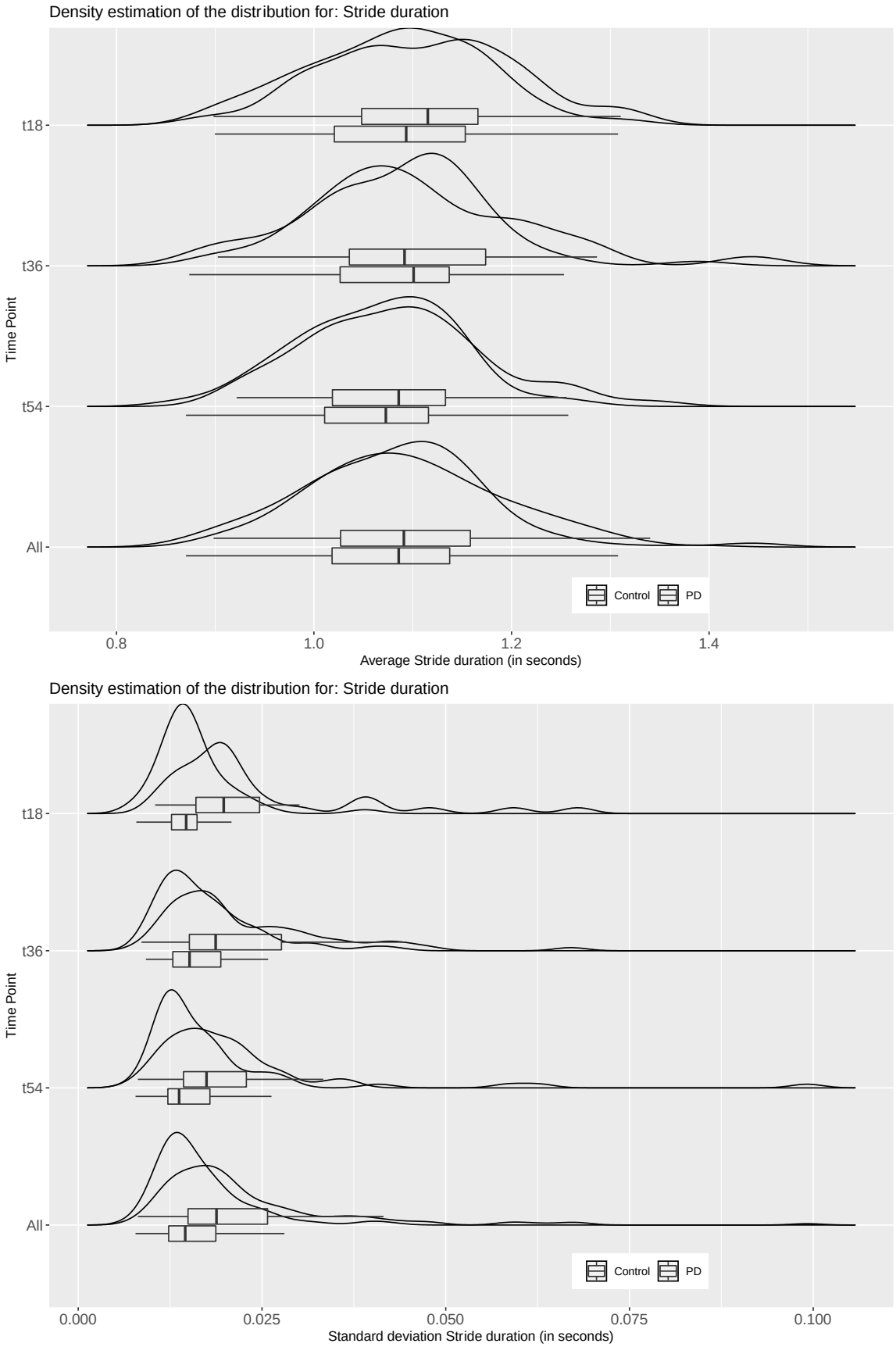

Figure 1. Visualization of distributions for averages (top) and standard deviations (bottom) of Stride Duration, by time point and group; dataset: ICICLE.

Summary:

Primary DMOs:

- most distributions show a departure from the symmetry of a Normal distribution in the PD group, with Stride duration and Stride length also exhibiting multiple modes;
- outliers are present in all distributions of the primary DMOs, except for Cadence which appears to be naturally skewed to the right;
- variability is larger in the PD group for all DMOs.
- mean relative change of DMOs over time are negligible in magnitude for controls; for cases, there is a steady mean decline in average Stride Length and Walking Speed.

Secondary DMOs:

- there is no evidence of difference between controls and cases (PD group) for the three DMOs measuring durations of phases of the gait cycle, that is Step, Stance, and Swing;
- variability is larger in the PD group for almost all DMOs;
- outliers are observed for all DMOs' distributions;
- Step Length for subjects affected by PD has a negative relative change over time of approximately -6% at time point t54 and -9% at t72, on average.

4.1.2 Multiple Sclerosis

I – MS Project

Table 6. Overview of DMOs' distributions characteristics for MS Project dataset.

	Apparent differences between control/MS	Distribution shape for cases (MS)	Distribution shape for controls	Larger variability	Skewness	Potential outliers
Primary DMOs						
Avg. Stride duration	Y	skewed	skewed	MS	right tail for both	Y
Avg. Stride length	Y	almost symmetrical	bi-modal	MS	left tail for controls	N
Cadence	Y	almost symmetrical	skewed	MS	left tail for controls	N
Walking speed	Y	skewed	skewed	MS	left tail for both	N

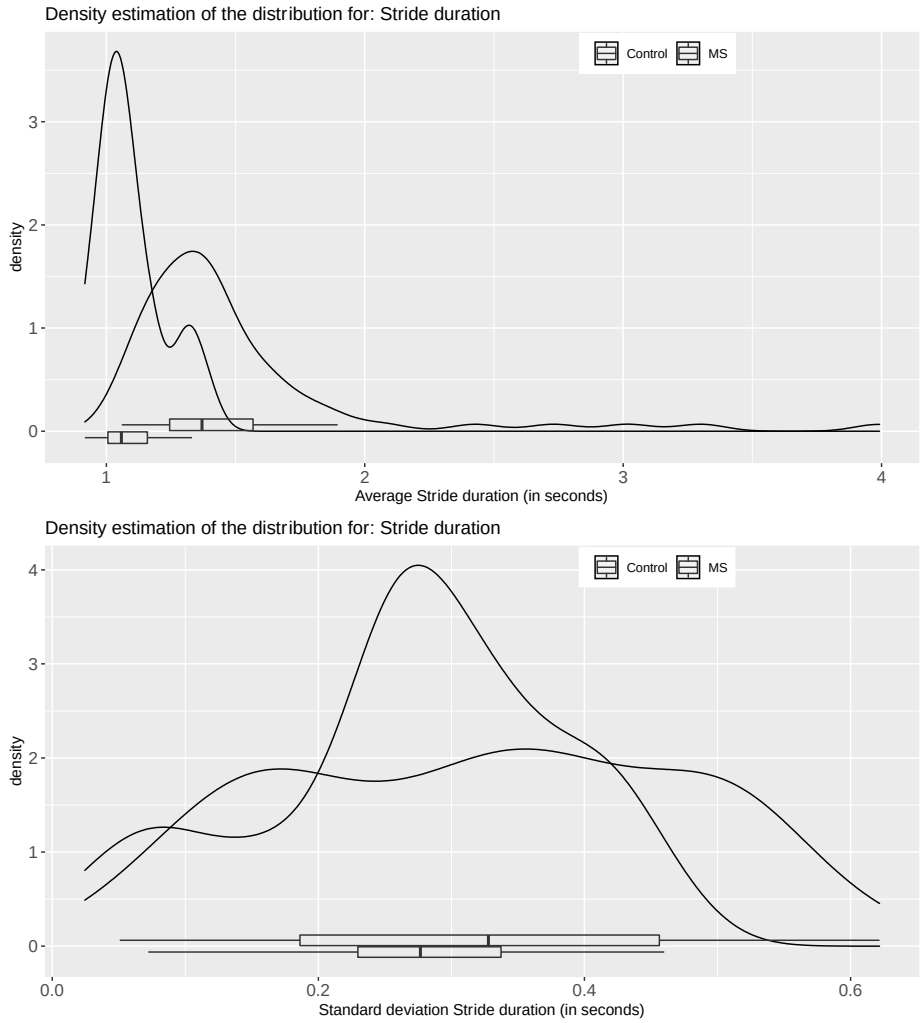

Figure 2. Visualization of distributions for averages and standard deviations of Stride Duration, by group, for the dataset MS Project.

Summary:

Primary DMOs:

- there is evidence of dissimilarities between control and MS group for all the DMOs observed in the dataset;
- variability is larger for subjects in the MS group.

4.1.3 Proximal Femur Fracture

I – Step-by-Step

Table 7. Overview of DMOs' distributions characteristics for Step-by-Step dataset.

	Apparent differences between Int./No. Int	Distribution shape for Int. group	Distribution shape for No. Int. group	Larger variability	Apparent changes over time: Int./No. Int	Skew.	Potential outliers
Primary DMOs (for walking bouts longer than 10 seconds)							
Daily total duration of WBs	Y	Skewed (t0, t2) Almost symmetrical (t1)	Bi-modal	No intervention group	Y/Y	Right tail	Y
Daily avg. duration of WBs	N	Almost symmetrical	Bi-modal (t0) Symmetrical (t1, t2)	No intervention group	N/N	Right tail	Y
Daily number of WBs	N	Skewed	Skewed	No intervention group	Y/Y	Right tail for both	Y
Daily number of strides	Y	Almost symmetrical (t0, t1) Bi-modal (t2)	Skewed (t0, t1) Bi-modal (t2)	No intervention group	Y/Y	Right tail	Y
Daily avg. stride duration	N	Bi-modal (t0, t2) Skewed (t1)	Bi-modal (t0) Almost symmetrical (t1, t2)	No intervention group	Y/Y	Right tail	Y
Daily cadence	N	Skewed (t0) Almost symmetrical (t1) Bi-modal (t2)	Bi-modal (t0, t1) Almost symmetrical (t2)	No intervention group	Y/Y	Left tail	Y
Walking speed (LAB)	N	Skewed (t0) Almost symmetrical (t1) Bi-modal (t2)	Almost symmetrical (t0) Skewed (t1, t2)	Same	Y/Y	Right tail	Y

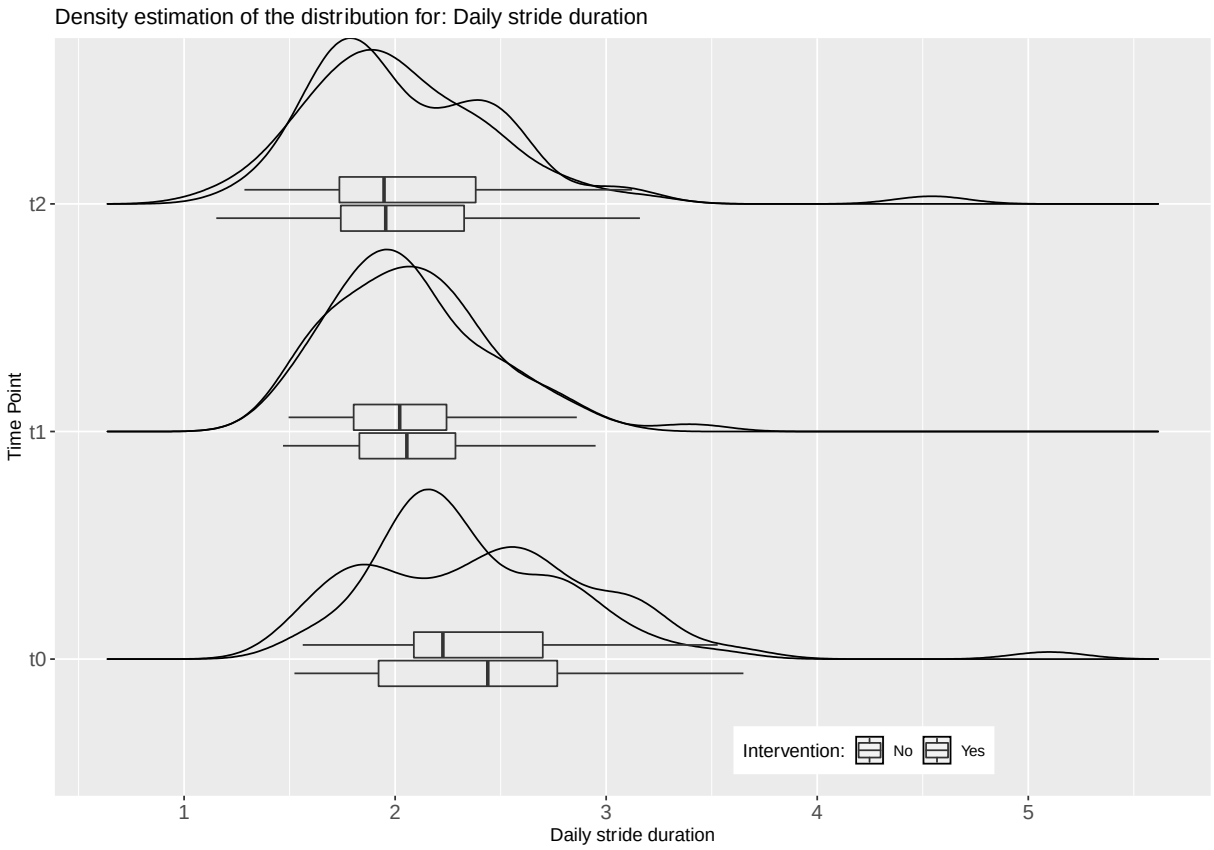

Figure 3. Visualization of distributions for averages of Daily Stride duration (in seconds) for bouts longer than 10 seconds, by time point and group; dataset: Step-by-Step.

Summary:

Primary DMOs:

- the two DMOs exhibit potential differences between the groups “No Intervention” and “Intervention” are the daily total duration of walking bouts (for bouts longer than 10 seconds) and the daily number of strides;
- all the distributions considered show the presence of outliers;
- the behaviour of the distribution, in terms of its average value and/or variability, changes over time for each DMOs considered except the daily average duration of walking bouts (longer than 10 seconds);
- most DMOs have distributions that are not symmetrical;
- variability is larger for subjects in the “No Intervention” group;
- bi-modal distributions are mostly observed at the second follow-up (t2).

4.1.4 Chronic Obstructive Pulmonary Disease

I – Urban Training

Table 8. Overview of DMOs' distributions characteristics for Urban Training dataset, Per-Protocol population.

	Apparent differences between Usual care/UT	Distribution shape for UT group	Distribution shape for control group	Larger variability	Apparent changes over time: control/UT	Skew.	Potential outliers
Primary DMOs (for walking bouts longer than 10 seconds)							
Daily total duration of WBs	Y	Skewed	Symmetrical	Same	N/Y	Right tail	Y
Daily number of WBs	N	Skewed	Almost symmetrical	same	N/Y	Right tail	Y
Daily avg. duration of WBs	N	Skewed	Skewed	same	N/Y	Right tail	Y
Daily avg. stride duration	N	Almost symmetrical	Almost symmetrical	same	N/N	-	Y
Daily number of strides	Y	Skewed	Symmetrical	UT	N/Y	Right tail	Y
Daily cadence	N	Almost symmetrical	Almost symmetrical	same	N/N	Left tail	Y
Primary DMOs (for all walking bouts)							
Daily total duration of WBs	Y	Skewed (t0) Bi-modal (t1)	Almost symmetrical	UT	N/Y	Right tail	Y
Daily number of WBs	Y	Skewed	Almost symmetrical	Same	N/Y	Right tail	Y
Daily number of strides	Y	Skewed	Symmetrical	UT	N/Y	Right tail	Y
Maximum daily walking duration	Y	Skewed	Skewed	UT	Y/Y	Right tail	Y

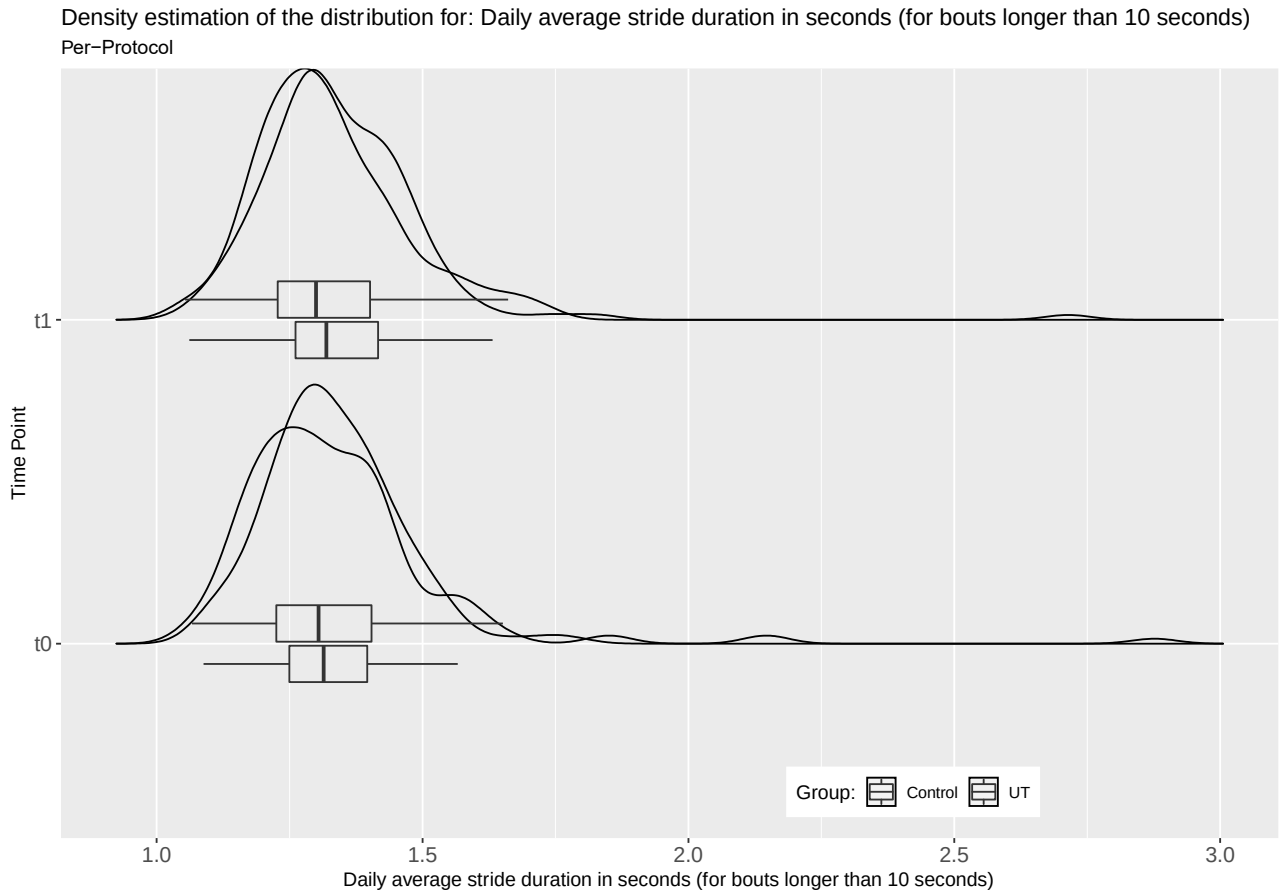

Figure 4. Visualization of distributions for averages of daily Stride Duration (in seconds) for bouts longer than 10 seconds, by time point and group; Per-Protocol population, dataset: Urban Training.

Summary:

Primary DMOs:

- the two DMOs providing some evidence of a difference between Usual Care and Urban Training group are the daily total duration of walking bouts (for bouts longer than 10 seconds) and daily number of strides;
- outliers are observed for all DMOs, whether they refer to walking bouts longer than 10 seconds or walking bouts of any duration;
- almost all distributions of DMOs for the Urban Training group are skewed;
- variability of DMOs is comparable between the two groups Usual care and Urban training when considering walking bouts longer than 10 seconds;
- there is no notable difference between DMOs' summaries and distributions for Per-Protocol or Intention-to-Treatment populations.

4.2 Marginal distribution of clinical outcomes (Objective 2)

4.2.1 Parkinson's Disease

I – ICICLE

Table 9. Summary quantities for clinical outcomes in ICICLE dataset (PD subjects only).

Clinical outcome	Time point				
	t0	**t18**	**t36**	**t54**	**t72**
Any fall (in a one-year window after time point considered), %	Yes=36 (34%) N=106	Yes=60 (64%) N=94	Yes=66 (74%) N=89	Yes=66 (74%) N=89	Yes=77 (87%) N=88
H&Y scale, % **[5 levels scale, 1 to 5]**	Stage 2=82% Stage 3=18% N=85	Stage 2=85% Stage 3=15% N=95	Stage 2=91% Stage 3= 9% N=85	Stage 2=82% Stage 3=18% N=68	Stage 2=71% Stage 3=29% N=57
Any freezing of gait (FoG), %	Yes=10 (9%) N=113	Yes=15 (15%) N=100	Yes=13 (18%) N=74	Yes=7 (12%) N=58	Yes=7 (15%) N=48
UPDRS-III, median (min-max) **[0-100 score]**	25 (7-50)	32.5 (10-60)	39 (12-65)	41 (13-70)	40 (10-69)
LEDD, median (min-max) **[in mg]**	120 (0-880)	358 (0-900)	438 (100-1175)	568 (120-1676)	607 (200-1550)
Number of FoG, median (min-max)	0 (0-18)	0 (0-19)	0 (0-25)	0 (0-23)	0 (0-21)

Summary:

- falls frequency in the PD group increases over time from the baseline (t0) to the last follow-up (t72)
- highest percentage of subjects affected by PD and in the category "3" of the H&Y scale (five levels scale with scores from 1 to 5) is observed at the last time point;
- distribution of UPDRS-III for PD subjects shifts to the right and departs from symmetry over time; this corresponds to larger mean values of the score and larger variability;
- distribution of levodopa equivalent daily dosage (LEED) follows the same pattern of the UPDRS-III distribution, with increased mean and media values over time, as well as variability; additionally, outliers and bi-modality is observed at the last two time points (54 months after follow-up and 72 months after follow-up).

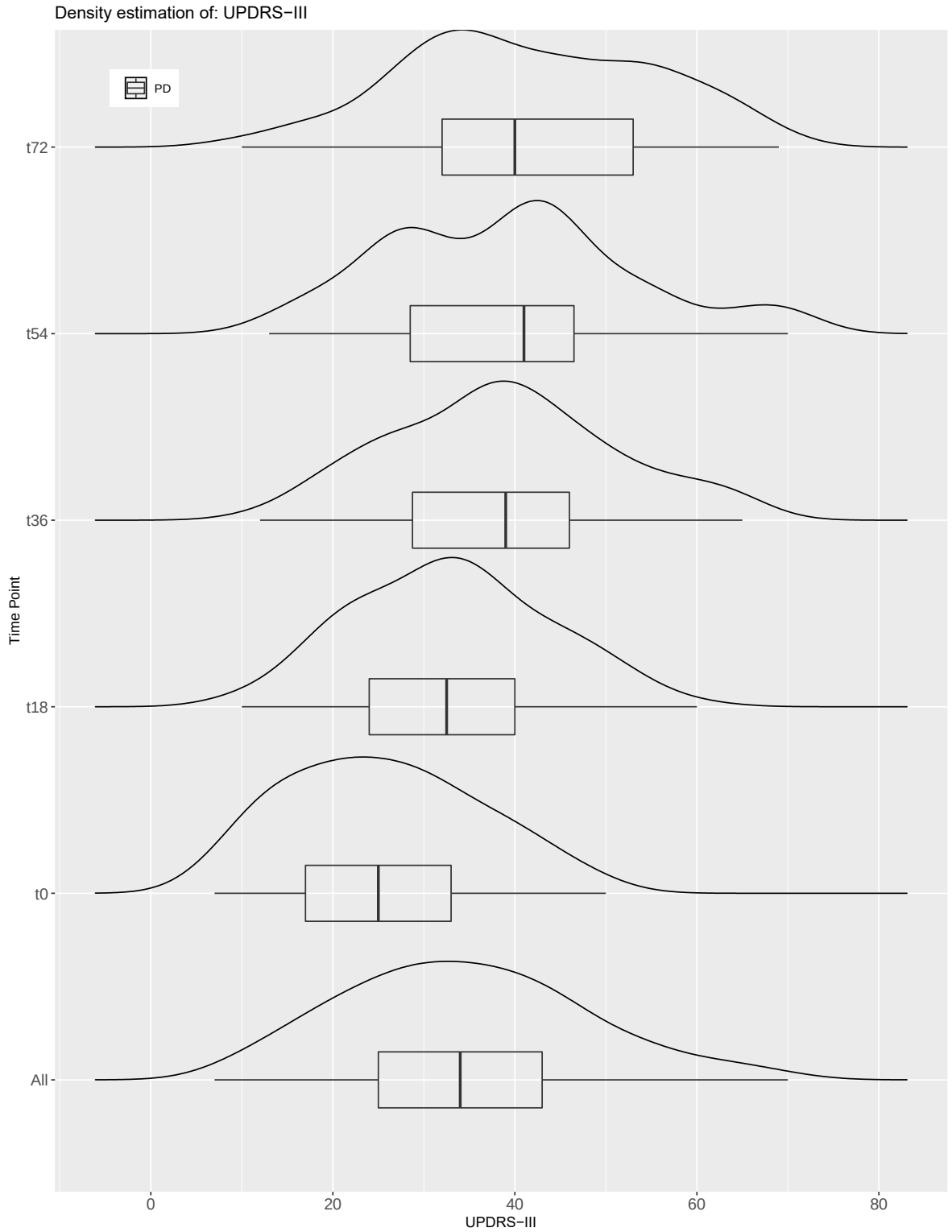

Figure 5. Visualization of distributions for UPDRS-III, by time point, for PD subjects only; dataset: ICICLE.

4.2.2 Multiple Sclerosis

I – MS Project

Table 10. Summary quantities for clinical outcomes in MS Project dataset (MS subjects only).

Clinical outcome	Group = "MS"
Use of walking aids, %	Yes=30 (45%) N=67
Expanded Disability Status Scale (EDSS), % [scale from 0 to 10 with 0.5 increments]	up to score 2.5=1 (1.4%) up to score 3.5=3 (4.3%) up to score 4.5=22 (31.8%) up to score 5.5=37 (53.5%) up to score 6.5=69 (100.0%) N=69
Multiple Sclerosis Impact Scale - 29 items (MSIS-29), median (min-max) [score from 0 to 100]	63.4 (30.2-100)
Multiple Sclerosis Walking Scale - 12 items (MSWS-12), median (min-max) [score from 0 to 100]	74.1 (33.3-100)

Summary:

- distribution of MSIS-29 for people affected by MS is bi-modal, with two modes respectively located at (approximately) 45 and 70;
- distribution of MSWS-12 in subject belonging to group MS is skewed, with a longer left tail;
- correlation value between MSIS-29 and MSWS-12, which are self-reported measurements of the impact of the disease on the mobility of the subject, is above 0.4;
- correlations values between both MSIS-29 and MSWS-12 and a non-self-reported measure, such as the timed 25 feet walking (t25fw) test on the same subjects, are below 0.2 in absolute value.

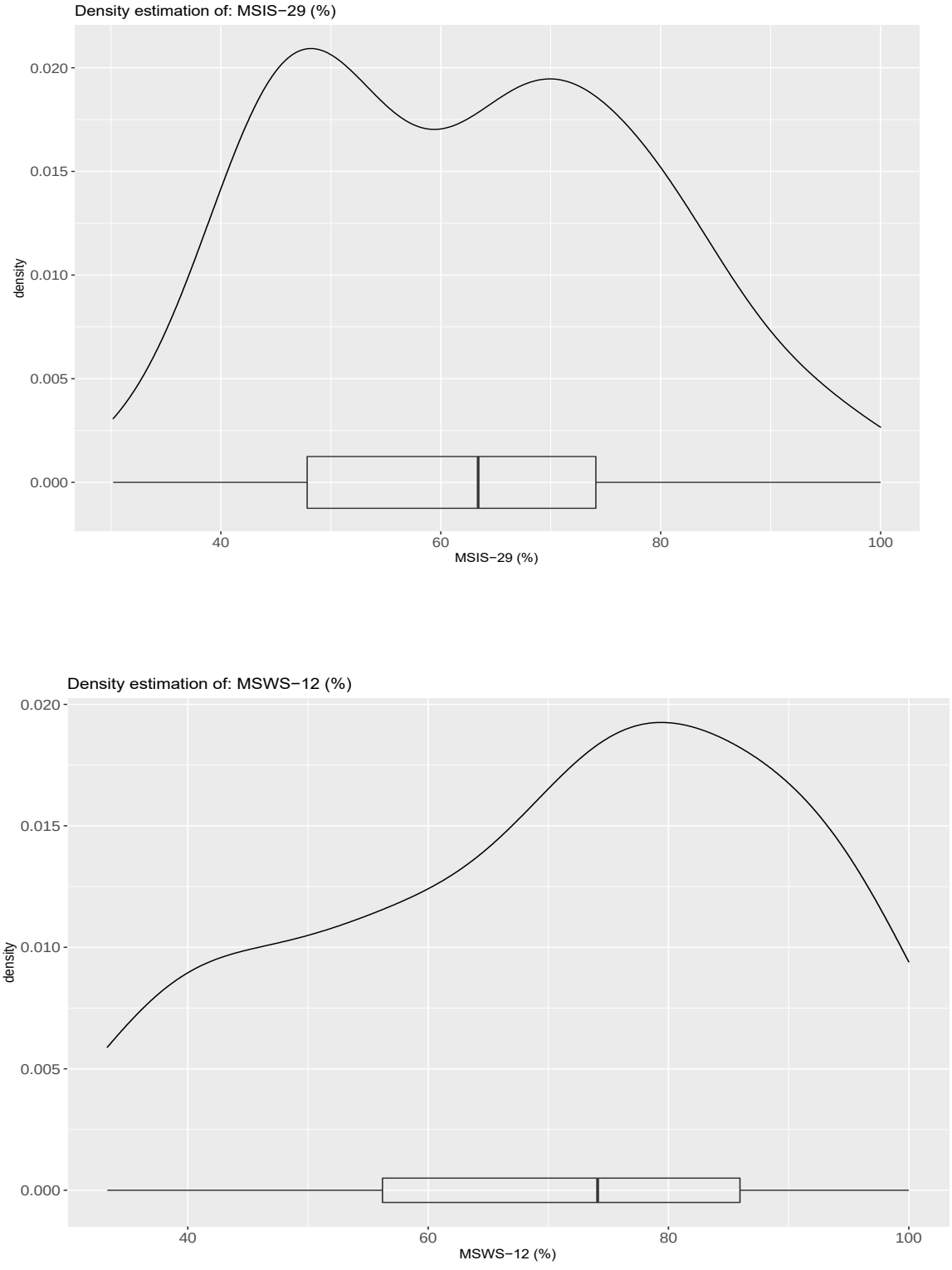

Figure 6. Visualization of distributions of MSIS-29 (%) and MSWS-12 (%) for MS subjects only: MS Project dataset.

4.2.3 Proximal Femur Fracture

I – Step-by-Step

Table 11. Summary quantities for clinical outcomes in Step-by-Step dataset. Numerical variables are summarized as median (min-max).

Clinical outcome	Group	Time point		
		t0	t1	t2
Short Falls Efficacy Scale – International (sFES-I), median (min-max)	No Intervention	15 (7-28)	12 (7-25)	11 (7-24)
	Intervention	16 (7-27)	13 (7-23)	11 (7-28)
Short Physical Performance Battery (SPPB), median (min-max)	No Intervention	3 (0-8)	4 (0-8)	4 (0-11)
	Intervention	3 (0-6)	4 (0-6)	5 (0-10)

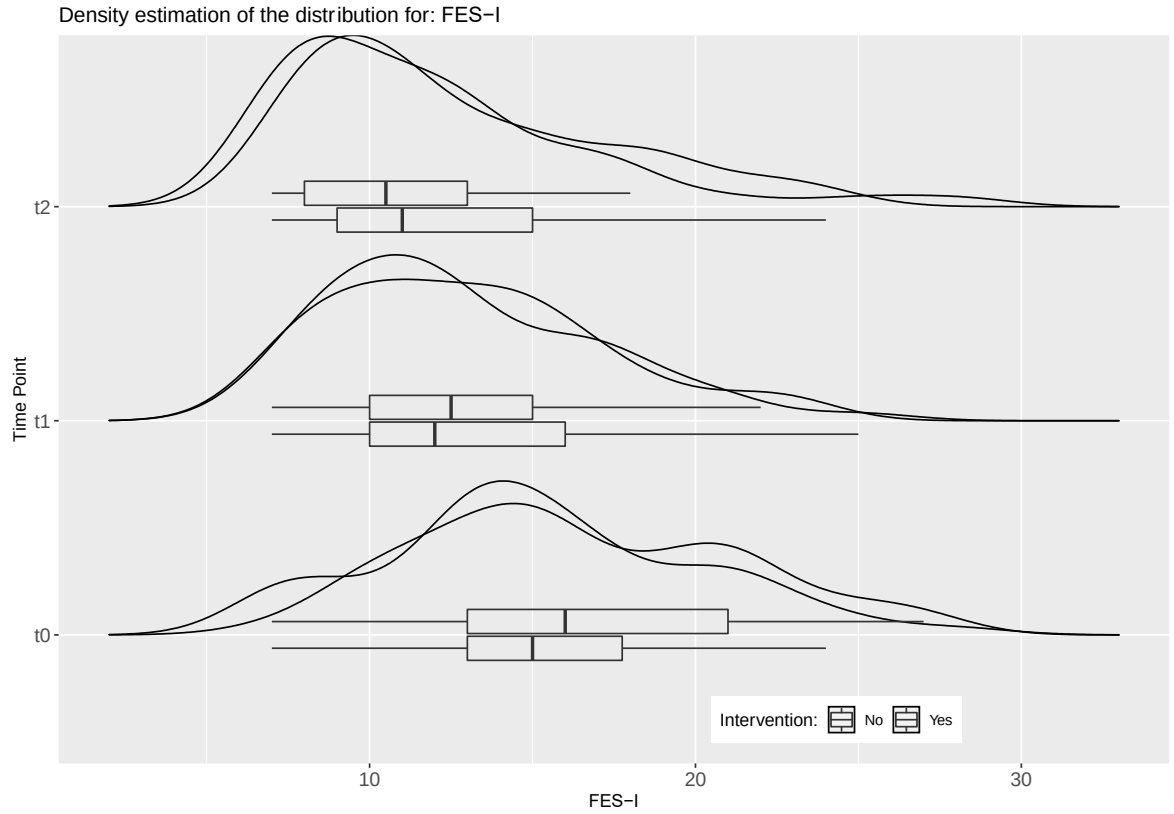

Figure 7. Visualization of distributions for sFESI-I, by time point and group: Step-by-Step dataset.

Summary:

- distribution of sFES-I is shifting to the left over time, corresponding to a decrease in the average score for both groups considered;
- there are some potential outliers observed for the variable sFES-I, especially at the second follow-up;
- distribution of SPPB is stable over time and shows no difference between the two group “No Intervention”/“Intervention”.

4.2.4 Chronic Obstructive Pulmonary Disease

I – Urban Training

Table 12. Summary quantities for clinical outcomes in Urban Training dataset, by time point and group, per Per-Protocol population. Categorical variables are summarized with frequency distributions of relevant categories, with format: event, % event, N available; numerical variables are summarized as median (min-max).

Clinical outcome	Group	Time point			
		t0		t1	
Modified Medical Research Council (mMRC) Dyspnea scale, %	Usual care	0 1 2 or higher N=142	33 (23%) 76 (54%) 33 (23%)	0 1 2 or higher N=143	32 (22%) 71 (50%) 40 (28%)
	Urban Training	0 1 2 or higher N=87	25 (29%) 41 (47%) 21 (24%)	0 1 2 or higher N=87	26 (30%) 36 (41%) 25 (29%)
Severe COPD exacerbations (in the previous 12 months), %	Usual care	Yes=20 (14%) N=141		Yes=23 (17%) N=139	
	Urban Training	Yes=4 (5%) N=86		Yes=12 (14%) N=85	
COPD Assessment Test (CAT), median (min-max)	Usual care	10 (0-37)		10 (0-33)	
	Urban Training	11 (0-35)		10 (0-29)	

Summary:

- the majority of the subjects report a mMRC Dyspnea score of 1 (at baseline: 53.5%, 49.7% at follow-up);
- frequency of severe COPD exacerbations in the 12 months prior to the time point considered range between 5% and 17%;
- distribution of CAT is slightly skewed in both groups considered, that is Usual care and Urban Training, both at baseline and follow-up.

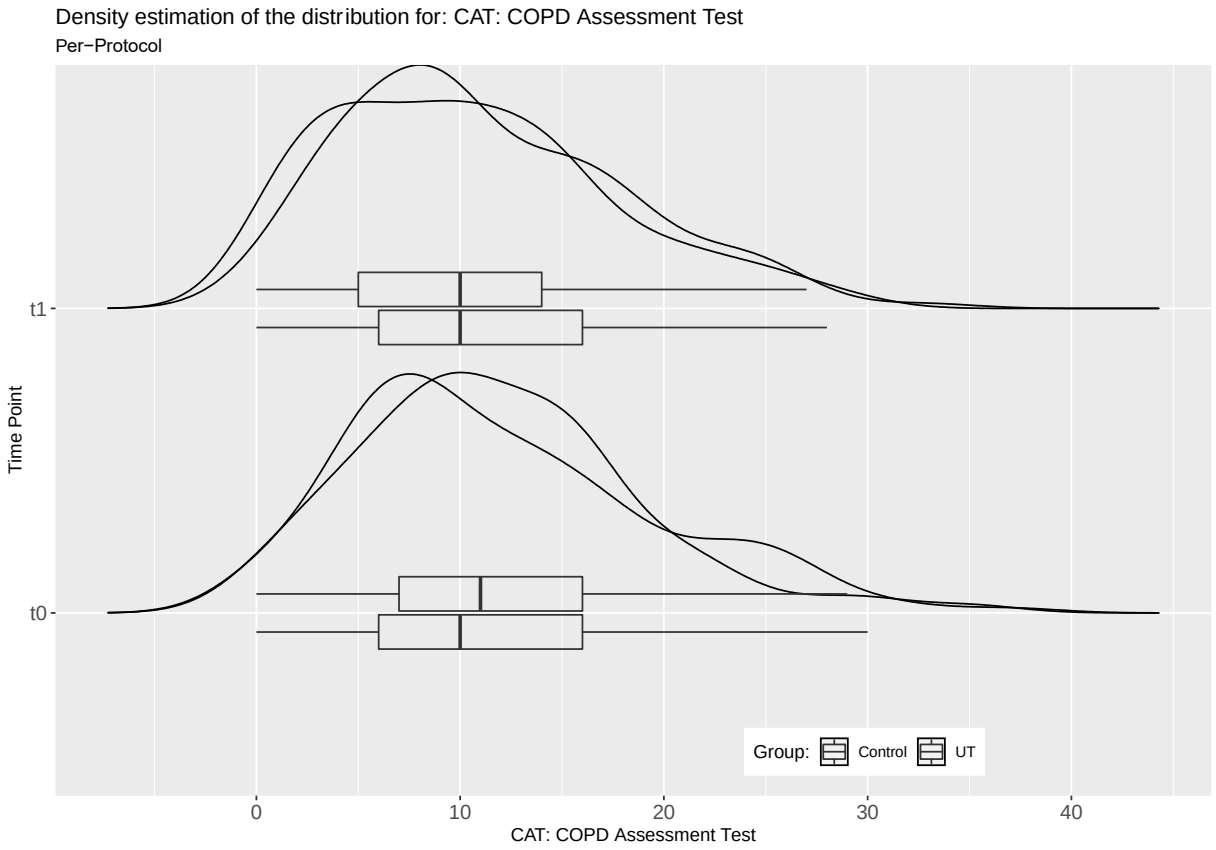

Figure 8. Visualization of distributions for CAT, by time point and group, Per-Protocol population: Urban Training dataset.

4.3 Characteristics of relationships between DMOs (Objective 3)

4.3.1 Parkinson's Disease

I – ICICLE

Table 13. Top (highest, in absolute value) Pearson's correlations among all pairs of DMOs, by group and time point; dataset: ICICLE.

Rank	Pair of variables		Group	t18	t36	t54
1	Stride duration	Cadence	Control	-0.99	-0.99	-1
			PD	-0.99	-0.99	-0.99
2	Stride length	Walking speed	Control	0.78	0.86	0.84
			PD	0.92	0.9	0.92
3	Stride duration	Walking speed	Control	-0.59	-0.57	-0.45
			PD	-0.64	-0.55	-0.48
4	Cadence	Walking speed	Control	0.59	0.54	0.44
			PD	0.64	0.54	0.49

N, control: t18, 48; t36, 63; and t54, 51.
N, PD: t18, 32; t36, 60; and t54, 51.

Summary:

Primary DMOs:

- correlation values close to 1 for the pair Cadence and Stride duration are explained by the functional dependency between the two DMOs, i.e. Cadence is computed as: 60 * number of steps/step duration, and stride and step are conceptually related;
- Spearman's correlations values are similar to those reported in Table 13;
- Pearson's correlations appear slightly larger in magnitude in the PD group with respect to the control group;
- in general, most of the DMOs are strongly correlated with one another.

Secondary DMOs:

- most of the Secondary DMOs are strongly (positively) correlated between one another (range 0.49 to 0.95);

4.3.2 Multiple Sclerosis

I – MS Project

Table 14. Top (highest, in absolute value) Pearson's correlations among all pairs of DMOs, by group; dataset: MS Project.

Rank	Pair of variables		Group	correlation
1	Stride duration	Cadence	Control	-0.99
			MS	-0.92
2	Stride length	Walking Speed	Control	0.47
			MS	0.87
3	Stride duration	Walking Speed	Control	-0.76
			MS	-0.7
4	Cadence	Walking Speed	Control	0.76
			MS	0.7
5	Stride duration	Stride length	Control	0.19
			MS	-0.35
6	Stride length	Cadence	Control	-0.18
			MS	0.28

N, control: 93;
N, MS: 93.

Summary:

Primary DMOs.

- there is some evidence of stronger correlations (in magnitude) between DMOs for subjects in the MS group, in particular for Stride length and Walking speed;
- the last two pairs of DMOs in Table 14 show a different direction of the correlation when comparing the two groups (Control vs MS); this is observed also when considering Spearman's correlations for the same pairs.

4.3.3 Proximal Femur Fracture

I – Step-by-Step

Table 15. Top (highest, in absolute value) Pearson's correlations among all pairs of DMOs, by group and time point, considering only walking bouts longer than 10 seconds; dataset: Step-By-Step.

Rank	Pair of variables		Intervention	t0	t1	t2
1	Cadence	Stride duration	No	-0.97	-0.98	-0.91
			Yes	-0.93	-0.97	-0.97
2	Daily number of strides	Daily total duration of WBs	No	0.95	0.94	0.92
			Yes	0.97	0.9	0.92
3	Daily number of strides	Daily stride duration	No	-0.54	-0.49	-0.58
			Yes	-0.52	-0.51	-0.65
4	Daily number of strides	Cadence	No	0.51	0.47	0.58
			Yes	0.64	0.49	0.63
5	Daily avg. duration of WBs	Cadence	No	0.45	0.43	0.43
			Yes	-0.02	0.4	0.41

Summary:

Primary DMOs:

- Pearson's correlation value of the pair Daily average duration of WBs and the Cadence is 0.45 at baseline for subjects belonging to the "No Intervention" while equal to -0.02 in the "Intervention" group; the very same pair of DMOs has Spearman's correlation values 0.56 in the "No Intervention" group and 0.28 in the "Intervention" group;
- signs of the correlations are stable across time points and reflect expected results for the DMOs recorded.

4.3.4 Chronic Obstructive Pulmonary Disease

I – Urban Training

Table 16. Top (highest, in absolute value) Pearson's correlations among all pairs of DMOs, by group and time point, considering only walking bouts longer than 10 seconds; dataset: Urban Training, Per-Protocol population.

				Per-protocol	
Rank	**Pair of variables**		**Group**	**t0**	**t1**
1	Daily total duration of WBs	Daily number of Strides	Control	0.99	0.99
			UT	0.99	0.99
2	Daily avg. duration of WBs	Cadence	Control	-0.91	-0.91
			UT	-0.97	-0.99
3	Daily total duration of WBs	Daily number of WBs	Control	0.77	0.74
			UT	0.7	0.64
4	Daily number of WBs	Daily number of Strides	Control	0.72	0.68
			UT	0.64	0.59
5	Daily avg. duration of WBs	Daily number of Strides	Control	0.54	0.3
			UT	0.61	0.71

Summary:

Primary DMOs:

- there is moderate to high correlation between DMOs in COPD patients.

4.4 Characteristics of DMOs conditional on disease characteristics (Objective 4)

4.4.1 Parkinson's Disease

I – ICICLE

Table 17. Summary on differences of DMOs' median values between categories of each confounder considered (categories are reported between brackets). ICICLE dataset, only for PD subjects.

Confounders and clinically-relevant subgroups	Median values of DMOs:			
	Avg. stride duration	Avg. stride length	Cadence	Walking bout speed
Age (Below median years*, Above median years) *69 years	Similar	Smaller in group "Above median years"	Similar	Smaller in group "Above median years"
Sex (Female, Male)	Larger in group "Male"	Larger in group "Male"	Smaller in group "Male"	Similar
Hoenn & Yarr scale (Mild, Moderate)	Larger in group "Moderate"	Smaller in group "Moderate"	Smaller in group "Moderate"	Smaller in group "Moderate"
Retrospective Fallers (No, Yes)	Similar	Smaller in group "Retr. Fallers"	Similar	Similar
Levodopa equivalent daily dosage (Below median dosage,* Above median dosage) *120 mg	Similar	Smaller in group "Above median dosage"	Similar	Smaller in group "Above median dosage"
Freezing of Gait (No, At least one)	Similar	Similar	Similar	Similar
PD duration (Below median duration*, above median duration) *6 months	Similar	Similar	Similar	Similar

Summary:

- subjects affected by PD and belonging to a group of higher scores on the H&Y scale exhibit, on average, Step duration, Stance duration, and Swing duration values larger than individuals with a lower H&Y score; additionally, the variability for these DMOs is increased in the “Moderate” H&Y group with respect to the “Mild” H&Y group.
- DMOs distributions by other clinical outcomes (retrospective falls prior to the baseline assessment, PD duration, and on the LEDD dosage), do not distinguish between groups.
- individuals experiencing freezing of gait have larger variability in terms of Gait Speed, but no meaningful departures on average from the values of other DMOs of those who do not have those episodes.
- the variable Age is a potential confounder as seems to be related to different mean values and standard deviations for DMOs: in particular, older subjects have on average lower Walking bout speed and shorter Stride Length.

4.4.2 Multiple Sclerosis

I – MS Project

Table 18. Summary on differences of DMOs' median values between categories of each confounder considered (categories are reported between brackets). MS Project dataset, only for MS subjects.

Confounders and clinically-relevant subgroups	Median values of DMOs:			
	Avg. stride duration	Avg. stride length	Cadence	Walking bout speed
Age (Below median* years, Above median years) *54 years	Similar	Similar	Similar	Similar
Sex (Female, Male)	Similar	Similar	Similar	Similar
Use of walking aids (None, Any)	Larger in group "Any walking aid"	Smaller in group "Any walking aid"	Smaller in group "Any walking aid"	Smaller in group "Any walking aid"
Extended Disability Status Scale EDSS (Below median*, Above median) *5.5 score	Larger in group "Above median"	Smaller in group "Above median"	Smaller in group "Above median"	Smaller in group "Above median"
Levodopa equivalent daily dosage (Below median dosage, Above median dosage)	Similar	Smaller in group "Above median dosage"	Similar	Smaller in group "Above median dosage"
Freezing of Gait (No, At least one)	Similar	Similar	Similar	Similar
PD duration (Below median duration, above median duration)	Similar	Similar	Similar	Similar

Summary:

- there is evidence of differences between sub-groups of subjects affected by MS conditional: use of a walking aids and the levels of EDSS (below or above the median of 5.5);
- there is no substantial difference between male and female MS patients, as well as between people younger or older than 54 years (mean value of the Age variable).

4.4.3 Proximal Femur Fracture

I – Step-by-Step

Table 19. Summary on differences of DMOs' median values between categories of each confounder considered (categories are reported between brackets), only for subject belonging to the "Intervention" group; Step-by-Step dataset.

Confounders and clinically-relevant subgroups	Median values of DMOs:			
	Stride duration	Number of walking bouts (longer than 10 seconds)	Avg. walking bout duration (longer than 10 seconds)	Cadence (steps/min)
Age (Below median years*, Above median years) *83 years	Similar	Similar	Similar	Larger in group "Above median years"
Sex (Female, Male)	Similar	Smaller in group "Male"	Larger in group "Male"	Similar
Additional fall episodes prior to the fracture (None, One or more falls)	Similar	Smaller in group "One or more falls"	Larger in group "One or more falls"	Similar

Summary:

- stride duration appears to have the same median values across different sub-groups defined on categorical variables: sex, age, and the number of prior falls;
- Cadence is the only DMOs for the "Intervention" group that shows a difference with respect to classes of Age (subjects below 83 years of age VS older subjects).

4.4.4 Chronic Obstructive Pulmonary Disease

I – Urban Training

Table 20. Summary on differences of DMOs' median values between categories of each confounder considered (categories are reported between brackets), only for subject belonging to the "Urban Training" group and "Per-Protocol" population; Urban Training dataset.

Confounders and clinically-relevant subgroups	Median values of DMOs:			
	Avg. stride duration	Number of walking bouts (longer than 10 seconds)	Avg. walking bout duration (longer than 10 seconds)	Cadence (steps/min)
Modified Research Council Dyspnea scale (Low: 0-1, Medium to Severe: 2-3-4)	Similar	Smaller in group "Medium to Severe"	Smaller in group "Medium to Severe"	Similar
COPD Assessment test (Below or equal to median score*, Above median score) *11 value	Similar	Smaller in group "Above median score"	Similar	Similar

Summary:

- among the sub-groups of subjects considered, for the "Urban Training" group only the number of walking bouts (for bouts longer than 10 seconds) shows a difference when comparing different levels of both the mMRC Dyspnea scale and CAT.
- median maximum observed walking duration (not reported) has lower values for subject in the control group with respect to the UT group, in both Per-Protocol and ITT populations, when considering the subjects with scores of the CAT below the median value of 11.

4.5 Relationships between DMOs and clinical outcomes (Objective 5)

4.5.1 Parkinson's Disease

I – ICICLE

Table 21. List of univariate linear regression models fitted on data at 36 months after baseline (t36), for subjects affected by PD. ICICLE dataset.

Dependent variable	Predictor in the model	Standardized predictor: Coefficient (95% CI)	Original-scale predictor: Coefficient (95% CI)	p-value	R^2
	Primary DMOs (N=55)				
UPDRS-III score	Average stride duration	5.34 (2.36, 8.31)	48.02 (21.26, 74.77)	0.001	0.19
UPDRS-III score	Average stride length	-5.29 (-8.21, -2.36)	-23.67 (-36.79, -10.56)	0.001	0.19
UPDRS-III score	Cadence	-5.27 (-8.15, -2.39)	-0.51 (-0.78, -0.23)	0.001	0.2
UPDRS-III score	Walking Speed	-6.11 (-8.74, -3.47)	-25.79 (-36.9, -14.67)	0.000	0.28

Table 22. List of univariate logistic regression models fitted on data at 36 months after baseline (t36), for subjects affected by PD. ICICLE dataset. Coefficients are reported on the odds scale; NC=not computable.

Dependent variable	Predictor in the model	Standardized predictor: Coefficient (95% CI)	Original-scale predictor: Coefficient (95% CI)	p-value
	Primary DMOs (N=47)			
Any fall? (0=No, 1=Yes)	Average stride duration	1.68 (0.87, 3.64)	107.09 (0.29, NC)	0.146
Any fall? (0=No, 1=Yes)	Average stride length	0.54 (0.26, 1.05)	0.07 (0, 1.26)	0.086
Any fall? (0=No, 1=Yes)	Cadence	0.63 (0.31, 1.18)	0.96 (0.89, 1.02)	0.169
Any fall? (0=No, 1=Yes)	Walking Speed	0.5 (0.23, 0.95)	0.05 (0, 0.81)	0.047

All models considered are fitted on a single time point, 36 months after baseline, as it provides the most amount of data without the presence of outliers.

Primary DMOs:

- in the univariate linear models with UPDRS-III as a response variable, Walking Speed has the highest negatively associated coefficient (-6.11), with a 95% confidence interval of (-8.74, -3.47) on the standardized scale, and (-25.79) (95%CI= -36.90,-14.67) on the original scale.

For logistic regression models on the falls frequency, the only relevant predictor is the average Walking Speed, with an OR of 0.5 (95%CI= 0.23, 0.95).

4.5.2 Multiple Sclerosis

I – MS Project

Table 23. List of univariate linear regression models fitted on data for subjects affected by MS; dependent variables are: MSWS-12 (%) and MSIS-29 (%); MS Project dataset.

Dependent variable	Predictor in the model	Standardized predictor: Coefficient (95% CI)	Original-scale predictor: Coefficient (95% CI)	p-value	R^2
	Primary DMOs (N=63)				
MSWS-12 (%)	Average stride duration	6.23 (1.73, 10.74)	11.7 (3.24, 20.15)	0.01	0.11
MSWS-12 (%)	Average stride length	-6.03 (-10.58, -1.47)	-26.68 (-46.84, -6.52)	0.01	0.1
MSWS-12 (%)	Cadence	-6.15 (-10.66, -1.63)	-0.34 (-0.59, -0.09)	0.01	0.11
MSWS-12 (%)	Walking Speed	-7.73 (-12.09, -3.37)	-29.15 (-45.59, -12.71)	0.000	0.17
MSIS-29 (%)	Average stride duration	0.61 (-3.3, 4.51)	1.14 (-6.19,8.47)	0.76	0
MSIS-29 (%)	Average stride length	-2.51 (-6.39, 1.37)	-11.1 (-28.28, 6.09)	0.21	0.03
MSIS-29 (%)	Cadence	-2.42 (-6.28, 1.44)	-0.14 (-0.35, 0.08)	0.22	0.02
MSIS-29 (%)	Walking Speed	-3.46 (-7.28, 0.35)	-13.06 (-27.44, 1.32)	0.08	0.05

- there is no evidence of a predictive capability of the measured DMOs for MSIS-29 through a linear regression model: 95% confidence intervals for univariate models with MSIS-29 as the response variable and each DMO, one at a time, as a predictor contain the value zero;
- all associated p-values are higher than the usual confidence levels (1%, 5%, 10%), with the only exception of the average Walking Speed (p-value 0.08).
- models with MSWS-12 as the response variable all have regression coefficients with intervals not including the zero, and significant p-values;
- the highest estimated effect associated to the average Walking Speed is equal to -7.73 (95%CI= -12.09, -3.37) for the standardized version of the predictor and -29.15 (95%CI= -45.59, -12.71) on the original scale.

4.5.3 Proximal Femur Fracture

I – Step-by-Step

Table 24. List of univariate linear regression models fitted on data for subjects affected by PFF and belonging to the "Intervention" group; dependent variables are time differences from baseline to first follow-up for: Short Physical Performance Battery (SPPB), short Falls Efficacy Scale – International (sFES-I) ; Step-by-Step dataset.

Dependent variable	Predictor in the model	Standardized predictor: Coefficient (95% CI)	Original-scale predictor: Coefficient (95% CI)	p-value	R^2
	Primary DMOs (N=57) for walking bouts longer than 10 seconds				
SPPB (t1-t0)	Total daily walking interval duration	0.33 (-0.2, 0.86)	0 (0, 0.001)	0.229	0.03
SPPB (t1-t0)	Average walking interval duration	0.049 (-0.489 0.587)	0.004 (-0.036, 0.044)	0.859	0
SPPB (t1-t0)	Number of walking intervals	0.472 (-0.04, 0.983)	0.017 (-0.001, 0.036)	0.077	0.06
SPPB (t1-t0)	Daily number of strides	0.485 (-0.025, 0.995)	0.001 (0, 0.002)	0.068	0.06
SPPB (t1-t0)	Daily stride duration	-0.189 (-0.722, 0.344)	-0.399 (-1.523, 0.725)	0.49	0.01
SPPB (t1-t0)	Daily Cadence	0.206 (-0.327, 0.738)	0.019 (-0.03, 0.068)	0.452	0.01
Primary DMOs (N=57) for walking bouts longer than 10 seconds					
sFES-I (t1-t0)	Total daily walking interval duration	-0.969 (-2.251, 0.314)	-0.001 (-0.003, 0)	0.145	0.04
sFES-I (t1-t0)	Average walking interval duration	0.701 (-0.595, 1.998)	0.052 (-0.044, 0.149)	0.294	0.02
sFES-I (t1-t0)	Number of walking intervals	-0.976 (-2.246, 0.295)	-0.036 (-0.082, 0.011)	0.138	0.04
sFES-I (t1-t0)	Daily number of strides	-0.887 (-2.163, 0.388)	-0.002 (-0.004, 0.001)	0.179	0.04
sFES-I (t1-t0)	Daily stride duration	1.293 (0.027, 2.559)	2.729 (0.057, 5.401)	0.051	0.08
sFES-I (t1-t0)	Daily Cadence	-1.165 (-2.44, 0.111)	-0.107 (-0.225, 0.01)	0.08	0.06

Univariate linear models are fitted on the subject-specific differences between response variables at time t1-t0, and at time t2-t0; these models, by construction, account for the discrepancies from two different time points that are due to subject-specific characteristics.

- for SPPB differences between t1 minus t0, only gait speed measure in laboratory has an associated significant coefficient, with a standardized value of the regression beta equal to 0.633 (95%CI= 0.188, 1.138), for both no intervention and “Intervention” group;
- when considering the differences computed as SPPB at time t2 minus the same variable at time t0, gait speed in laboratory has a significant coefficient only within the “Intervention” group, with a value of 0.706 (95%CI= 0.236, 1.175) for the standardized predictor.
- when the response variable considered is sFES-I change from t0 to t1, gait speed in laboratory is again the only predictor with a significant coefficient, only however in the “No Intervention” group.

4.5.4 Chronic Obstructive Pulmonary Disease

I – Urban Training

The summary table of predictor’s coefficients is omitted upon request by the dataset providers, due to fitted models being results on unpublished data.

- in the Per-Protocol population, models (logistic univariate regressions) with the variable COPD Exacerbation (No, Yes) all have regression coefficients with confidence intervals including the zero, thus pointing towards non-significant effects at the pre-specified 5% level of type I error; the only exceptions is Cadence computed for walking bouts longer than 10 seconds.
- in the ITT population, not only for Cadence but also the average daily stride duration for walking bouts longer than 10 seconds has a significant coefficient in the model.

5 Discussion

Results reported in Section 4 cover the main findings from the full statistical analysis performed on the four datasets. Although specific outputs of the analysis should be confined to the datasets considered and not extrapolated, there are some considerations on the features of DMOs and clinical endpoints that can be transferred and applied to other tasks and analysis of Mobilise-D Project. These key points are listed below.

- careful consideration should be payed to extreme values in the distributions of the DMOs, particularly in labelling them as either 'expected' for that particular DMO, and thus 'outliers'/rare events in the classical statistical sense, or outliers due to technical issues in either the recording sensor or the algorithm processing the signal; this is especially true if maximum of DMOs or high percentiles are quantities to be used in modelling the relationship between DMOs and clinical endpoints;
- additionally, given the aggregation level considered in the analysis, which lead to computing individual averages of measurements already averaged across walking bouts, even more extreme values are expected if a finer level of aggregation is used, i.e. looking at daily DMOs observations of a person;
- outliers in the distributions of Primary and Secondary DMOs have been observed in all datasets, both from laboratory and free-living environments, which means there is no specific set of DMOs to pinpoint with respect to the previous comments;
- another feature of the distribution of some DMOs was bi-modality, that is the absence of a single 'most probable' values to observe on average, with two different candidates. In particular, among Primary DMOs, bi-modality was observed for:
    - (*ICICLE dataset*) avg. stride duration and avg. stride length;
    - (*Step-by-Step dataset*) total and avg. duration of walking bouts, number of strides and their average duration, as well as Cadence;
    - (*Urban Training dataset*) total duration of walking bouts;
- behind bi-modality there is usually a hidden variable not considered in the computation and visualization of the distribution of the DMO, which effectively partitions the observations in sub-groups that induces one (or more) modes in the "marginal" view of the DMO; however, there are also additional considerations specific to each dataset:
    - (*Step-by-Step*) the observational window at the first follow-up is not the same for every subject, and this temporal mismatch might be reflected in the distribution of the DMOs;
    - (*Urban Training dataset*) the data presented in the report are aggregated as described in Section 3: observations for each subject over a full week timeframe have been averaged into a single measurement at each time point, which means that two potential behaviour describing mobility during weekdays and mobility during weekends are confounded with one another, potentially inducing a bi-modal "marginal" distribution of the DMOs.
- similar considerations on bimodality are valid for clinical endpoints, where the variable might be impacted by different 'severity groups' producing different modes on the distribution of the clinical outcome:
    - (*ICICLE*) UPDRS-III and LEDD at the last two follow-up, 54 months and 72 months after baseline;

  - (*MS Project*) MSIS-29;
  - (*Step-by-Step*) sFES-I;

- most observed clinical endpoints' distributions were not close to symmetry, and variability of these outcomes for dataset with multiple time points was not constant over time: this implies that statistical models or procedures assuming Normality of the variable might not be appropriate for the endpoints considered in this report. For example, a broader class of predictive models that can account for non-normality of the response variable and thus potentially more adequate for these clinical outcomes are the generalized linear models;
- binary clinical endpoints analyzed in the available pre-existing dataset did not show any unbalanced class issue, with enough observations for both levels of the categories, i.e. falls frequency, or retrospective fallers; categorical clinical outcomes on scales, such as EDSS, Hoehn & Yahr score, or with many levels, could be either categorized into a binary version or collapsed into fewer levels, with two distinct benefits: (i) less parameters to be estimated in a statistical model when the clinical endpoints are used either as response variables or predictors; (ii) an increase in the effective number of observations per class;
- when inspecting correlations between DMOs, both Pearson's and Spearman's correlation should be computed: the latter is more robust with respect to outliers, and can capture also non-linear (monotone) relationships between two variables; some of the DMOs are functionally related, such as Cadence and Stride Duration, due to how are computed, and thus should not be used together as predictors; some options to avoid redundancy of information and problems of multicollinearity are: (i) selecting the pairs of least correlated DMOs; (ii) choose a DMO defined as an average, i.e. average duration of walking bouts as a summary quantities for two other related DMOs, which are the total duration of walking bouts and the number of walking bouts;
- statistical models with clinical endpoints as response variable should consider the design of the study the data come from, and account as much as possible for all known sources of variability: in particular, for dataset with repeated measures and different time points, mixed effect models allow the use of random effect to capture the subject-specific variability and the temporal variability in the data, that could otherwise potentially mask the effect of the DMOs in predicting a response variable;

Finally, as expected, data coming from a free-living environment are in general more variable and less reliable in finding evidence on the relationship between DMOs and clinical endpoints. In fact, although not fully comparable, DMOs were predictive of the outcome in linear models fitted on dataset from the laboratory, such as ICICLE, whereas for dataset from free-living environment the same DMOs had no significant coefficient. In this sense, free-living data are: (i) unavoidably affect by more sources of noise; (ii) a less validated setting with respect to DMO extraction (i.e., DMOs from the pre-existing dataset were not extracted with Mobilise-D algorithms); (iii) the best aggregation modalities of DMOs, when considering continuous recording, are not yet known. In free-living, the main primary outcome of Mobilise-d (real-world walking speed) was not available in these pre-existing datasets. However, from these exploratory analyses, two DMOs showed promising characteristics: daily total duration of walking bouts and daily number of strides, in two very different populations (PFF and COPD), exhibited capability in differentiating between the "No Intervention"/"Intervention" groups.

6 Conclusions

This report provided an overview of the exploratory statistical analyses performed on pre-existing datasets provided by partners of the Mobilise-D Project. The task was to study the distribution of Digital Mobility Outcomes (DMOs), clinical outcomes, and their relationship, to increase knowledge for future use of these outcomes in other Tasks of the Project, as well as to plan further statistical analyses accordingly. Considerations on the distributions of the DMOs, as well as features of clinical endpoints and their joint behaviour can inform decisions on future steps of Mobilise-D, especially in terms of which statistical tools and models to use, as well as any preparation and selection of variables considered for the analysis.

Appendix A. *Technical details on datasets and systems used to obtain DMOs.*

General note:

For ICICLE and MS project datasets, the laboratory-based DMOs were not derived from the wearable sensor on the lower back (like the one that will be used in Mobilise-D) but from reference systems. Reference systems were gold standard (stereophotogrammetric system) or silver standard (two sensors on shanks). This was done in order to have DMO values that could follow as close as possible Mobilise-D definitions and guidelines.

– **ICICLE**

Setting: Laboratory

Test: 2 minute walking test over 25 meters circuit.

System used to obtain the DMOs:

Gaitrite electronic walkway. DMOs were computed for each pass on the instrumented walkway, which was positioned on one side of the circuit.

References:

Del Din S, Godfrey A, Rochester L. Validation of an Accelerometer to Quantify a Comprehensive Battery of Gait Characteristics in Healthy Older Adults and Parkinson's Disease: Toward Clinical and at Home Use. IEEE J Biomed Health Inform. 2016;20(3):838-47.

– **MS Project**

Setting: Laboratory

Test:

6 minute walking test over 10 meters (straight)

System used to obtain the DMOs:

Two IMUs, placed bilaterally on the shanks. The sensors were Opal (APDM), with accelerometer and gyroscope, 128 Hz sampling frequency. The algorithm to process the data and compute DMOs was provided by EPFL (see references).

References - dataset:

Angelini, L., Hodgkinson, W., Smith, C. *et al.* Wearable sensors can reliably quantify gait alterations associated with disability in people with progressive multiple sclerosis in a clinical setting. *J Neurol* **267,** 2897–2909 (2020). https://doi.org/10.1007/s00415-020-09928-8

Angelini, L.; Carpinella, I.; Cattaneo, D.; Ferrarin, M.; Gervasoni, E.; Sharrack, B.; Paling,

D.; Nair, K.P.S.; Mazzà, C. Is a Wearable Sensor-Based Characterisation of Gait Robust Enough to Overcome Differences Between Measurement Protocols? A Multi-Centric Pragmatic Study in Patients with Multiple Sclerosis. *Sensors* **2020**, *20*, 79.

References – System:

Salarian, A., Burkhard, P. R., Vingerhoets, F. J., Jolles, B. M., & Aminian, K. (2012). A novel approach to reducing number of sensing units for wearable gait analysis systems. *IEEE Transactions on Biomedical Engineering*, *60*(1), 72-77.

Salarian, A., Russmann, H., Vingerhoets, F. J., Dehollain, C., Blanc, Y., Burkhard, P. R., & Aminian, K. (2004). Gait assessment in Parkinson's disease: toward an ambulatory system for long-term monitoring. *IEEE transactions on biomedical engineering*, *51*(8), 1434-1443.

– **Step-by-Step**

Setting: Free-living

Monitoring:
The 3 monitoring periods have different characteristics

- t0 and t1: 24-h monitoring during in-patient rehabilitation
- t2: seven-day-monitoring (7x24 hours) taking place in the older adults' home environment.

System used to obtain the DMOs:

A thigh-worn inertial sensor including a tri-axial accelerometer (activPAL3™ (PAL Technologies Ltd., Glasgow, UK)).

References:

Kampe K, Kohler M, Albrecht D, Becker C, Hautzinger M, Lindemann U, Pfeiffer K. Hip and pelvic fracture patients with fear of falling: development and description of the "Step by Step" treatment protocol. Clin Rehabil. 2017 May;31(5):571-581. doi: 10.1177/0269215517691584. Epub 2017 Feb 17. PMID: 28415881.

– **Urban Training**

Setting: Free-living

Monitoring:
Patients were instructed to wear the sensor for a week.

System used to obtain the DMOs::
A Dynaport accelerometer (McRoberts BV, The Hague, The Netherlands), previously validated for COPD patients (see references). It was worn on the centre of lower back with an elastic strap. A valid measurement was defined as a minimum of 3 days with at least 8 h of wearing time within waking hours.

References – dataset

Arbillaga-Etxarri A, et al., Long-term efficacy and effectiveness of a behavioural and community-based exercise intervention (Urban Training) to increase physical activity in patients with COPD: a randomised controlled trial. Eur Respir J. 2018 Oct 18;52(4):1800063. doi: 10.1183/13993003.00063-2018. PMID: 30166322; PMCID: PMC6203405.

References – system

Rabinovich RA, Louvaris Z, Raste Y, *et al*. Validity of physical activity monitors during daily life in patients with COPD. *Eur Respir J* 2013; 42: 1205–15.
Van Remoortel H, Raste Y, Louvaris Z, *et al*. Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. *PLoS One* 2012; 7: e39198.