Mobilise-D Technical Implementation Manual: How to validate and measure digital mobility outcomes from a wearable device

Mobilise-D

Technical Implementation Manual

How to validate and measure digital mobility outcomes from a wearable device

imi innovative medicines initiative
efpia

Table of Contents

Referencing

Mobilise-D Technical Implementation Manual, no. 1, Mobilise-D Consortium, September 2025. Available at: https://doi.org/10.5281/zenodo.14901135.

DOI: 10.5281/zenodo.14901135

Contributing authors

Cameron Kirk [1], Paolo Tasca [2], Kirsty Scott [1], Arne Kuederle [3], Lisa Alcock [1,4], Anisoara Paraschiv-Ionescu [5], Alison Keogh [6], Arne Mueller [7], Claudia Mazzà [8], Andrea Cereatti [2], Silvia Del Din [1,4], Lynn Rochester [1,4,9] and the Mobilise-D Consortium

[1] Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle Upon Tyne UK. [2] Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy. [3] Machine Learning and Data Analytics Lab, Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. [4] National Institute for Health and Care Research (NIHR) Newcastle Biomedical Research Centre (BRC). [5] Laboratory of Movement Analysis and Measurement, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland. [6] University College Dublin, Dublin, Ireland. [7] Novartis Institutes of Biomedical Research, Novartis Pharma AG, Basel, Switzerland. [8] Department of Mechanical Engineering and Insigneo Institute for in Silico Medicine, The University of Sheffield, Sheffield, UK. [9] Newcastle University and the Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle Upon Tyne, UK.

List of abbreviations

CAD	Cadence estimation
CHF	Congestive heart failure
COPD	Chronic obstructive pulmonary disease
CI	Confidence interval
CVS	Clinical Validation Study
DMOs	Digital mobility outcomes
GNSS	Global navigation satellite system
GSD	Gait sequence detection
HA	Healthy adult
ICD	Initial contact detection
IMU	Inertial measurement unit
INDIP	INertial module with DIstance sensors and Pressure insoles. This reference system was specifically developed to enable real-world validation of DMOs in Mobilise-D.
L5	Fifth lumbar vertebra
MAE	Mean absolute error
MM	Move monitor
MRE	Mean relative error
MS	Multiple sclerosis
PD	Parkinson's disease
SL	Stride length
SP	Stereophotogrammetric system
TUG	Timed Up and Go
TVS	Technical Validation Study
WB	Walking bout

Key definitions

Inertial measurement unit: an electronic device that tracks acceleration, angular velocity and the earth's magnetic field using triaxial accelerometers, gyroscopes and magnetometers.

Metrological: Metrological analysis is the process of comparing a new method/system of measurement with a standard known measure. It involves assessing how well the new method/system measures quantities and ensuring it meets required standards of accuracy, precision and stability, determining if the calibration is sound and general quality control. By defining and validating processes for metrological analysis of a new measure, measurement standards can be met and transferred to prospective users.

Real-world setting: Real-world relates to the context in which walking takes place—that is free-living, unsupervised, uncontrolled and non-standardised. As such, it is unscripted as there are no instructions to the participant who does not need to interact with the wearable device(s). Real-world actions occur in non-simulated everyday situations in unconstrained environments with minimal thought of being tested. It is equivalent to actions at home or in the community over continuous periods of time. Synonymous terms are (environment of) daily living or relating to daily life. Real-world is distinct from laboratory-based, supervised (=fully controlled and observed), and semi-controlled (walking 'freely' but with supervision) tests. Real-world walking is different from scripted/instructed walking, which can take place in the home or lab (such as walking tests like the 4x10m test, 6-minute walk test (6MWT) and Timed Up and Go (TUG)).

Stride: A stride is equivalent to the gait cycle and every stride contains two steps. Stride duration is the interval between two successive initial contacts of the same foot.

Turns: The process of turning consists of decelerating the forward motion, rotating the body as a whole, and stepping out toward the new direction (Hase and Stein, 1999). Thus, turning includes a change of walking direction and change in angular orientation including a rotational movement of the body around the longitudinal axis. Turning, curvilinear walking, and straight walking involve different neuromotor strategies and need to be discriminated.

Walking bout (WB): a walking sequence containing at least two consecutive strides of both feet (e.g., R-L-R-L-R-L or L-R-L-R-L-R). The start and end of a walking bout are determined by a resting period or any other activity (non-walking period). The initial stride of a WB follows a non-walking period, and the final stride precedes the next non-walking period (Appendix 1).

For a full list of definitions please see (Kluge et al., 2021).

Navigating the manual

This manual has been designed to clearly summarise the steps taken and insight gathered during the Mobilise-D Technical Validation Study. Regardless of whether your aim is to leverage the publicly available algorithmic pipeline (https://github.com/mobilise-d/Mobgap) and/or dataset (https://zenodo.org/records/13899386) or to reflect the methodology and standards in new research, we hope this manual will be of benefit. To facilitate easier navigation regardless of aim, signposting at the start of each section to relevant content and references has been added.

1. Publishable summary

This manual is for anyone (i.e., developers, researchers, clinicians) that intends to validate and measure digital mobility outcomes from a single wearable device integrating a three-axial accelerometer and gyroscope.

The framework described in this manual is based upon the lessons learnt by the Mobilise-D consortium during the state-of-the-art Technical Validation Study (TVS) (Mazzà et al., 2021). The primary objective of the TVS was to explore the validity of a single wearable device attached to the lower back and the selected algorithms to estimate a select list of digital mobility outcomes in laboratory and real-world settings in different patient populations.

A device agnostic approach ensured that a method was developed that was generalisable to any device with the required technical requirements and attachment modalities. To date, this is the largest and most harmonised technical validation study on mobility and represents the efforts of a large group of multi-disciplinary experts from more than 30 academic, technical, and pharmaceutical industry organisations throughout the globe.

This manual draws from the experiences and lessons learnt of the state-of-the-art validation procedure to provide a comprehensive overview of the most valid methods of data capture, optimal approaches to technical validation and specific real-world scenarios where digital mobility outcomes can be quantified with the highest accuracy and precision.

The aim of this Technical Implementation Manual is to provide a guidance reference to encourage the wide-spread and consistent adoption of proven methods required for the valid measurement of DMOs across a range of health conditions and a much-needed harmonisation of best practice.

2. Introduction

Measuring changes in real-world walking is an important aspect of health to patients, due to associations with physical, emotional, and social experiences of walking (Delgado-Ortiz et al., 2023).

In this regard, digital health technologies (DHTs), including wearable devices, offer an objective, low-cost and ecologically valid approach of continuously monitoring real-world walking activity, through quantification of digital mobility outcomes (DMOs) (Rochester et al., 2020). Tools to facilitate remote visits and monitoring have never been more necessary, and DMOs could be deployed to remotely monitor aspects of gait characteristics and real-world mobility behaviour which are not captured in routine clinic-based assessments. Such information would allow human movement specialists and clinicians to target and manage aspects of mobility disability that are of utmost importance to patients (Deane et al., 2014; Delgado-Ortiz et al., 2023; Port et al., 2021).

The ability to objectively evaluate patients remotely has significant advantages for both clinical research and clinical management. DMOs could be remotely assessed, offering an inclusive and complete evaluation of mobility before the patient has even arrived for a face-to-face visit in the clinic. This reduces the length of in-person clinical assessments, allowing more visits, increasing the amount of time available to clinicians to perform other aspects of the visit, and augment the information about the health status of their patients.

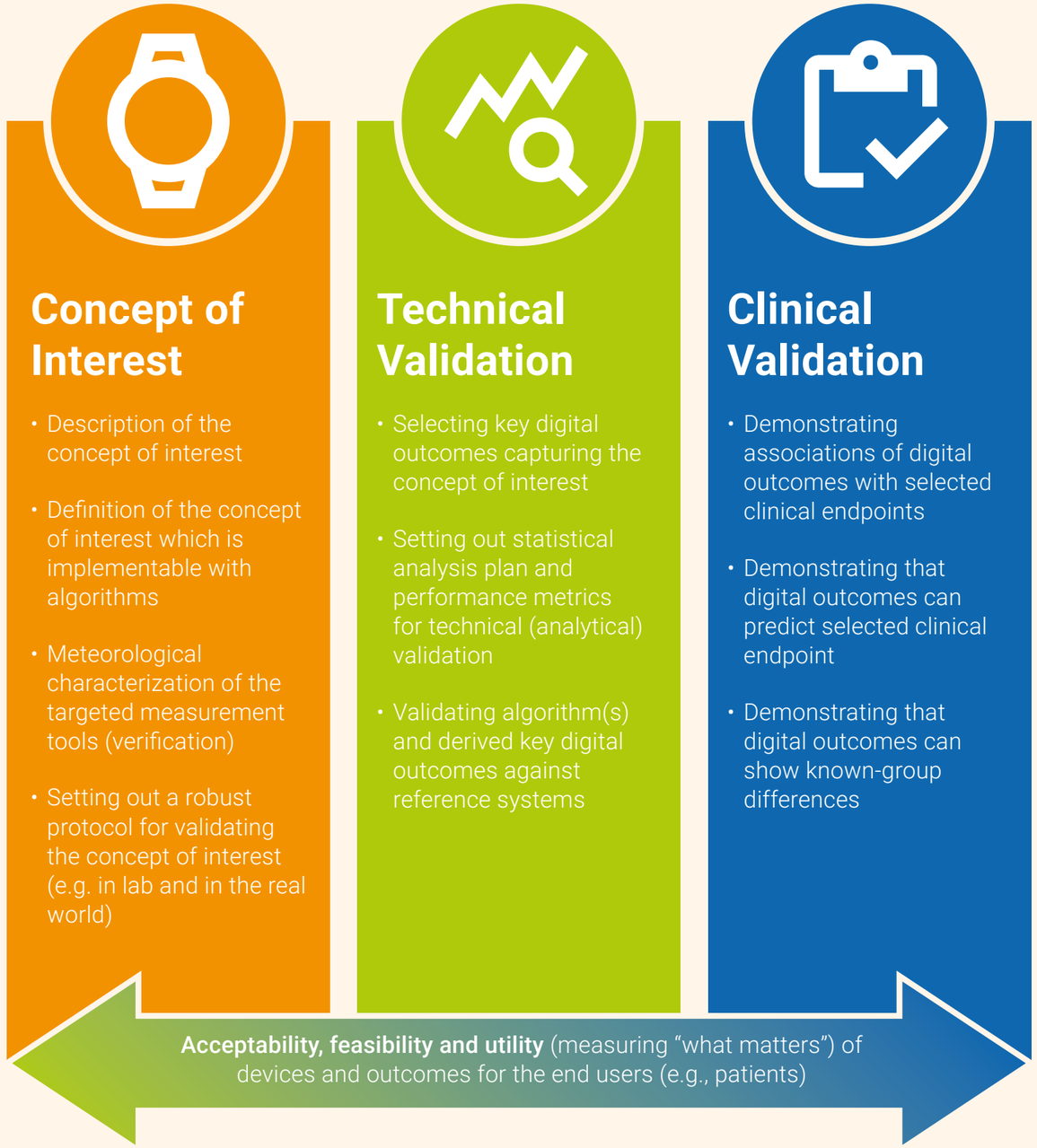
As described by Rochester et al., 2020, adoption of DHTs in clinical trials and ultimately health care poses significant challenges because of the multiple types of expertise (e.g., technical, clinical and regulatory) and phases (including both technical and clinical validation) required to deliver the evidence needed for acceptance and adoption. For DHTs and DMOs to be adopted, considerable multidisciplinary efforts are needed. Important contributions have been made to bring scientific experts and technology users onto the same page, e.g., harmonisation of terms for biomarkers and endpoint guidelines (Cagney et al., 2018) for the development of novel digital endpoints (Babrak et al., 2019; Coran et al., 2019), frameworks for validation and primers (Goldsack et al., 2020), and guidelines for interactions with regulators (Cerreta et al., 2020).

However, the challenge for anyone aiming to deliver digital tools approved for use in research and ultimately clinical practice is to bring these recommendations together in one seamless roadmap from inception to adoption. The Mobilise-D consortium aimed to learn from and adopt these best practices whilst innovating in this space – leading with the development of standards, disease and hardware agnostic approaches, open access tools and databases amongst other things. Technical and clinical validation, evaluation of acceptability and user engagement were undertaken within a multidisciplinary framework of collaboration and transparency. Figure 1 sets out the critical steps towards adoption (Alcock et al., 2024). Throughout, embedded user engagement, external scientific advisory input, and patient and public involvement enriched the work of the Mobilise-D consortium.

Figure 1. Framework and recommendations for the use of wearable devices and paired algorithms to assess gait analysis. This includes defining the concept of interest, technical validation of algorithms and clinical validation of digital mobility outcomes, and assessing the acceptability, feasibility and utility of devices and outcomes by the end users. This figure was adapted from a figure published in a book chapter, for more information see Alcock et al., 2024.

This manual focuses on two of the key challenges within the roadmap: (i) to define the concept of interest, in our case mobility, which is quantified with algorithms and measurement devices and (ii) technical validation of the identified computational pipeline and device combination. This is based around the knowledge and insights gained from the Mobilise-D TVS, focused upon the verification, analytical validation, and user acceptability of a single device approach to continuously assess and quantify a range of real-world DMOs in diverse clinical populations. This manual provides a detailed use case for real-world mobility assessment, lessons learnt and guidelines to enable others to apply the methods from Mobilise-D to other technologies and similar and different concepts of interest.

The TVS was undertaken to address the limitations of existing approaches to technical validation. The study has identified the most thorough approach to technically validating DMO measurements to the highest methodological standard. This manual will provide a framework which could be adopted to ensure the most accurate and reliable approach to passively and continuously measure DMOs from a single device that is acceptable to the user. This manual is organised in three macro sections: verification, usability and analytical validation adapted from the V3 framework (Bakker et al., 2025; Goldsack et al., 2020).



3. Overall recommendations

This section provides an overview of recommendations for anyone seeking to measure and validate DMOs from a single wearable device. These recommendations, detailed in Table 1, are based upon the learnings from the Mobilise-D TVS and are intended as a guide that may be adapted to a specific case rather than being a check list.

Table 1. Recommendations to acquire technically valid DMOs based upon lessons learnt from the Mobilise-D TVS.

Technical validation study protocol: Mazzà et al., 2021 Scott et al., 2022 **Validation of individual algorithms:** Micó-Amigo et al., 2023 **Complete pipeline validation:** Kirk et al., 2024	**Minimum requirements of measurement device** Sampling frequency, range and resolution: Accelerometer: 100Hz, ≥ ± 8g, 1 mg (at ± 8g) Gyroscope: 100Hz, ≥ ± 2000 dps, 70 mdps (at ± 2000 dps) Battery duration: Must be able to continuously record for a minimum of seven days. ***Errors, short static acquisition:*** • Norm of acceleration ideally in range from 0.98 to 1.02 g, and the noise (standard deviation) ideally in range from 1.40 to 2.52 mg. • Mean and standard deviation from gravity between 0.9 to 3.1 degrees and 0.4 to 1.2 degrees. • Bias (mean relative error) of the angular velocity measured from devices between 0.73 to 0.90 degrees/second and noise between 0.05 to 0.16 degrees/second. ***Errors, long static acquisition:*** Bias instability (root Allan variance) between 9.63 to 16.28 degrees per hour within each day.
Supporting references: **Technical validation study protocol:** Mazzà et al., 2021	**Device position and attachment** ***Number of devices:*** a single wearable device is recommended for continuous recording over a minimum of seven days and for repeated research/clinical assessments. ***Device position:*** The lower back at the level of the fifth lumbar vertebra (L5). ***Attachment modality:*** The wearable device can be either body worn (i.e., a belt), or body attached to the skin at the L5 location. The method is dependent upon the model of device and person's preference. Mobilise-D recommends that the same device and attachment method should be used for the entire study duration, to reduce bias in DMO calculation between different methods. The device must be attached symmetrically (i.e., on the lumbar vertebra). If the device is off-centre or shifts, the information recorded by the accelerometer (peaks related to foot initial contacts) will be mostly related to strides rather than steps and this may introduce errors in the processing pipeline. Attachment modalities that reduce the chance of attaching the sensor incorrectly or having it shift during wear is preferred.
Supporting references: **Data standardization procedure:** Palmerini et al., 2023	**Data standards** ***Data format:*** Raw data recorded by the adopted wearable device are recommended to be in the most universal data format: .csv (comma-separated values) file, however other formats may also be accepted (cwa, omx etc.,), dependent upon device. Data should always be stored with all relevant meta data. ***Info needed for algorithm processing:*** • Type of test (Laboratory or real-world). Required for cleaning thresholds. • Condition of clinical cohort. Required for validation analysis. • Height of wearable device. Required for calculation of stride length.

Supporting references: **Usability validation:** Keogh, et al. 2023	**Usability assessment** Ask and report findings from a subset of patients concerning their views and experiences of wearing the device following a period of continuous real-world monitoring using questions linked to the usability and degree of comfort of the device. An example of questions can be found in Appendix 5. The measurement approach should maintain high usability standards to minimise patient burden, enhance compliance, and improve data quality. When applicable, existing data or evidence on usability can be used to inform assumptions. However, if no prior data is available for a specific use case, do not assume usability findings are transferable across different patient populations. Instead, new usability data should be gathered before proceeding.
Validation of individual algorithms: Micó-Amigo et al., 2023 **Complete pipeline validation:** Kirk et al., 2024	**Assessment setting** In general, it's recommended to include both real-world and laboratory assessments within measurement protocols. However, in certain cases, a range of laboratory tasks can represent the same challenges to algorithm performance as real-world recordings.
Validation of individual algorithms: Micó-Amigo et al., 2023 **Complete pipeline validation:** Kirk et al., 2024	**Cohort** Each patient population presents a unique challenge to the algorithms for DMO estimation. Therefore, ideally, users should undertake a technical validation of DMOs upon patients that were not included in the Mobilise-D TVS.

Validation of individual algorithms: Micó-Amigo et al., 2023 **Complete pipeline validation:** Kirk et al., 2024	**Walking bout duration** A sample size calculation should be performed to evaluate the minimum number of WBs needed to ensure reliability and validity for real-world gait monitoring in a single patient cohort. It is recommended that analysis only be completed on WBs longer than 10s. As such, the removal of WBs under 10s should be factored into any sample size calculation. As an example, the sample size calculated in the Mobilise-D TVS was 401 per patient population of interest. .
Validation of individual algorithms: Micó-Amigo et al., 2023 **Complete pipeline validation:** Kirk et al., 2024	**Walking speed** Errors in DMO estimation were largest at slow walking speeds and/or the usage of walking aids. Therefore, the choice of an algorithm for DMO estimation should consider its sensitivity to walking speed/impairment, given its confounding effect on gait analysis, and population of interest.
Validation of individual algorithms: Micó-Amigo et al., 2023 **Complete pipeline validation:** Kirk et al., 2024	**Contexts** When possible, it is recommended to adopt complementary technologies (e.g. smartphone) to enable collection of real-world contextual information that could further help to understand the most challenging real-world scenarios for DMO measurement. In the TVS smartphones capture GPS information needed to estimate contextual location. See 6.6.

4. Selection of measurement devices

The DMOs are derived from the electronic signals recorded by the sensors integrated in an inertial measurement unit (IMU) or referred to in this manual as the ‘device’. First, the device must be selected.

While a broad range of devices exist, they must meet the minimum technical requirements outlined below. This ensures sufficient accuracy, inclusion of essential data, time resolution, and compatibility of the data collected with the Mobilise-D analytical pipeline while meeting usability preferences of the studied patient populations. The following section details these requirements, which have been developed from extensive testing and knowledge from technical experts within the Mobilise-D consortium.

Using the Mobilise-D TVS dataset?

Prior to exploring further validation with the Mobilise-D TVS dataset (https://doi.org/10.5281/zenodo.13899385), please be mindful of the device specification and the metrological characterisation completed on all lower back devices. This information can be found in **Table 2.**

Collecting new data to use the Mobilise-D pipeline?

To ensure the technical validity of the algorithms used in the Mobilise-D pipeline, it is essential that device specification, device location and method of attachment, data standard, and the defined metrological characterisation satisfy the defined requirements. A brief overview of these requirements can be found in **Table 2,** with further detailed descriptions in the subsequent sections of this chapter.

4.1 Sensor requirements

To implement the Mobilise-D pipeline, researchers can utilise any measurement device in their study, if it satisfies the minimum technical requirements outlined in Table 1. A single wearable device is recommended for continuous recording over a minimum of seven days and for repeated research/clinical assessments, to reduce patient burden and increase compliance. The wearable IMU should include a tri-axial accelerometer that captures accelerations (gravitational acceleration) and a triaxial gyroscope, that captures angular velocities.

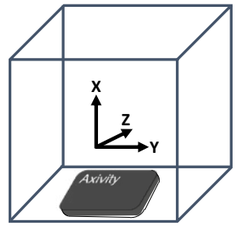
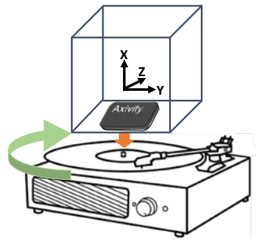
The device may also include other types of sensors such as magnetometers, barometers, and temperature sensors which could add further validity and context. Mobilise-D included the DynaPort Move Monitor (MM) + (McRoberts, the Hague, Netherlands) and the Axivity AX6 (Axivity Ltd, Newcastle upon Tyne, UK) as the selected measurement devices. Calibration of the devices before their use should be performed to ensure that the devices meet the performance requirements to capture quality data.

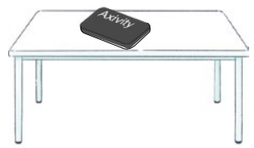
4.2 Technical characterisation of the device (metrological characterisation)

It is vital for researchers to ensure that each wearable device used to collect data in their study be completely characterized. To start, each device should be tested across multiple checks, to ensure data integrity critical for clinical and research applications. Failure to do so, could result in loss of data through battery failure, or bias and noise in IMU signals which impacts upon reliability, accuracy, and precision of the device. The aim of the metrological characterization is to verify that the wearable devices are suitable for data collection and establishes reference values for equivalent devices. An overview of this process is given below (Table 3), Mazzà et al., 2021.

We recommend all devices undergo complete characterisation prior to commencement of a study followed by a brief characterisation before each data collection use.

Table 3. Summary of the characterisation protocol tests completed on the wearable device (modified from Mazzà et al., 2021).

Acquisition:	Setup	Description	To note	Output
3D static		Perspex cube. The cube is repeatedly reorientated to **align each axis of the devices LCS** with gravity. The device/cube is left static for ~60sec for each alignment.	**Each side of the cube should be in contact with table for at least 60s.**	• Short noise evaluation (3D) • Static bias (3D) • Angular deviation from vertical (3D)
3D dynamic		The device is securely attached to the record player and repeatedly reorientated to align each axis of the devices LCS with gravity. The secured device is left to spin at 33.3RPM for ~60sec for each alignment. An example method of attachment is utilising a Perspex cube securely fitted over the spindle of the record player.	**Each side of the cube should be in contact with the moving record player for at least 60s.**	• Angular velocity error • Integral angle

Acquisition:	Setup	Description	To note	Output
Long static		The device is placed level on a table and left to record a static test for 7-days.	**It is important that any external noise (i.e., knocking or vibration of the equipment and table) be kept to a minimum during these recordings.**	• Long noise evaluation (norm) • Angular deviation from vertical (one axis) • Battery check
Quick static		The device is placed level on a table and left to record a static test for ~60sec.	**It is important that any external noise (i.e., knocking or vibration of the equipment and table) be kept to a minimum during these recordings.**	• Short noise evaluation (norm) • Static bias (norm) • Angular deviation from vertical (one axis)

However, if a study has a large sample size, it may be impractical to characterize every unit. As a practical measure, we recommend performing a complete characterisation on a random subset for each batch of devices (Figure 1). For example, in the Mobilise-D project, 35 Axivity AX6 devices were randomly tested from an initial batch of 100 devices delivered. All wearable devices tested during the Mobilise-D project satisfied the required performance thresholds. Based on the above analysis, we recommend following the same procedure and testing 20% of devices received. If the subset of devices tested satisfies the complete device characterisation, the whole batch can be considered suitable for use. However, to ensure integrity in the measure and monitor any deterioration in the device's performance (a common occurrence in repetitive and prolonged use) it is recommended that the quick characterisation is performed before each use.

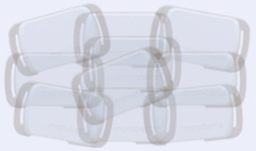
	Complete device characterisation	Quick characterisation
	Before deployment to the clinical centres. 7+days - 5 recordings ✓ Short time noise evaluation ✓ Long time noise evaluation ✓ Measurement norm during static ✓ Angular deviation from vertical ✓ Static bias ✓ Angular velocity error ✓ Integral angle error ✓ Battery test	Before data acquisition. 3 minutes- 1 recordings ✓ Short time noise evaluation ✓ Measurement norm during static ✓ Static bias
TVS		
CVS		

Figure 1.
The Wearable Device characterisation protocol adopted during the TVS and the CVS.

An example of the results from a complete characterisation are reported in Appendix 2, Appendix 3, and Appendix 4. When assessing devices based on short static acquisitions both the bias and the noise of the accelerometer and gyroscope norms of the Axivity AX6 were either statistically similar or lower than the DynaPort MM+.

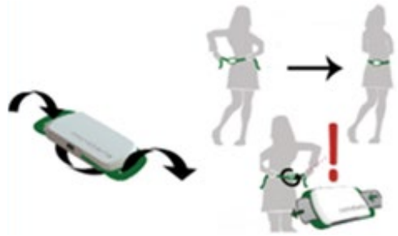
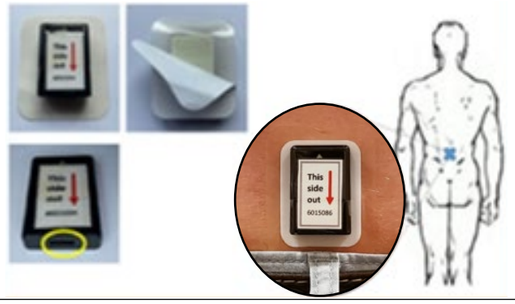
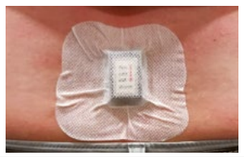
4.3 Device position and attachment

To accurately quantify DMOs using the Mobilise-D algorithms, the device must be worn on the lower back at the level of the fifth lumbar vertebra (L5). This position, close to the body's centre of mass, allows the measurement of gait events, and the overall movement pattern of the right and left limbs. Based upon previous research (Keogh et al., 2023) and demonstrated in the Mobilise-D Clinical Validation Study with four large cohorts of people with multiple sclerosis, Parkinson's disease, COPD, and hip fracture, participants stated that they found the low back a comfortable location to wear the device for the seven days.

The wearable device can be worn either on the body by a belt around the waist or attached directly to the skin at the level of L5. The method of attachment is dependent upon the device and model. An example of a body worn device is the DynaPort MM+, which can be worn via a waist-belt that is fed through specialised rubber loop straps, that are built into the device. In contrast, the Axivity AX6, is attached directly to the skin. In the Mobilise-D Clinical Validation Study, a custom hydrocolloid adhesive patch was designed specifically for long wear time periods (Design: Dr Lisa Alcock Newcastle University, Manufacturer: Linxens; formerly Nile AB). The device is secured in place with a plaster patch (Hypafix, BSN medical GmbH, Germany). An overview of both attachment modalities can be viewed in Figure 2.

Figure 2.
Tools, fixation instructions and the final setup of wearable devices included in the Technical Validation Study.

The belt should be securely fitted and regularly inspected between assessments to ensure the device maintains the appropriate vertical alignment and location over the fifth lumbar vertebrae. A usability assessment should be conducted as part of a feasibility evaluation in a new patient population of interest. It is crucial that the wearable device is positioned symmetrically and vertically over the fifth lumbar vertebra in the correct orientation. This is particularly important as certain orientation errors cannot be corrected as part of data processing, making proper placement essential to obtain valid, quality data.

	DynaPort MM+. *Body Worn*	Axivity AX6. *Body Attached*
Tools		
Fixation instructions		
Final setup		

Recommendations for the body attached solution:

The body attachment solution for the Axivity as detailed above, includes two components that are both provided by Nile (Ängelholm, Sweden): a medical grade adhesive and a rounded plaster patch (Figure 2). The adhesive patch is used to attach the Axivity AX6 to the skin while the rounded plaster goes over the Axivity AX6 for i) securing the wearable device on the skin creating a seal around the Axivity AX6, ii) help with comfort and iii) prevent excess slippage. Two large pull tabs on the adhesive patch have been included for ease of use and to help with compliance (especially for participants with dexterity or coordination issues).

Should the body-attachment method be adopted, it is advised that participants are provided with 2/3 adhesive and plaster patches to cover a period of at least seven days. It would be highly normal and expected that the skin under the patch will be slightly red after extended wear time. The skin should return to a normal appearance within several days. Participants should be instructed to contact the research staff with any questions. For preservation of resources, all unused applications can be returned to the researchers within the pre-paid envelope that will be used for posting the device back.

4.4 Learnings from Mobilise-D: Using a different wearable device.

A core objective throughout the Mobilise-D technical validation period was to produce an algorithmic pipeline that could sustain wide adoption past the lifespan of the project. As such, agreement in the technical validity in estimating DMOs when using different wearable devices (i.e., electronics and/or mode of attachment) needed to be challenged and recommendations made. Initial evaluation considered different wearable device combinations, which varied in one of three ways to the original device used in the technical validation (Dynaport MM+):

1) Same wearable device and a different mode of attachment (i.e., body worn or body attached)

2) Different electronics (i.e., different wearable devices that met the device requirements in Table 2) and the same mode of attachment.

3) Different electronics (i.e., different wearable devices that met the device requirements in Table 2) and a different mode of attachment (i.e., body worn or body attached)

Data collection for these different wearable device combinations occurred within a subsample of the technical validation study participants. This reduced the need for further recruitment, reduced participant burden, and ensured comparable data to the original device used and the reference systems (described in Chapter 6).

Preliminary analysis has shown excellent agreement between the walking speed estimated using different wearable device combinations for walking bouts (WB) lasting longer than 10 s. This initial assessment concluded that algorithm performance was not affected by any of the wearable device combinations evaluated. However, to reduce potential variability of the measurement and maintain methodological rigor, it is recommended that the same device and attachment method be used for all assessments on an individual during the entire study duration.

5. Usability validation and acceptability of the measurement device and protocol

To promote the use of wearable devices, it is important that they are usable by and offer value to all users. In this respect, researchers should explore the acceptability of the wearable device adopted for data collection in their specific population of interest.

Using the Mobilise-D TVS dataset?

In Mobilise-D, acceptability was assessed in 40% (n = 48) of participants, which is considered acceptable in most populations (Boddy, 2016; Robinson, 2014; Sim et al., 2018). The qualitative approach to measure acceptability during Mobilise-D is outlined in **Figure 3** and results are presented and detailed in **Keogh et al., 2023**.

Collecting new data to use the Mobilise-D pipeline?

We recommend asking and reporting the findings from a subset of patients about their views and experiences of wearing the device following a period of continuous real-world assessment using questions linked to the usability and degree of comfort of the device. The following chapter will briefly summarise the protocol and data analysis used in the Mobilise-D TVS, with example questionnaires available in Appendix 5 and Appendix 6.

5.1 How can acceptability be assessed?

This section summarises the qualitative approach used to measure the acceptability of the wearable technology during the Mobilise-D TVS. Further information on the outcomes of this study can be found in Keogh et al., 2023.

The qualitative approach consisted of three different methodologies, with the questions from **Part 1** and **2** available in Appendix 5 and interview structure and questions for **Part 3** available in Appendix 5

Part 1: The comfort rating scale (CRS) (Knight et al., 2002)

The CRS is a six-item 21-point Likert scale questionnaire which was developed and tested for face validity and reliability in adult populations. People are asked to report whether they have low agreement (i.e., 0) to high agreement (i.e., 20), for each of the six questions covering topics of emotion, attachment, harm, perceived change, movement, and anxiety.

Part 2: The Rabinovich questionnaire (Rabinovich et al., 2013)

This is a 12-item questionnaire on a 5-point ordinal scale. Participants were asked to note which statement reflected their experiences best, with questions covering topics such as whether the device is comfortable to wear at night, whether technical problems were experienced and how easy the device was to use. Responses per question were then transformed into a numerical scale from '1' to '5' whereby '1' signified low acceptance and '5' high acceptance. Section B simply asks respondents to verbally give the device a single score between 0 and 100 and asks them to comment on what features of the device they both liked and didn't like.

Part 3: Semi-structured interviews

In addition, Mobilise-D researchers interviewed study participants. The structure of interview used in the TVS was split into two parts. The first section aimed to further explore the participants' experiences of wearing the McRoberts Dynaport MM+ device. Previous literature has suggested that usefulness, comfort and ease of use are all critical factors (Mercer et al., 2016) thus, these were selected as the categories for the device assessment.

5.2 Insights from Mobilise-D: Analysis of acceptability data

To assess acceptability of the device, qualitative and quantitative data collected among users were analysed. In the TVS, Shapiro-Wilk tests determined data were not normally distributed ($p < 0.05$). Thus, non-parametric descriptive statistics were conducted for the questionnaires (median [inter-quartile range (IQR)], minimum-maximum) and reported overall and per patient cohort. A Kruskall-Wallis test was used to determine differences between cohorts for each questionnaire.

Interviews can be analysed using a content analysis approach. In the TVS, the first section of the interviews was categorized deductively based on the previously listed sections of comfort, ease of use, and perceived usefulness, in line with the Technology Acceptance Model (Davis and Davis, 1989). Perceived usefulness was defined as 'the degree to which a person believes that using a device would enhance their health'. Perceived use was defined as 'the degree to which a person believes that using a device was free of effort'. Comfort was defined as 'a state of physical ease and freedom from pain or constraint'.

The second part of the interview was analysed inductively to explore participants' current healthcare experiences and their use of and opinions towards technology in the management of their health and condition. Codes that related to these concepts were identified within the text and grouped together into meaningful themes. Specific quotations, which were deemed to represent the most important aspects of participants' experiences, were selected for inclusion in the analysis by the researcher.

Given the primary aim of this study, data saturation was considered in relation to the acceptability of the device. this part of the interview was considered exploratory in nature and given the purposive method of sampling and the broad, complex topic under consideration; we cannot be certain that saturation was reached for this component across all cohorts. Thus, although the sample size used in this study falls within the range that saturation is typically found in studies (Hennink and Kaiser, 2022), the themes related to the second part of the interviews should be interpreted with caution. Nonetheless, saturation was reached for the acceptability component of the interviews as no new information was seen after three interviews across the cohorts for each of the explored themes.

6. Analytical validation - the need for a robust protocol

Validating the accuracy, relability and robustness of derived DMOs against a gold standard is essential to ensure rigour in the outcomes and possible inferences made. The acceptability and outcomes derived from the device and software must be validated to the specific population of interest and context in which they will be measured (e.g., the real-world) (Maetzler et al., 2016).

This requires an extensive and robust protocol deployed in the population of interest and should aim to include the range and variation in walking speeds, environment, postural transitions, and manoeuvres, such as turns, that would be expected in the context of real-world walking. This approach ensures that the DMOs assessed provide a technically valid measurement of real-world walking behaviour and identify the effects of confounding factors which affect the levels of accuracy and reliability. Mobilise-D developed a comprehensive and multi-stage protocol for the TVS to evaluate these factors and included a variety of tests conducted in laboratory and real-world settings (Mazzà et al., 2021). Figure 3 summarises the TVS protocol phases including in-lab and 2.5-hour free-living assessments and measurement tools that have been utilised for the validation of DMOs.

Using the Mobilise-D TVS dataset?

A substantial number of publications and resources have been produced to outline the development, validation and requirments produced from the Technical Validation Study protocol (See Table 1 for complete references). Please be mindful, if using the gold standard data as comparison that citation is made to the original paper where the method was validated. For the stereophotogrammetric system, please reference Bonci et al., 2022. For the multi-sensor system (INDIP), please reference Salis et al., 2023.

Collecting new data to use the Mobilise-D pipeline?

A substantial number of publications and resources have been produced to outline the development, validation and requirments produced from the Technical Validation Study protocol (See Table 1 for complete references). Please be mindful, if using the gold standard data as comparison that citation is made to the original paper where the method was validated. For the stereophotogrammetric system, please reference Bonci et al., 2022. For the multi-sensor system (INDIP), please reference Salis et al., 2023.

Mobilise-D Technical Validation Protocol

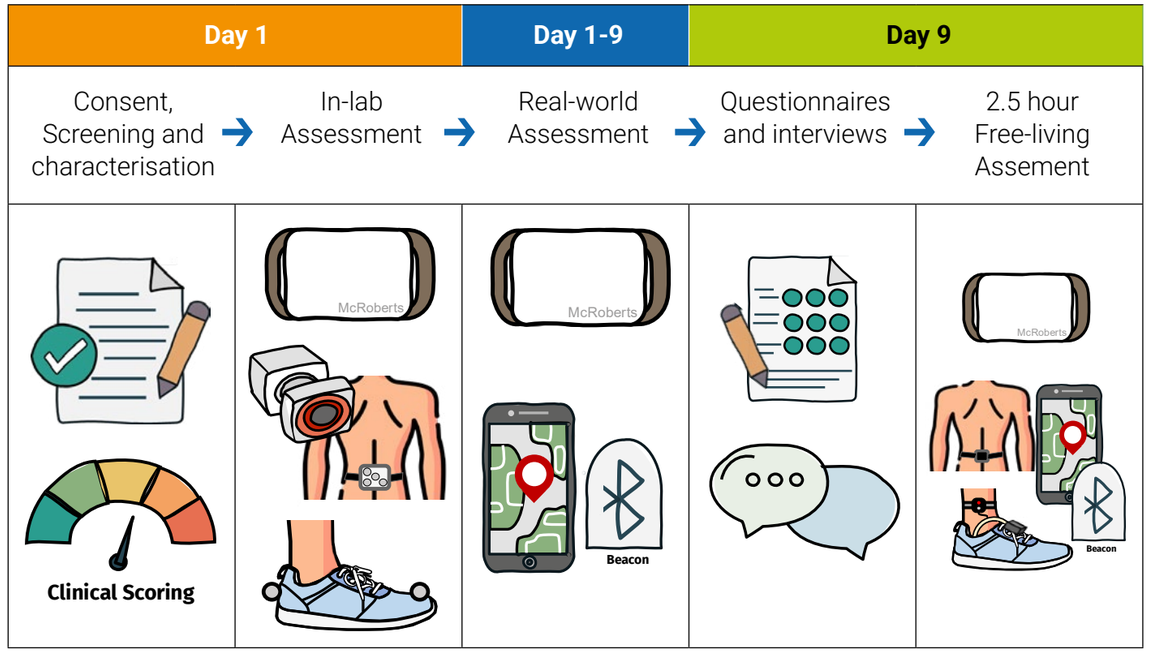


Figure 3.
Schematic of phases of the Mobilise-D TVS protocol, including in-lab and 2.5-hour free-living assessments whose data have been used for the validation of the Mobilise-D digital mobility outcomes.

6.1 Study population and study size

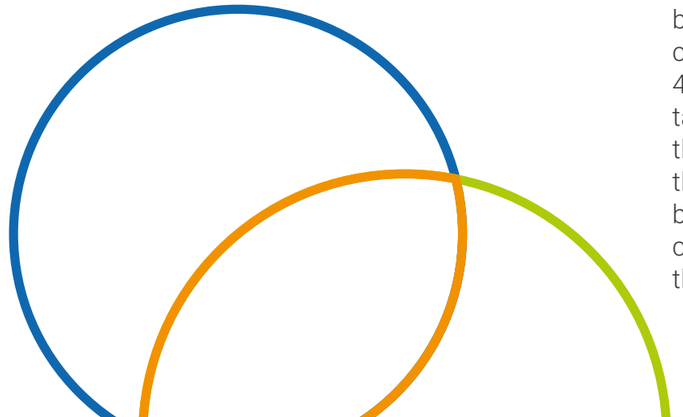
To obtain robust statistical analysis, sample size evaluation for technical validation should be based on a minimum number of identified walking bouts (WBs) (including in-lab and 2.5-hour assessments) per cohort, supplemented by a sufficient number of participants (Table 4).

To establish whether the 2.5 hours of unsupervised recording was sufficient to reach an adequate sample size from a statistical point of view, a preliminary statistical power analysis was conducted to define the minimum number of WBs for each cohort required to validate the estimate of walking speed. In order to obtain a confidence interval smaller than 0.1 when comparing two different instruments with an $\alpha = 0.05$ and a power $ß=0.9$, an analysis was performed in Stata 16.1 (Stata Corp LP; College Station, Texas, USA; command line: sampicc 0.7 2, alpha (0.05) power (0.9) w (0.1) ci), which showed that the minimum number of WBs needed in each cohort for an Intraclass Correlation Coefficient (ICC) ≥ 0.7 would be 401 walking bouts per population of interest (Scott et al., 2022).

6.2 In-lab assessment

A laboratory-based assessment should be utilised to quantify validity and consistency within and between groups and different types of walking tasks under ideal controlled conditions. The protocol should specifically include structured and task-based mobility activities and a simulated daily activity session, mimicking habitual movements. To satisfy the abovementioned requirements, the following tasks were performed in a laboratory setting during the Mobilise-D TVS:

- Straight Line Walking: Participants were asked to walk a predefined path of 5 m from a standing start to a standing end. This task was performed twice at three self-selected speeds: comfortable, slow, and fast (Figure 4.1a).
- Timed Up and Go (TUG): At a comfortable speed, participants were asked to rise from the chair and walk 3 m to the cone, make a 180° left hand turn around the cone, walk back to the chair and sit down (Figure 4.1b). The TUG is a validated, standard clinical assessment of mobility for patients with varying levels of walking impairment. Although the task setup conforms to the published guidelines (Podsiadlo & Richardson, 1991a), the primary reason for including this task in the TVS protocol was the sit-to-walk/walk-to-sit posture and turn movements.
- L-Test: At a comfortable speed, participants were asked to rise from the chair and walk 4 m to the first cone, make a 90° turn to the left around the cone and walk straight to the second cone 3 m away, turn 180° to the left around the cone and walk straight before making a final 90° turn to the right of the first cone and walking back to sit in the chair (Figure 4.1c). Like the TUG, this test has previously been validated as a clinical assessment in patient cohorts (Deathe and Miller, 2005) however, the main purpose of including this test in the TVS protocol was the variation of turns included in a short walkway.
- Surface Test: Participants were asked to walk at a comfortable speed around the circuit by walking around the cones and stepping over the carpeted mat completing the circuit twice. Participants started and stopped the tasks in a standing position (Figure 4.1d).
- Hallway Test: From a standing start, participants were asked to walk at a comfortable speed along the predefined walkway stepping up and down from the step. At the end of the walkway, participants completed a 180° turn and walked back along the walkway (again stepping up and down the step) before coming to a stop at the end of the walkway (Figure 4.1e).
- Simulated Daily Activities: Participants began sitting in a chair (Figure 4.1f) and complete a series of tasks defined in Figure 4.1g. while moving around the room. The tasks were split into separate steps, with the next set of instructions only given to the participant after the previous step had been completed. All steps for this task were completed at the preferred walking speed of the participant.



Recommendation for the straight walking task

Based on the protocol validation in Scott et al., 2022, it is advised that the walkway be extended to the full length of the capture volume or room to accommodate longer strides. Further detail can be found in the abovementioned paper.

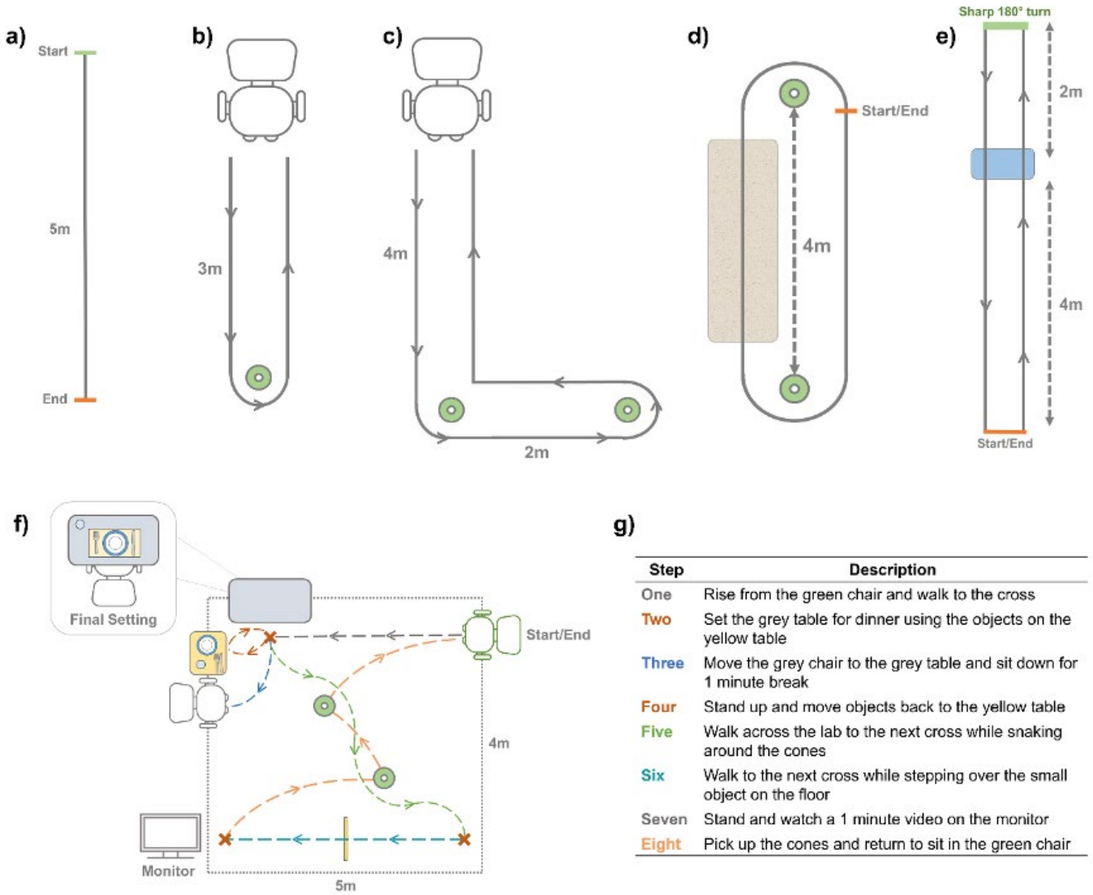


Figure 4.
Schematics of the seven structured tasks: a) straight walking, b) TUG, c) L-test, d) Surface test and e) Hallway test completed in the laboratory-based assessment, as well as a schematic of the simulated daily activities f) and description of the eight steps performed during this task g).

6.3 2.5-hour real-world assessment

The performance of DMOs should be assessed in real-world settings with unsupervised and unstructured activities. In the TVS, to account for the battery life constraint of the reference system and to be able to collect enough walking bouts to validate the DMOs (Scott et al., 2022) (Table 5), real-world data were captured across one session of 2.5 hours. Following the laboratory assessment, researchers visited the participants' homes and prepared the combination of devices for data capture (Table 5). The participants were given a guide outlining different tasks to perform during the 2.5 hours that would aid validation of the DMOs in specific contexts not feasible to complete in the laboratory setting (Figure 5). Participants chose what activities to complete.

Figure 5.
Excerpt from the information sheet given during the home-based validation outlining the suggested tasks to complete.

2.5-hour Home Monitoring

During the 2.5 hours, we would like you to try and complete the tasks on this list. For your safety please do not attempt to complete tasks that you are not comfortable or confident with.

GENERAL TASKS

- Rise from a chair and walk to another room ☐
- Walk to the kitchen and make a drink ☐
- Walk up and down a set of stairs (if possible) ☐
- Walk outside (if possible, for a minimum of 2 minutes) ☐
- If walking outside, walk up and down an inclined path ☐

6.4 Continuous real-world assessment

DMOs should be assessed over a period of continuous recording, in this case seven days. A seven-day duration was chosen because it is thought to be a sufficient amount of time to accurately represent a person's real-world behaviour. The participant was equipped with the wearable device set to collect data for 8 days, to ensure a full 7-days of data capture. Following this, participants returned the device to the site using a pre-paid envelope along with any spare attachments. Participants were instructed to carry on their normal, daily activities across the 7-days. In the absence of a suitable reference system capable of continuous recording over seven days, this dataset can be included in exploratory analyses. Collecting continuous data over less than seven days can also be applied to characterise DMOs when exploring a new population.

6.5 Reference systems

To determine the accuracy and capability of a new measurement tool, comparison to the best available method of measure is required. During the technical validation process, the best available measures (i.e., reference systems) of the DMOs are needed. Due to the protocol structure consisting of multiple stages and context changes, appropriate reference systems for each measurement setting are required (Table 5, Figure 3).

Table 5.
Summary of the experimental protocol and reference systems used to validate the algorithms in the laboratory and real world.

Context of assessment	Reference Systems	Tested device	Mobility Tasks
Laboratory	Stereophotogrammetry (SP)	DynaPort MM+ INDIP	Structured mobility tasks and daily living activities
Real world (2.5 hours)	INDIP1 Mobile Phone with Aeqora App Beacon**	DynaPort MM+	Unsupervised real-world activities (including predetermined tasks)
Real-world (7 days)	Mobile Phone with Aeqora App Beacon**	DynaPort MM+	Unsupervised daily living

**Denotes reference system was not included in the Mobilise-D TVS due to technological issues (see 6.7).

In-lab assessment

In the TVS, a multi-camera stereophotogrammetric system (SP) was used in the laboratory setting. Although SP systems are a robust measure and traditionally considered the gold standard within the field, accuracy may vary when considering changes in system type (e.g., manufacturer, number and model of cameras, and laboratory environment), calibration effectiveness and experience of the operator. To this end, initial work within the TVS focused on creating requirements that would mitigate differences in various systems being used in a multi-centric study (Mazzà et al., 2021; Scott et al., 2022) and establishing the estimation of accurate spatiotemporal outcomes for comparison to the wearable device (Bonci et al., 2022). Resources and source code for future use are available through the mentioned publications.

Real-world assessment

In the real-world, the Mobilise-D consortium developed a wearable multi-sensor system, referred to as the INDIP system (INertial module with DIstance sensors and Pressure insoles) (Salis et al., 2023). It contained two magneto-IMUs positioned over the instep and fixed to shoelaces with clips, and a third magneto-IMU attached to the lower back with Velcro (Figure 3). Pressure insoles specific for each participant's foot size were inserted into the shoes. Utilization of the INDIP system as a reference during the real-world protocol proved to be successful in overcoming limitations in accuracy, battery life and usability, all of which are common restrictions of real-world reference systems previously adopted in the literature (e.g., wearable camera and global navigation satellite system systems (GNSS)). The INDIP system had been validated against the SP system across a range of conditions and in the TVS cohorts, with excellent results and reliability in quantification of DMOs (MAE laboratory ≤ 0.02 m/s, simulated daily activities = 0.03 to 0.05 m/s). A complete overview of the validation results can be found in (Salis et al., 2023). The INDIP and the wearable device were synchronized using their timestamps (± 10 ms). Thus, the INDIP system could be used in both laboratory and real-world settings to enable a concurrent validation of DMOs such as the walking speed, as provided in the present study (Kirk et al., 2024; Micó-Amigo et al., 2024). The recording duration of 2.5 hours with the INDIP system enabled recording of a wide range of activities and walking speeds.

6.6 Contextual factors

To estimate contextual factors participants were asked to carry a smartphone device with the Aeqora App installed (Ciravegna et al., 2019) for both the 2.5 hours free-living assessment and 7-days real-world assessments. This enabled capture of GNSS data needed to estimate indoor/outdoor location, weather, and elevation, which could all potentially impact the validity of DMOs. Additional factors such as walking aid usage could also affect validity of DMO estimation, due to changing the walking signal measured by the wearable device. To capture this in the TVS, a Bluetooth beacon was also fixed to the participants walking aid, to understand whether DMOs can be estimated with acceptable accuracy whether the individual utilises a walking aid.

6.7 Learning from Mobilise-D: Technical validation reference systems

In the TVS, we were aware that, using a 2.5-h window of activity in the real world was insufficient for validation purposes. Therefore, it may not have captured change and higher variability in mobility due to fatigue or the inconsistent nature of activity. We recommend that, if battery life and technical challenges can be overcome, a longer duration of real-world validation data collection is preferred.

Additional reference systems were used to capture contextual information such as walking aid usage or GPS location information from Bluetooth beacons and smart phones. The Bluetooth beacons enabled a more granular insight of walking aid usage upon algorithm performance, through identification of which specific walking bouts were completed with assistance of a walking aid, rather than broadly categorising individuals as walking aid users. The beacons in this scenario did not work, so more robust tools should be considered. Specifically, the beacons did not have sufficient granularity to detect if people used their walking aid or an aid was just close by. Comparison of algorithm performance indoors and outdoors, as measured by GPS coordinates captured from smart phones, would have added more contextual value to the specific real-world settings. This additional information improves the accuracy and reliability of the derived DMOs. However, many individuals did not take the smartphone with them when they left their houses, thus compliance to carrying the device was low.

6.8 Learning from Mobilise-D: Recruitment and validation of sample size

An original sample size of 120 was defined according to Consensus-based Standards for the selection of health Measurement Instruments guidelines for measurement properties (COSMIN, https://www.cosmin.nl/). This sample size allowed for 'excellent' methodological quality of non-inferiority studies, and is endorsed by the COSMIN checklist, a standardized tool for assessing the methodological quality of studies on measurement properties. In the TVS, participants was enrolled via their clinical care team or research registers to represent six cohorts: congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), healthy older adults (HA), multiple sclerosis (MS), Parkinson's disease (PD), and proximal femoral facture (PFF). Table 6 displays the specific number of participants recruited from each cohort. Each cohort was recruited across multiple sites to ensure generalizability (e.g. differing cultures and contexts).

Based on the required number of WBs, the targeted sample size was achieved for all cohorts including CHF. As such, we recommend that researchers include a range of gait-related tasks, that ensure a sufficient number of WBs are captured. **Table 6** reports the number of WBs included for each cohort.

Table 6.
Number of participants recruited in the TVS and the total number of WBs identified by the reference systems for each cohort. Sample size based on the minimum number of WBs required for robust statistical analyses was achieved (>630, corresponding to ICC>0.7 and CI=0.1).

Cohort	Planned Initially	Total Enrolled	Withdrawn	Actual	Total number of identified Walking Bouts (in-lab & 2.5 hour assessment)
Healthy Adults (HA)	20	20	-	20	1716
Multiple Sclerosis (MS)	20	20	-	20	1056
Parkinson's Disease (PD)	20	22	2	20	1138
Proximal Femoral Fracture (PFF)	20	20	1	19	900
Chronic Obstructive Pulmonary Disease (COPD)	20	17	-	17	1320
Congestive Heart Failure (CHF)	20	12	-	12	969
Overall	**120**	**111**	**3**	**108**	**7099**

7. Algorithms and the computational pipeline

The analytical pipeline has been re-implemented into Python, as part of the Mobgap package (https://github.com/mobilise-d/Mobgap), ensuring greater public availability.

The re-implemented Mobgap pipeline enables recommendation of a process to follow if users integrate new algorithms. The Mobilise-D pipeline consists of several algorithmic blocks that were developed by world leading research teams consisting of experts in the field of gait analysis (Kirk et al., 2024; Micó-Amigo et al., 2024). Once interested users have collected data from their wearable device, they can input the raw data into the Mobgap computational pipeline. The raw data is processed across several blocks and each block provides a given output. This pipeline provides results on a stride and WB level. Their definitions and respective thresholds for inclusion can be viewed in Table 7. The turning algorithm was not evaluated as part of the technical validation but was included in the pipeline GitHub repository.

Using the Mobilise-D TVS dataset?

If you are applying a new algorithm pipeline upon the Mobilise-D TVS dataset, please reference the TVS dataset accordingly based upon the guidelines in the dataset Zenodo repository. Mobgap is also recommended to load the TVS data, to ensure ease of data conversion and compatible input formats for pipeline automaticity.

Collecting new data to use the Mobilise-D pipeline?

This computational pipeline has now been re-implemented and is publicly available as part of the Mobgap package in python (https://github.com/mobilise-d/Mobgap, https://Mobgap.readthedocs.io/en/latest/README.html). If you apply the Mobilise-D existing pipeline upon your dataset please reference Micó-Amigo et al., 2024 for the individual pipeline block or Kirk et al., 2024 for walking speed from the complete pipeline

Walking bouts	A walking bout (WB) is a walking sequence containing at least two consecutive strides of both feet (e.g., R-L-R-L-R-L or L-R-L-R-L-R). The start and end of a walking bout are determined by a resting period or any other activity (non-walking period) > 3 seconds between detected steps. The initial step of a WB follows a non-walking period, and the final step precedes the next non-walking period.
Strides	A stride is the interval between two consecutive initial contacts of the same foot. As such, a stride is equivalent to the gait cycle and every stride contains two steps. Left strides start with an initial contact of the left foot, right strides with an initial contact of the right foot. Only strides of duration > 0.2 seconds and < 3 seconds were considered.
Turns	The process of turning consists of decelerating the forward motion, rotating the body as a whole, and stepping out toward the new direction (Hase and Stein, 1999). Thus, turning includes a change of walking direction and change in angular orientation including a rotational movement of the body around the longitudinal axis. Turning, curvilinear walking, and straight walking involve different neuromotor strategies and need to be discriminated. In Mobilise-D only turns of duration > 0.5 and < 10 seconds and rotation angle > 45o are considered.

Table 7.
Key operational definitions of walking bouts, strides and turns that were determined from a comprehensive review of consensus definitions (Kluge et al., 2021). A more detailed description can be viewed in Appendix 1.

The individual algorithmic blocks of the pipeline are shown in Figure 5, which uses raw IMU data as the input. Researchers should first consider the validity of algorithms from each block separately, before moving onto the validity of the entire pipeline combined.

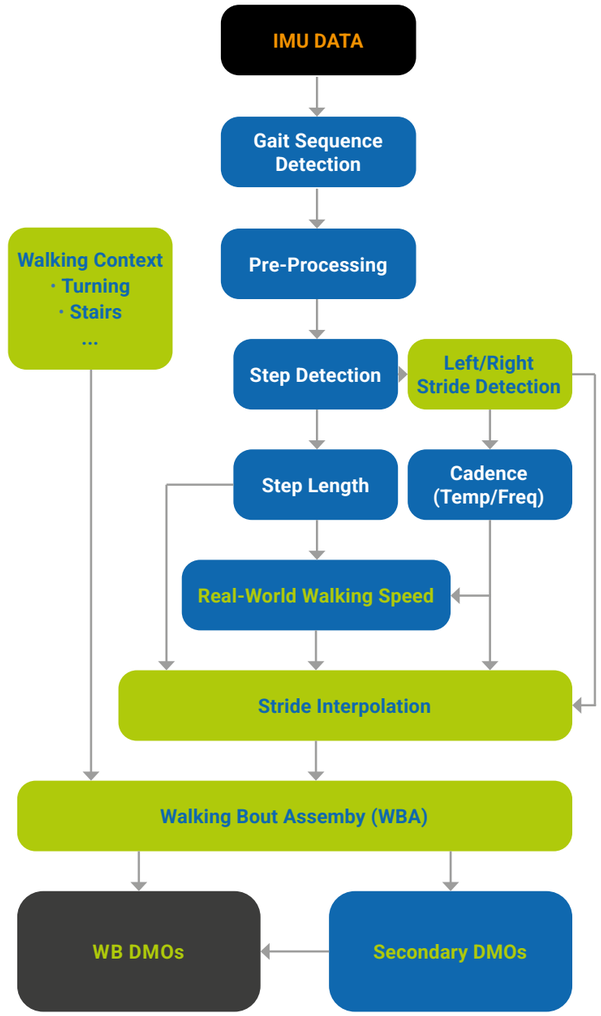
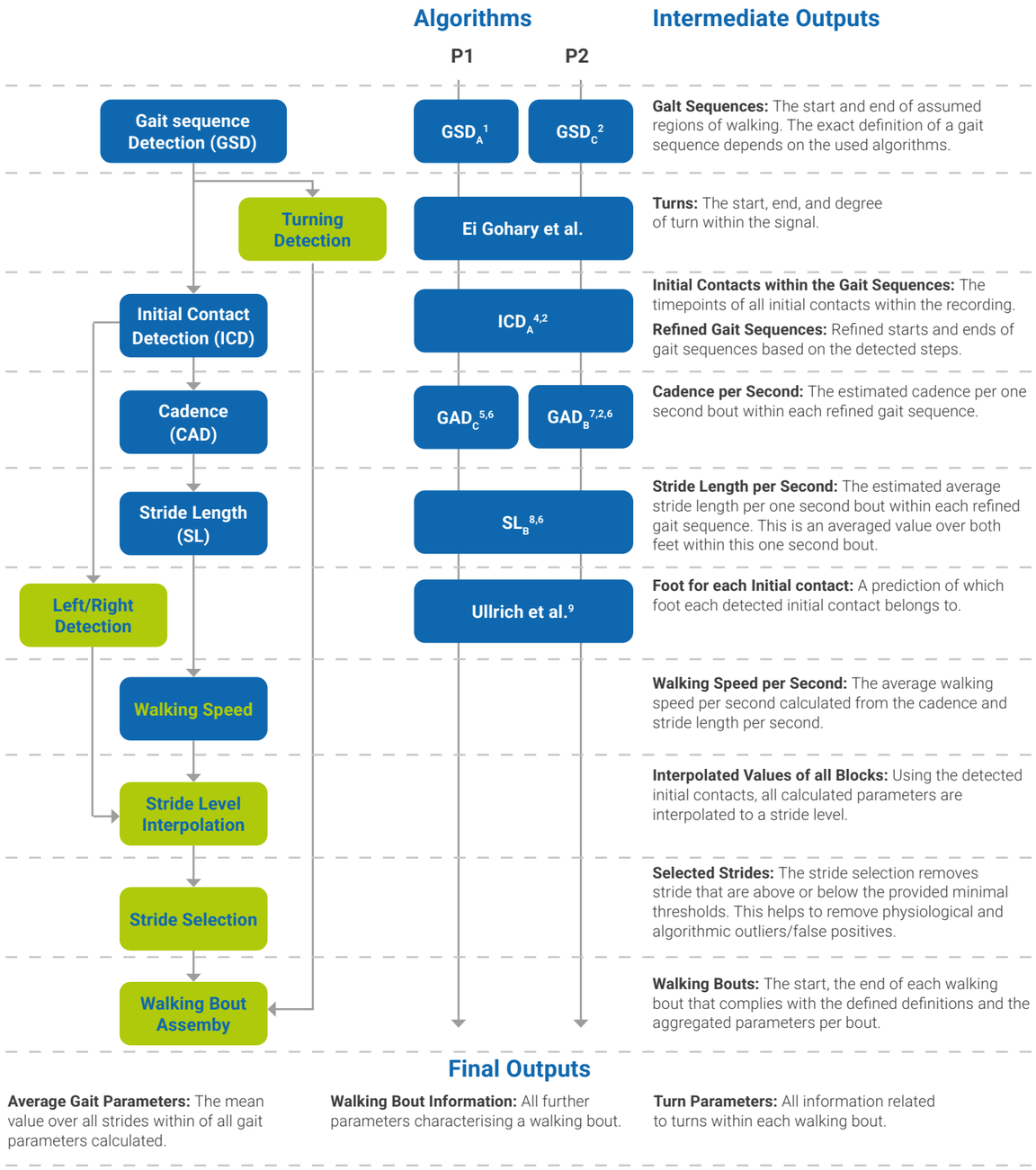


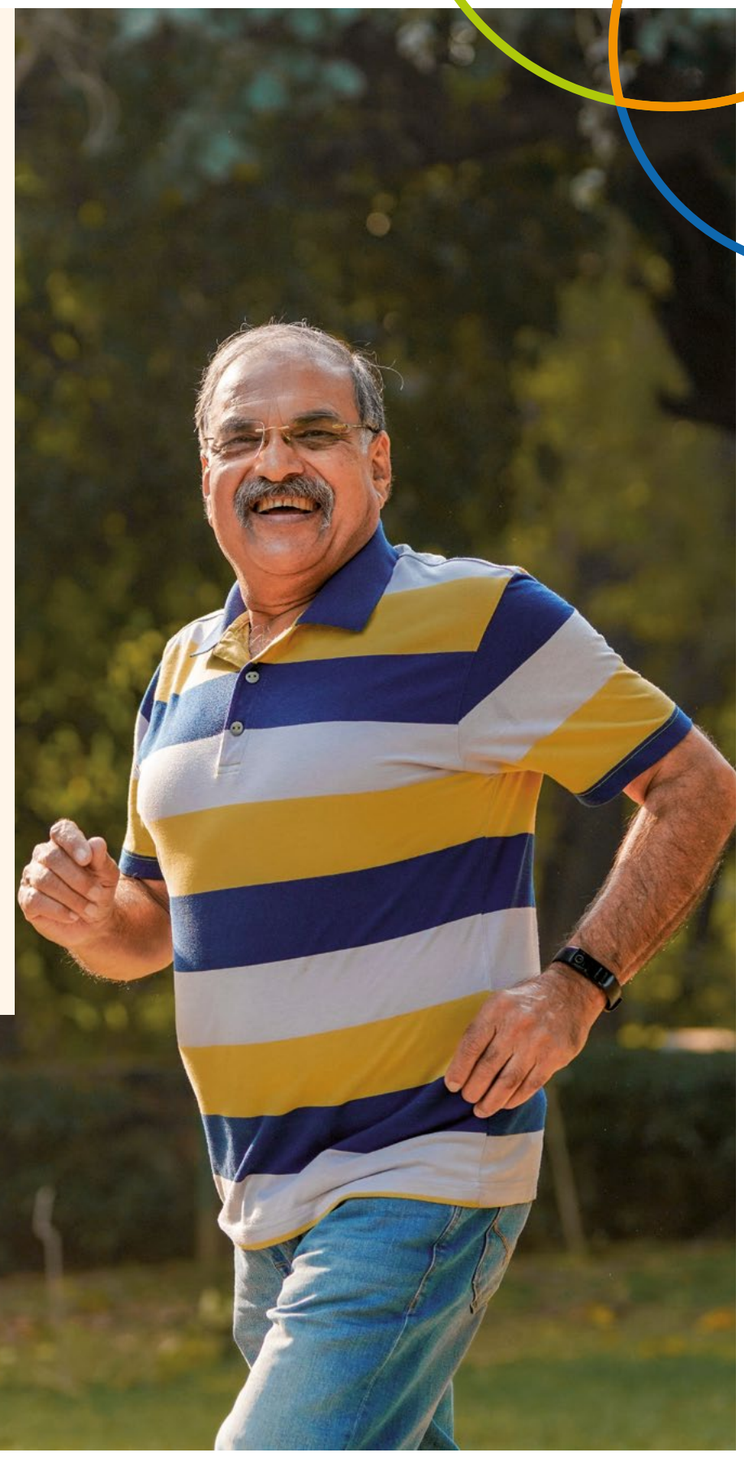
Figure 5. Mobilise-D pipeline to derive the DMOs.
For each block, a large number of algorithms were identified from a literature search that was developed from biomechanics, machine learning and signals processing techniques. These algorithms were implemented and improved for slow or impaired walking using appropriate pre-processing of IMU data and, were ranked on their performance based upon data collected in the TVS. Criterion validity metrics were estimated to compare the performance of all algorithms at each block of the DMO pipeline against the laboratory and real-world reference systems. For each block, the algorithm that had the strongest validity was determined as the top-ranked for that specific block and included in the pipeline. The results are described in detail in Micó-Amigo et al., 2023. The top ranked algorithms differed depending upon the cohort for certain DMOs.

The best performing algorithms from each block were combined and validated within a single-analytical pipeline to estimate walking speed. In the TVS, two analytical pipelines were compiled to represent cohort specific needs: P1 provides the optimal combination of algorithms selected for HA, COPD, and CHF conditions, and P2 provides the optimal combination for cohorts with more impaired gait patterns (PD, MS, and PFF). Both pipelines can be viewed in Figure 6, alongside descriptions of the processing steps and final outputs.



References:
1. Iluz et al., "Automated Detection of Missteps during Community Ambulation in Patients with Parkinson's Disease."
2. Paraschiv-Ionescu et al., "Locomotion and cadence detection using a single trunk-fixed accelerometer: validity for children with cerebral palsy in daily life-like conditions."
3. El-Gohary et al., "Continuous Monitoring of Turning in Patients with Movement Disability."
4. McCamley et al., "An Enhanced Estimate of Initial Contact and Final Contact Instants of Time Using Lower Trunk Inertial Sensor Data."
5. Shin and Park, "Adaptive Step Length Estimation Algorithm Using Optimal Parameters and Movement Status Awareness."
6. Soltani et al., "Algorithms for Walking Speed Estimation Using a Lower-Back-Worn Inertial Sensor."
7. Lee et al., "Computational Methods to Detect Step Events for Normal and Pathological Gait Evaluation Using Accelerometer."
8. Zijlstra and Hof, "Assessment of Spatio-Temporal Gait Parameters from Trunk Accelerations during Human Walking."
9. Ullrich et al., "Machine Learning-Based Distinction of Left and Right Foot Contacts in Lower Back Inertial Sensor Gait Data."

Figure 6.
Overview of the different algorithmic steps of the analytical pipeline with short explanations of intermediate and final outputs of each of the algorithmic blocks; gait sequence detection (GSD), initial contact detection (ICD), cadence estimation (CAD) and stride length estimation (SL). The algorithm column indicates the used algorithms for the two pipelines P1 (HA, COPD, CHF) and P2 (MS, PD, PFF). Short citations for the algorithms are provided below the figure. For more details see Table 1 in Micó-Amigo et al., 2024. These two pipelines allow the derivation of all the DMOs, including walking speed. A complete list of the primary and secondary DMOs were validated within the TVS, and together with their respective definitions they can be viewed in Appendix 7.



8. Statistical analysis plan and performance metrics for analytical validation

To validate the performance of the Mobilise-D pipeline upon a different population or context use case, follow the steps outlined below.

First it is fundamental to establish and clearly describe the statistical analysis plan framework for the validation of DMOs; to enable reproducibility and results comparison, this should include detailed information of the study context, protocol, and population, followed by a section that clearly defines all outcome measures and variables, and finally a detailed section that explains the overview of the planned statistical analyses.

Using the Mobilise-D TVS dataset?

The validation results of the existing Mobilise-D pipeline based upon the TVS dataset is published in Micó-Amigo et al., 2024 for the individual pipeline block and Kirk et al., 2024 for the validation of the complete pipeline. Validation of the Mobgap version of the pipeline is currently under development.

Collecting new data to use the Mobilise-D pipeline?

We recommend a complete validation of the mobilise-D pipeline in the clinical context that you intend to implement the DMOs (population, assessment setting, context etc.), using the framework outlined below.

Algorithms should not be limited to validation in isolation, and instead the complexity of the comprehensive multi-stage pipeline needed to estimate walking speed should be considered. This first requires the identification of walking activity, followed by the estimation of DMOs (e.g., steps, cadence, stride length, walking speed). The final outputs of the pipeline are the result of the interactions between all the algorithms (referred to as blocks).

The errors of the final DMOs depend on the errors of each block that accumulates across each step of the pipeline. As such, a robust validation of walking speed requires the estimation of walking speed from each algorithmic block combined in a single pipeline at a walking bout level. An overview of the validation results of the computational pipeline can be found in (Kirk et al., 2024).

8.1 Data cleaning

Prior to statistical analysis we recommend reviewing the distribution of all variables to check for missing values, inconsistencies, and outliers, or to identify patterns detected.

8.2 Statistical analysis

Several statistical analyses and performance metrics can be adopted to characterise the DMOs and evaluate selected criterion and concurrent validity metrics. Each of the statistical analyses and performance metrics can then be tailored to the nature of the specific DMOs and their level of aggregation/minimum granularity (e.g. step or stride, walking bout, test, etc). For example, in the TVS, to characterise the primary and secondary DMOs (see Appendix 7), various concurrent validity performance metrics were utilised to compare performance to the reference system. This approach is described in more detail by (Bonci et al., 2020). Data can be analysed as follows for the DMOs that present multiple values per WB (e.g., with a minimum granularity level set to a step/stride level, see Appendix 6).

- **1st level of aggregation** - at WB level: the **mean and standard deviation (SD)** across strides of a WB, together with the **absolute and relative errors of the mean and the SD**. In the case of DMOs whose analyses are based on true positives (e.g. initial and final contact events), we also obtained the **maximum absolute and relative errors, and the root-mean-square of the absolute errors** of each WB.
- **2nd level of aggregation** - at patient level (across all WBs): **descriptive statistics (mean, SD, median, inter-quartile range, minimal, maximal and root-mean-square)** of each of the values obtained at the 1st level of aggregation, for each participant, across all WBs of each lab test or real-world assessment.
- **3rd level of aggregation** - at cohort level: across all participants with the same disease/condition **(mean, 95% confidence interval (CI), SD, median, inter-quartile range, intra-class-correlation coefficient (when feasible) and p-value of normal distribution)** of all metrics obtained at the 2nd level of aggregation.

Note that for each of the DMOs, the 95% CI was evaluated only at cohort level (3rd level of aggregation). And in the case of ICC, analysis at cohort level was performed, working on DMOs that are true positive events – so single sensor (SS) events – that correspond to reference system (RS) events (e.g. step durations based on true positives per WB, for all WBs, for all participants of the cohort).

Categorical data analysis: Categorical data were analysed with ad hoc metrics and tools. For instance, laterality labels of the foot performing the initial contact events left or right, were reported using the absolute number of correctly identified labels, and the relative number (i.e. normalized to the total number of initial contact events). Differently to numerical analysis, instead of ICC we used the Cohen's kappa at the highest level of aggregation (cohort level), by including all initial contact events labels, based on the true positives. For more detailed descriptions of the statistical analyses see (Bonci et al., 2020; Kirk et al., 2024; Micó-Amigo et al., 2024)

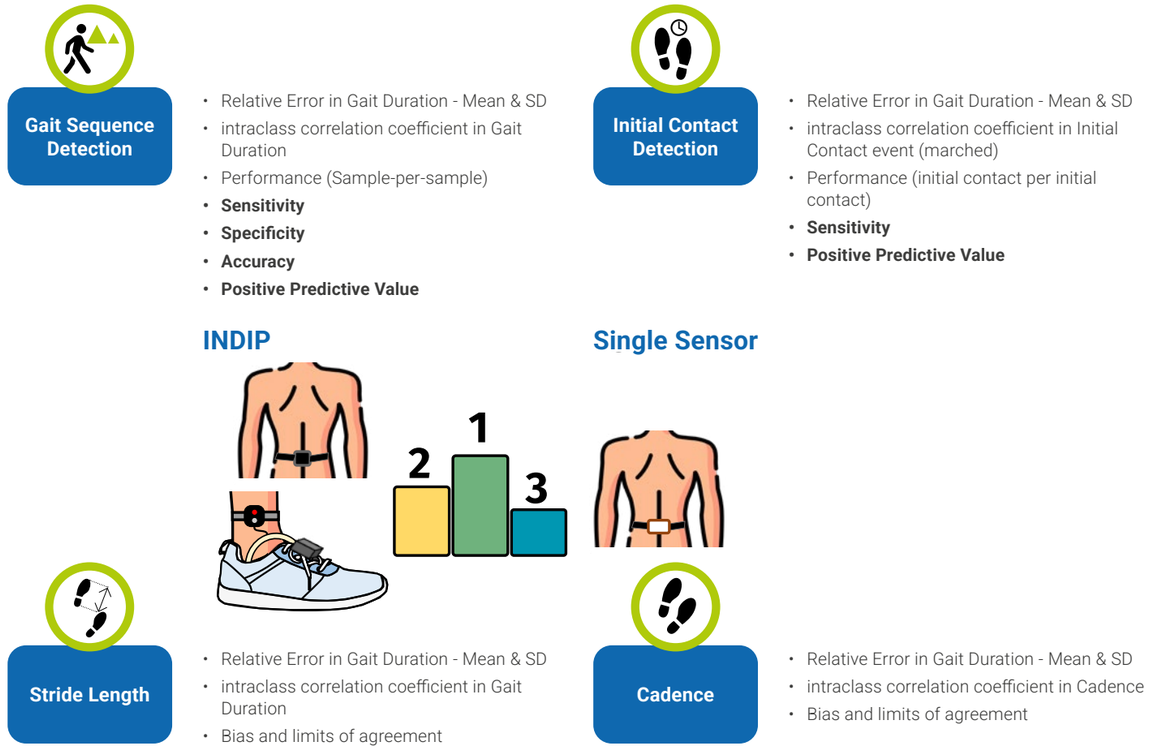


Figure 7.
Example of validity measures that were adopted for the concurrent analysis of DMOs within the TVS in laboratory and 2.5 hours real-world conditions (Micó-Amigo et al., 2024).

Although we strongly recommend setting the granularity of statistics as implemented for the Mobilise-D TVS, different levels of granularity of statistics can be adopted to provide multiple outcomes per individual walking bout and individual participant. This approach has been covered in more detail in previous literature (Bonci et al., 2020; Kirk et al., 2024; Micó-Amigo et al., 2024).

Acceptable values for performance/criterion validity metrics: As a general indication to provide a first assessment of the quality of the estimated DMOs, it was agreed that acceptable DMOs were characterized by performance metrics above 0.7 (i.e. Sensitivity, Positive Predictive Value, Accuracy, etc), with criterion validity metrics allowing for acceptable relative errors on the DMOs estimation of 20% or less (see Appendix 6 for details).

9. Technical validation insights: lessons learnt and recommendations

Based upon the results of the Technical Validation Study, we provide some key recommendations below on how to address the influence of confounding factors. The results of the individual algorithms of the DMOs pipeline and the analytical pipeline to estimate walking speed are extensively described by Micó-Amigo et al., 2024 and Kirk et al., 2024, respectively.

9.1 Importance of conconsidering computational pipeline

Users must evaluate the impact of individual pipeline blocks and consider the influence of DMOs error propagation. In our case, given the complexity of the respective algorithms, SL has demonstrated to have a larger contribution than cadence to the observed walking speed. Thus, future researchers could develop methods to improve SL estimation. Furthermore, the wearable device seems to record shorter WBs than the reference system. Detection of initial contact and left-right detection is more difficult under challenging conditions (e.g., turns or stairs), which may limit reliable stride information separating longer periods of walking into multiple shorter WBs. This could be the result of the wearable device not being positioned properly on the lower back, whereas the real-world reference system (INDIP) included feet sensors as well as a lower back sensor, thus being more robust to detect gait events across longer periods. WB identification is influenced by the values of the step-by-step DMOs which concur to meet WB criteria. Therefore differences in measurements of single temporal or spatial gait variables (e.g. stride duration or stride length) can have an impact on the comparison between the Mobilise-D computational pipeline and the reference system.

This motivates further research in the determination of initial contacts and their laterality under challenging real-world conditions.

An overview of the block-by-block results for the different DMOs can be observed in Table 8. This demonstrates how algorithm performance provides its own unique challenge for each DMO, and thus each block and the entire pipeline should be validated separately.

Table 8.
Comparison of digital mobility outcome validity from laboratory and real-world settings (Kirk et al., 2024; Micó-Amigo et al., 2024). Results presented across all cohorts unless specified.

Digital mobility outcome	Laboratory	Real-world
Gait sequence detection	Very good accuracy (> 0.82), F1-score (>0.79) and specificity (>0.69).	Excellent accuracy (>0.95), F1-score (>0.79) and specificity (>0.97).
Initial contact detection	Mean relative errors <9.44%, moderate to good ICC (0.59-0.76).	Mean relative error <11.88% and an ICC between 0.50 and 0.64.
Cadence estimation	Excellent results with a low relative error (4.66-8.66%) and good agreement between the reference system and the device, across all cohorts (ICC: 0.72-0.85).	Very good results showing a mean relative error between 6.59% and 9.90%, ICC ranged between 0.54 and 0.75.
Stride length estimation	Relative error ranged from 14% to 30% for PFF, the most impaired cohort. Errors were ~20% for MS and PD, where gait was impaired and participants presented with greater asymmetry and shuffling episodes. Agreement was moderate to good across all cohorts (ICC: 0.50-0.81).	Excluding very short walking bouts (≤10 seconds of duration) showed a relative error between 19% and 28% and a moderate to good agreement (ICC: 0.55-0.78), with the lowest ICC found for the PFF cohort.
Walking speed	Mean absolute error over all tests of 0.12 m/s and a mean relative error of 17.82%, with ICC showing good agreement (0.80).	Mean absolute error was higher across all participants (0.13 m/s), with a mean relative error of 24.48% and ICC showing moderate agreement across all cohorts (0.33), when estimated from all walking bouts combined.

9.2 Assessment setting

Users must validate DMOs within the intended assessment setting, as it was observed that real-world and laboratory environments present very different challenges. There is a need to test the algorithms under the most challenging and diverse conditions including turning and changes of direction, fragmented and slow walking sequences and other non-walking activity that could be misinterpreted by the algorithm as walking. Unsurprisingly, real-world data created a greater challenge for algorithm performance as it contains a greater variety in walking and non-walking activities, turning, non-walking breaks and very short walking bouts which provided the biggest challenge to algorithm performance.

In the instance of turning, real-world data results exceeded the performances of the in-lab results. This can be explained by the laboratory protocol being specifically designed to capture a varied profile of different types of turns, angles and timings. The more challenging turns may not have been undertaken by participants in the real-world, explaining why real-world results were more favourable. Table 9 and Table 10 demonstrate the differences in DMO performance between laboratory and real-world assessment.

9.2.1 Cohort

DMOs should be validated on performance data from people with the specific disease or condition of interest. In the TVS, DMO validity differed between cohorts, which is demonstrated in Table 8 and Table 9. In general, cohorts characterised by greater gait impairments (reflected by slower overall walking speed), such as the PFF cohort in the TVS, provided the greatest challenge for DMO estimation.

DMO	COPD	CHF	HA	MS	PD	PFF
Gait sequence detection (Mean and 95 % CI of gait sequence duration (seconds), accuracy, F1 score, specificity)	11.9 [10.7 , 13.0] 0.85, 0.80, 0.89	12.4 [10.9 , 13.9] 0.84, 0.82, 0.85	10.6 [9.8 , 11.5] 0.83, 0.75, 0.88	11.3 [10.3 , 12.3] 0.87, 0.85, 0.89	11.3 [10.2 , 12.4] 0.84, 0.81, 0.85	11.0 [10.0 , 11.9] 0.78, 0.75, 0.75
Step time estimation (Mean and 95 % CIs (seconds), mean relative error %, ICC)	0.68 [0.66 , 0.70] 2.99%, 0.92	0.67 [0.65 , 0.69] 6.75%, 0.40	0.64 [0.62 , 0.65] 6.12%, 0.72	0.67 [0.66 , 0.69] 6.27%, 0.77	0.69 [0.67 , 0.70] 3.57%, 0.90	0.67 [0.65 , 0.68] 4.52%, 0.88
Cadence estimation (Mean and 95 % CIs (steps/min), mean relative error %, ICC)	91.7 [89.7 , 93.6] 4.49%, 0.92	92.5 [90.3 , 94.7] 6.57%, 0.80	96.5 [94.5 , 98.4] 5.23%, 0.73	95.3 [93.2 , 97.3] 5.25%, 0.85	92.8 [91.0 , 94.6] 3.80%, 0.94	94.8 [92.9 , 96.7] 4.23%, 0.80
Stride length estimation (Mean and 95 % CIs (metres), mean relative error %, ICC)	1.11 [1.09 , 1.13] 11.59%, 0.77	1.11 [1.08 , 1.14] 15.71%, 0.80	1.01 [0.99 , 1.03] 13.93%, 0.62	1.00 [0.97 , 1.04] 17.04%, 0.60	1.02 [0.99 , 1.06] 13.06%, 0.75	0.94 [0.91 , 0.96] 23.39%, 0.81
Walking speed (Mean and 95 % CIs (metres/sec), mean relative error %, ICC)	0.87 [0.55, 1.13] 2.27%, 0.91	0.87 [0.52, 1.24] -2.06%, 0.82	0.85 [0.57, 1.15] 0.55%, 0.86	0.84 [0.52, 1.23] 6.58%, 0.81	0.82 [0.53, 1.20] 7.99%, 0.79	0.72 [0.43, 1.04] 14.34%, 0.82

Table 9.
Validity of DMOs when estimated from all laboratory tasks combined across each cohort of interest

DMO	COPD	CHF	HA	MS	PD	PFF
Gait sequence detection (Mean and 95 % CI of gait sequence duration (seconds), accuracy, F1 score, specificity)	16.6 [13.1 , 20.1] 0.95, 0.70, 0.96	28.2 [15.6 , 40.7] 0.89, 0.62, 0.88	26.5 [18.0 , 35.0] 0.88, 0.58, 0.88	18.4 [14.7 , 22.0] 0.91, 0.56, 0.92	23.7 [16.9 , 30.4] 0.90, 0.58, 0.91	16.2 [11.8 , 20.7] 0.94, 0.61, 0.94
Step time estimation (Mean and 95 % CIs (seconds), mean relative error %, ICC)	0.75 [0.74 , 0.76] 5.95%, 0.70	0.71 [0.70 , 0.72] 6.29%, 0.58	0.75 [0.74 , 0.75] 5.83%, 0.80	0.72 [0.71 , 0.73] 6.27%, 0.79	0.72 [0.71 , 0.73] 6.04%, 0.70	0.76 [0.75 , 0.78] 7.55%, 0.48
Cadence estimation (Mean and 95 % CIs (steps/min), mean relative error %, ICC)	86.6 [85.8 , 87.4] 5.64%, 0.73	92.2 [90.9 , 93.5] 4.55%, 0.88	87.4 [86.5 , 88.3] 5.53%, 0.820.73	88.5 [87.1 , 89.9] 6.40%, 0.82	89.5 [88.4 , 90.7] 5.25%, 0.82	86.2 [85.0 , 87.5] 5.41%, 0.82
Stride length estimation (Mean and 95 % CIs (metres), mean relative error %, ICC)	1.08 [1.06 , 1.09] 21.70%, 0.49	1.04 [1.02 , 1.07] 14.31%, 0.79	0.97 [0.95 , 0.98] 13.83%, 0.75	0.99 [0.96 , 1.01] 20.26%, 0.57	0.97 [0.94 , 0.99] 15.83%, 0.69	0.89 [0.88 , 0.91] 19.98%, 0.44
Walking speed (Mean and 95 % CIs (metres/sec), mean relative error %, ICC)	0.66 [0.52, 0.78] 22.71%, 0.57	0.82 [0.66, 1.04] 1.29%, 0.82	0.61 [0.51, 0.71] 10.53%, 0.88	0.71 [0.57, 0.83] 19.56%, 0.63	0.71 [0.51, 0.86] 17.79%, 0.72	0.64 [0.53, 0.77] 10.63%, 0.66

Table 10.
Validity of DMOs when estimated from 2.5-hour real-world assessment across each cohort of interest

9.2.2 Walking bout duration

Real-world walking is comprised of many bouts which differ in their duration and context throughout a day and from day to day. As such, a thorough analysis should be based on the duration of the WB (evaluated by the reference system), considering only the real-world assessment data. Specifically, WBs should be divided into groups according to WB duration. Below is an example of the thresholds that were adopted for analysis in the TVS:

- Group 1: WB duration ≤ 10 seconds
- Group 2: WB duration > 10
- Group 3: 10 seconds ≥ WB duration ≤ 30 seconds
- Group 4: 30 seconds ≥ WB duration ≤ 60 seconds
- Group 5: 60 seconds ≥ WB duration ≤ 120 seconds
- Group 6: WB duration > 120 seconds

We found that the WB duration directly influenced the validity of the algorithms. For initial contact detection, an overall error decrease was observed when the walking bout duration increased. Overall, higher errors (> 50%) were observed only in the 0.9% of the detected walking bouts; these bouts were characterised by a short duration (8.37 ± 4.71 s) and slow walking speed (0.44 ± 0.24 m/s). Highest errors for cadence algorithms also occurred within short WBs, with a mean duration of 8.88 seconds. Stride length errors were also dependent upon the duration, as shortest WBs < 10 seconds produced mean errors of 32.6%, while longest WBs had errors of 9.3%. Unsurprisingly, walking speed estimates were influenced by the WB duration, as very short WBs (< 10 seconds) presented the largest error (Figure 8).

9.2.3 Walking speed

Algorithms should be tested across the entire range of gait speeds that may typically be observed within the real world. The influence of gait speed can be explored by performing a subgroup analysis of the clinical cohorts. In the TVS, such analysis was based on the average speed of a gait stride quantified by the reference system in each assessment context. Participants were stratified on three levels: fast speed: > 1 m/s, medium speed: 0.5m/s – 1 m/s, slow speed < 0.5 m/s).

Gait speed had an impact upon performance metrics and therefore validity in the TVS. In general, the performance of all algorithms was lower for walking speeds below 0.5 m/s, which was considered as a threshold between slow and medium speed walkers. This could partially explain why the lowest performances were observed in the PFF cohort, as they had the slowest walking speed (Figure 8). In general, moderate walking speeds provided the greatest accuracy.

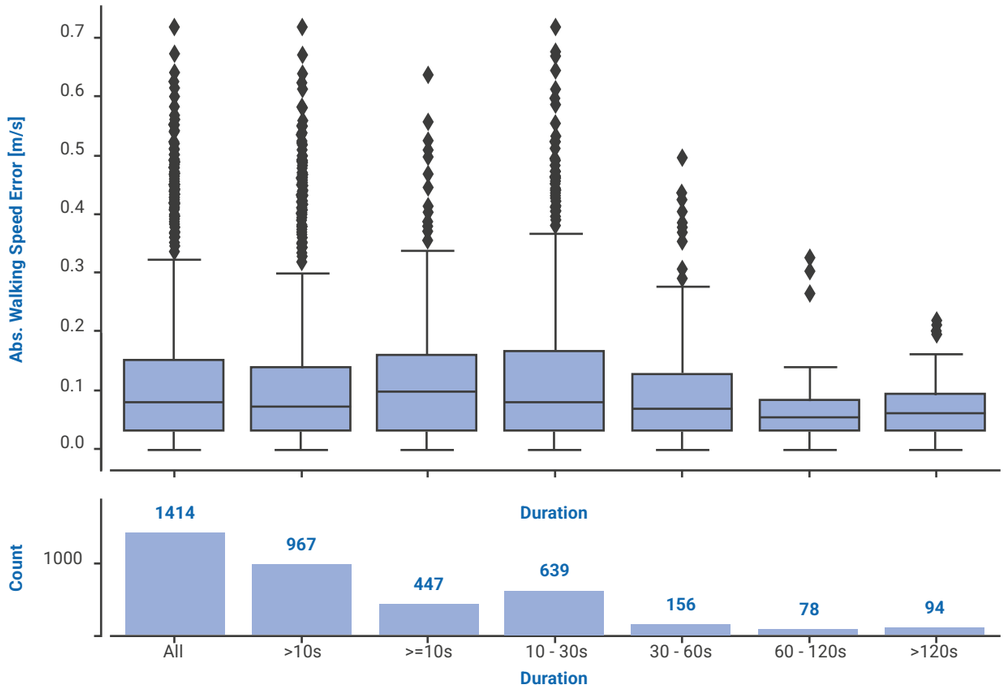


Figure 8.
The boxplots show how estimates of walking speed from the reference system and device were distributed over all WB duration groups. In the top, WB errors are grouped by various duration bouts. In the bottom the number of bouts within each duration group is visualized.

9.2.4 Walking bout complexity

Users should also evaluate DMO performance based upon the complexity of tasks/activities. In the TVS, the lowest complexity task was straight walking, and the task with greatest complexity was the simulated daily activities. The influence of complexity was cohort dependent and had the largest influence upon error for the MS, PD and PFF cohorts (processed with the P2 pipeline). We would expect those cohorts to experience more gait impairments than the CHF, COPD and HA cohorts (estimated from the P1 pipeline). However, whether the observed effect is caused by specific gait properties of the respective cohorts or by shortcomings in the selected algorithms, cannot be concluded based on the analysis performed.

Despite the challenges posed by outdoor environments (changes in terrain, weather and traffic negotiation due to humans and vehicles), outdoor environments captured more prolonged and uninterrupted walks in comparison to indoor environments, such as the household, which represent more confined and cluttered spaces with limited capacity for completing sequences of straight walking. Thus, error ranges from long uninterrupted outdoor walks are expected to be lower than results from confined indoor environments. Therefore, the combination of gait parameters with further contextual information might help to take this into account during data interpretation.

10. Key resources

The resources below are provided to supplement the recommendations set out in this document.

Verification of measurement tools

- Mobilise-D Webinar series – Session 2 – The Experimental tools
- Bonci T, Keogh A, Del Din S, et al. An Objective Methodology for the Selection of a Device for Continuous Mobility Assessment. Sensors 2020; 20: 6509.

Data Standardization

- Palmerini, L., Reggi, L., Bonci, T et al. Mobility recorded by wearable devices and gold standards: the Mobilise-D procedure for data standardization. Sci Data 2023; 10, 38. https://doi.org/10.1038/s41597-023-01930-9

Technical validation study protocol

- Mobilise-D webinar series – Session 3 – The Technical Validation Study Insights
- Mazzà C, Alcock L, Aminian K, et al. Technical validation of real-world monitoring of gait: a multicentric observational study. BMJ Open 2021; 11: e050785.
- Scott K, Bonci T, Salis F, et al. Design and validation of a multi-task, multi-context protocol for real-world gait simulation. Journal of NeuroEngineering and Rehabilitation 2022; 19: 141.
- Algorithms
- F Salis, S Bertuletti, T Bonci, et al. A Method for Gait Events Detection based on Low Spatial Resolution Pressure Insoles Data. Journal of Biomechanics, 127, 110687, 2021.
- F. Salis , S. Bertuletti, T. Bonci, et al., A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. Frontiers in Bioengineering and Biotechnology, 11, 2023.

Technical validation algorithms and results

- Mobilise-D Webinar series – Session 1 – The Algorithms
- Mobilise-D Webinar series – Session 5 – The TVS Early Insights
- Kirk C, Kuederle A, Micó-Amigo ME, et al. Author correction: Estimating real-world walking speed from a single wearable device: analytical pipeline, results and lessons learnt from the Mobilise-D technical validation study. 2024. Sci Rep 14, 28878. https://doi.org/10.1038/s41598-024-79454-4
- Micó-Amigo ME, Bonci T, Paraschiv-Ionescu A, et al. Correction: Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. Journal of NeuroEngineering and Rehabilitation 21, 71. https://doi.org/10.1186/s12984-024-01361-6

User acceptability

- What can we learn from patient perspectives on digital health technologies? Laura Delgado speaker at the Mobilise-D Summer School, 2023
- How to involve patients and the public in research and innovation. Dr Alison Keogh speaking at the Mobilise-D Summer School, 2023
- Keogh A, Alcock L, Brown P, et al. Acceptability of wearable devices for measuring mobility remotely: Observations from the Mobilise-D technical validation study. DIGITAL HEALTH 2023; 9: 205520762211507.
- Keogh A, Taraldsen K, Caulfield B, et al. It's not about the capture, it's about what we can learn": a qualitative study of experts' opinions and experiences regarding the use of wearable sensors to measure gait and physical activity. Journal of NeuroEngineering and Rehabilitation 2021; 18: 78.

11. Funding

This work was supported by the Mobilise-D project that has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 820820.

This JU received support from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations (EFPIA). LA, LR and SDD were also supported by the National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre (BRC) based at The Newcastle upon Tyne Hospital NHS Foundation Trust, Newcastle University and the Cumbria, Northumberland and Tyne and Wear (CNTW) NHS Foundation Trust. LA, LR and SDD were also supported by the NIHR/Wellcome Trust Clinical Research Facility (CRF) infrastructure at Newcastle upon Tyne Hospitals NHS Foundation Trust.

All opinions are those of the authors and not the funders. The content in this publication reflects the authors' view, and neither IMI nor the European Union, EFPIA, NHS, NIHR, DHSC, or any associated partners are responsible for any use that may be made of the information contained herein.

12. Acknowledgements

The authors acknowledge all the members of the Mobilise-D Work Packages 2, 3 and 6 for continuous discussion and critical input.

The authors want to specially acknowledge Dr Nunzio Camerlingo, Prof Brian Caulfield, Prof Bjoern Eskofier, Prof Jeffrey M. Hausdorff, Dr Ani Ionescu, Arne Kuederle, Dr Luca Palmerini, Dr Daniel Rooks and Pirinka Tuttle for critical review of this manual.

The authors are particularly grateful to the participants in the Technical Validation Study for their time and enthusiastic contribution, especially during the pandemic.

13. References

Alcock, L., Micó-Amigo, E.M., Bonci, T., Del Din, S., 2024. “Algorithms for gait: technical and clinical validity” in Gait, Balance and Mobility Analysis : Theoretical, Technical, and Clinical Applications.

Babrak, L.M., Menetski, J., Rebhan, M., Nisato, G., Zinggeler, M., Brasier, N., Baerenfaller, K., Brenzikofer, T., Baltzer, L., Vogler, C., Gschwind, L., Schneider, C., Streiff, F., Groenen, P.M.A., Miho, E., 2019. Traditional and Digital Biomarkers: Two Worlds Apart. Digit Biomark 3, 92–102. https://doi.org/10.1159/000502000

Bakker, J.P., Barge, R., Centra, J., Cobb, B., Cota, C., Guo, C.C., Hartog, B., Horowicz-Mehler, N., Izmailova, E.S., Manyakov, N.V., McClenahan, S., Motola, S., Patel, S., Paun, O., Schoone, M., Sezgin, E., Switzer, T., Tandon, A., van den Brink, W., Vairavan, S., Vandendriessche, B., Vrijens, B., Goldsack, J.C., 2025. V3+ extends the V3 framework to ensure user-centricity and scalability of sensor-based digital health technologies. npj Digit. Med. 8, 1–13. https://doi.org/10.1038/s41746-024-01322-2

Bonci, T., Keogh, A., Del Din, S., Scott, K., Mazzà, C., on behalf of the Mobilise-D consortium, 2020. An Objective Methodology for the Selection of a Device for Continuous Mobility Assessment. Sensors 20, 6509. https://doi.org/10.3390/s20226509

Bonci, T., Salis, F., Scott, K., Alcock, L., Becker, C., Bertuletti, S., Buckley, E., Caruso, M., Cereatti, A., Din, S., Gazit, E., Hansen, C., Hausdorff, J., Maetzler, W., Palmerini, L., Rochester, L., Schwickert, L., Sharrack, B., Vogiatzis, I., mazzà, C., 2022. An Algorithm for Accurate Marker-Based Gait Event Detection in Healthy and Pathological Populations During Complex Motor Tasks. Frontiers in Bioengineering and Biotechnology 10, 1. https://doi.org/10.3389/fbioe.2022.868928

Cagney, D.N., Sul, J., Huang, R.Y., Ligon, K.L., Wen, P.Y., Alexander, B.M., 2018. The FDA NIH Biomarkers, EndpointS, and other Tools (BEST) resource in neuro-oncology. Neuro-Oncology 20, 1162–1172. https://doi.org/10.1093/neuonc/nox242

Cerreta, F., Ritzhaupt, A., Metcalfe, T., Askin, S., Duarte, J., Berntgen, M., Vamvakas, S., 2020. Digital technologies for medicines: shaping a framework for success. Nat Rev Drug Discov 19, 573–574. https://doi.org/10.1038/d41573-020-00080-6

Ciravegna, F., Gao, J., Ireson, N., Copeland, R., Walsh, J., Lanfranchi, V., 2019. Active 10: Brisk Walking to Support Regular Physical Activity, in: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. Presented at the PervasiveHealth’19: The 13th International Conference on Pervasive Computing Technologies for Healthcare, ACM, Trento Italy, pp. 11–20. https://doi.org/10.1145/3329189.3329208

Coran, P., Goldsack, J.C., Grandinetti, C.A., Bakker, J.P., Bolognese, M., Dorsey, E.R., Vasisht, K., Amdur, A., Dell, C., Helfgott, J., Kirchoff, M., Miller, C.J., Narayan, A., Patel, D., Peterson, B., Ramirez, E., Schiller, D., Switzer, T., Wing, L., Forrest, A., Doherty, A., 2019. Advancing the Use of Mobile Technologies in Clinical Trials: Recommendations from the Clinical Trials Transformation Initiative. Digital Biomarkers 3, 145. https://doi.org/10.1159/000503957

Davis, F., Davis, F., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13, 319. https://doi.org/10.2307/249008

Deane, K.H.O., Flaherty, H., Daley, D.J., Pascoe, R., Penhale, B., Clarke, C.E., Sackley, C., Storey, S., 2014. Priority setting partnership to identify the top 10 research priorities for the management of Parkinson’s disease. BMJ Open 4, e006434. https://doi.org/10.1136/bmjopen-2014-006434

Deathe, A.B., Miller, W.C., 2005. The L Test of Functional Mobility: Measurement Properties of a Modified Version of the Timed “Up & Go” Test Designed for People With Lower-Limb Amputations. Physical Therapy 85, 626–635. https://doi.org/10.1093/ptj/85.7.626

Delgado-Ortiz, L., Polhemus, A., Keogh, A., Sutton, N., Remmele, W., Hansen, C., Kluge, F., Sharrack, B., Becker, C., Troosters, T., Maetzler, W., Rochester, L., Frei, A., Puhan, M.A., Garcia-Aymerich, J., 2023. Listening to the patients’ voice: a conceptual framework of the walking experience. Age Ageing 52, afac233. https://doi.org/10.1093/ageing/afac233

Goldsack, J.C., Coravos, A., Bakker, J.P., Bent, B., Dowling, A.V., Fitzer-Attas, C., Godfrey, A., Godino, J.G., Gujar, N., Izmailova, E., Manta, C., Peterson, B., Vandendriessche, B., Wood, W.A., Wang, K.W., Dunn, J., 2020. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). npj Digit. Med. 3, 1–15. https://doi.org/10.1038/s41746-020-0260-4

Hase, K., Stein, R.B., 1999. Turning Strategies During Human Walking. Journal of Neurophysiology 81, 2914–2922. https://doi.org/10.1152/jn.1999.81.6.2914

Hennink, M., Kaiser, B.N., 2022. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. Soc Sci Med 292, 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Hickey, A., Din, S.D., Rochester, L., Godfrey, A., 2016. Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis. Physiol. Meas. 38, N1. https://doi.org/10.1088/1361-6579/38/1/N1

Keogh, A., Alcock, L., Brown, P., Buckley, E., Brozgol, M., Gazit, E., Hansen, C., Scott, K., Schwickert, L., Becker, C., Hausdorff, J., Maetzler, W., Rochester, L., Sharrack, B., Vogiatzis, I., Yarnall, A., mazzà, C., Caulfield, B., 2023. Acceptability of wearable devices for measuring mobility remotely: Observations from the Mobilise-D technical validation study. DIGITAL HEALTH 9, 205520762211507. https://doi.org/10.1177/20552076221150745

Kirk, C., Küderle, A., Micó-Amigo, M.E., Bonci, T., Paraschiv-Ionescu, A., Ullrich, M., Soltani, A., Gazit, E., Salis, F., Alcock, L., Aminian, K., Becker, C., Bertuletti, S., Brown, P., Buckley, E., Cantu, A., Carsin, A.-E., Caruso, M., Caulfield, B., Cereatti, A., Chiari, L., D'Ascanio, I., Garcia-Aymerich, J., Hansen, C., Hausdorff, J.M., Hiden, H., Hume, E., Keogh, A., Kluge, F., Koch, S., Maetzler, W., Megaritis, D., Mueller, A., Niessen, M., Palmerini, L., Schwickert, L., Scott, K., Sharrack, B., Sillén, H., Singleton, D., Vereijken, B., Vogiatzis, I., Yarnall, A.J., Rochester, L., Mazzà, C., Eskofier, B.M., Del Din, S., 2024. Author Correction: Mobilise-D insights to estimate real-world walking speed in multiple conditions with a wearable device. Sci Rep 14, 28878. https://doi.org/10.1038/s41598-024-79454-4

Kluge, F., Din, S.D., Cereatti, A., Gaßner, H., Hansen, C., Helbostad, J.L., Klucken, J., Küderle, A., Müller, A., Rochester, L., Ullrich, M., Eskofier, B.M., Mazzà, C., Consortium, on behalf of the M.-D., 2021. Consensus based framework for digital mobility monitoring. PLOS ONE 16, e0256541. https://doi.org/10.1371/journal.pone.0256541

Knight, J.F., Baber, C., Schwirtz, A., Bristow, H.W., 2002. The comfort assessment of wearable computers, in: Proceedings. Sixth International Symposium on Wearable Computers,. Presented at the Sixth International Symposium on Wearable Computers (ISWC 2002), IEEE, Seattle, WA, USA, pp. 65–72. https://doi.org/10.1109/ISWC.2002.1167220

Maetzler, W., Klucken, J., Horne, M., 2016. A clinical view on the development of technology-based tools in managing Parkinson's disease. Mov Disord 31, 1263–1271. https://doi.org/10.1002/mds.26673

Mazzà, C., Alcock, L., Aminian, K., Becker, C., Bertuletti, S., Bonci, T., Brown, P., Brozgol, M., Buckley, E., Carsin, A.-E., Caruso, M., Caulfield, B., Cereatti, A., Chiari, L., Chynkiamis, N., Ciravegna, F., Del Din, S., Eskofier, B., Evers, J., Garcia Aymerich, J., Gazit, E., Hansen, C., Hausdorff, J.M., Helbostad, J.L., Hiden, H., Hume, E., Paraschiv-Ionescu, A., Ireson, N., Keogh, A., Kirk, C., Kluge, F., Koch, S., Küderle, A., Lanfranchi, V., Maetzler, W., Micó-Amigo, M.E., Mueller, A., Neatrour, I., Niessen, M., Palmerini, L., Pluimgraaff, L., Reggi, L., Salis, F., Schwickert, L., Scott, K., Sharrack, B., Sillen, H., Singleton, D., Soltani, A., Taraldsen, K., Ullrich, M., Van Gelder, L., Vereijken, B., Vogiatzis, I., Warmerdam, E., Yarnall, A., Rochester, L., 2021. Technical validation of real-world monitoring of gait: a multicentric observational study. BMJ Open 11, e050785. https://doi.org/10.1136/bmjopen-2021-050785

Mercer, K., L, G., E, S., P, C., M, L., K, G., 2016. Acceptance of Commercially Available Wearable Activity Trackers Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods Evaluation. JMIR mHealth and uHealth 4. https://doi.org/10.2196/mhealth.4225

Micó-Amigo, M.E., Bonci, T., Paraschiv-Ionescu, A., Ullrich, M., Kirk, C., Soltani, A., Küderle, A., Gazit, E., Salis, F., Alcock, L., Aminian, K., Becker, C., Bertuletti, S., Brown, P., Buckley, E., Cantu, A., Carsin, A.-E., Caruso, M., Caulfield, B., Cereatti, A., Chiari, L., D'Ascanio, I., Eskofier, B., Fernstad, S., Froehlich, M., Garcia-Aymerich, J., Hansen, C., Hausdorff, J.M., Hiden, H., Hume, E., Keogh, A., Kluge, F., Koch, S., Maetzler, W., Megaritis, D., Mueller, A., Niessen, M., Palmerini, L., Schwickert, L., Scott, K., Sharrack, B., Sillén, H., Singleton, D., Vereijken, B., Vogiatzis, I., Yarnall, A.J., Rochester, L., Mazzà, C., Del Din, S., for the Mobilise-D consortium, 2024. Correction: Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. Journal of NeuroEngineering and Rehabilitation 21, 71. https://doi.org/10.1186/s12984-024-01361-6

Micó-Amigo, M.E., Bonci, T., Paraschiv-Ionescu, A., Ullrich, M., Kirk, C., Soltani, A., Küderle, A., Gazit, E., Salis, F., Alcock, L., Aminian, K., Becker, C., Bertuletti, S., Brown, P., Buckley, E., Cantu, A., Carsin, A.-E., Caruso, M., Caulfield, B., Cereatti, A., Chiari, L., D'Ascanio, I., Eskofier, B., Fernstad, S., Froehlich, M., Garcia-Aymerich, J., Hansen, C., Hausdorff, J.M., Hiden, H., Hume, E., Keogh, A., Kluge, F., Koch, S., Maetzler, W., Megaritis, D., Mueller, A., Niessen, M., Palmerini, L., Schwickert, L., Scott, K., Sharrack, B., Sillén, H., Singleton, D., Vereijken, B., Vogiatzis, I., Yarnall, A.J., Rochester, L., Mazzà, C., Del Din, S., for the Mobilise-D consortium, 2023a. Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. Journal of NeuroEngineering and Rehabilitation 20, 78. https://doi.org/10.1186/s12984-023-01198-5

Micó-Amigo, M.E., Bonci, T., Paraschiv-Ionescu, A., Ullrich, M., Kirk, C., Soltani, A., Küderle, A., Gazit, E., Salis, F., Alcock, L., Aminian, K., Becker, C., Bertuletti, S., Brown, P., Buckley, E., Cantu, A., Carsin, A.-E., Caruso, M., Caulfield, B., Cereatti, A., Chiari, L., D'Ascanio, I., Eskofier, B., Fernstad, S., Froehlich, M., Garcia-Aymerich, J., Hansen, C., Hausdorff, J.M., Hiden, H., Hume, E., Keogh, A., Kluge, F., Koch, S., Maetzler, W., Megaritis, D., Mueller, A., Niessen, M., Palmerini, L., Schwickert, L., Scott, K., Sharrack, B., Sillén, H., Singleton, D., Vereijken, B., Vogiatzis, I., Yarnall, A.J., Rochester, L., Mazzà, C., Del Din, S., for the Mobilise-D consortium, 2023b. Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. Journal of NeuroEngineering and Rehabilitation 20, 78. https://doi.org/10.1186/s12984-023-01198-5

Palmerini, L., Reggi, L., Bonci, T., Del Din, S., Micó-Amigo, M.E., Salis, F., Bertuletti, S., Caruso, M., Cereatti, A., Gazit, E., Paraschiv-Ionescu, A., Soltani, A., Kluge, F., Küderle, A., Ullrich, M., Kirk, C., Hiden, H., D'Ascanio, I., Hansen, C., Rochester, L., Mazzà, C., Chiari, L., 2023. Mobility recorded by wearable devices and gold standards: the Mobilise-D procedure for data standardization. Sci Data 10, 38. https://doi.org/10.1038/s41597-023-01930-9

Port, R.J., Rumsby, M., Brown, G., Harrison, I.F., Amjad, A., Bale, C.J., 2021. People with Parkinson's Disease: What Symptoms Do They Most Want to Improve and How Does This Change with Disease Duration? J Parkinsons Dis 11, 715–724. https://doi.org/10.3233/JPD-202346

Rabinovich, R., Louvaris, Z., Raste, Y., Langer, D., Van Remoortel, H., Giavedoni, S., Burtin, C., Regueiro, E., Vogiatzis, I., Hopkinson, N., Polkey, M., Wilson, F., Macnee, W., Westerterp, K., Troosters, T., 2013. Validity of physical activity monitors during daily life in patients with COPD. The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology 42. https://doi.org/10.1183/09031936.00134312

Rochester, L., Mazza, C., Mueller, A., Caulfield, B., McCarthy, M., Becker, C., Miller, R., Piraino, P., Viceconti, M., Dartee, W.P., Garcia-Aymerich, J., Aydemir, A.A., Vereijken, B., Arnera, V., Ammour, N., Jackson, M., Hache, T., Roubenoff, R., 2020. A Roadmap to Inform Development, Validation and Approval of Digital Mobility Outcomes: The Mobilise-D Approach. Digit Biomark 4, 13–27. https://doi.org/10.1159/000512513

Salis, F., Bertuletti, S., Bonci, T., Caruso, M., Scott, K., Alcock, L., Buckley, E., Gazit, E., Hansen, C., Schwickert, L., Aminian, K., Becker, C., Brown, P., Carsin, A.-E., Caulfield, B., Chiari, L., D'Ascanio, I., Din, S., Eskofier, B., Cereatti, A., 2023. A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. Frontiers in Bioengineering and Biotechnology 11. https://doi.org/10.3389/fbioe.2023.1143248

Scott, K., Bonci, T., Salis, F., Alcock, L., Buckley, E., Gazit, E., Hansen, C., Schwickert, L., Aminian, K., Bertuletti, S., Caruso, M., Chiari, L., Sharrack, B., Maetzler, W., Becker, C., Hausdorff, J.M., Vogiatzis, I., Brown, P., Del Din, S., Eskofier, B., Paraschiv-Ionescu, A., Keogh, A., Kirk, C., Kluge, F., Micó-Amigo, E.M., Mueller, A., Neatrour, I., Niessen, M., Palmerini, L., Sillen, H., Singleton, D., Ullrich, M., Vereijken, B., Froehlich, M., Brittain, G., Caulfield, B., Koch, S., Carsin, A.-E., Garcia-Aymerich, J., Kuederle, A., Yarnall, A., Rochester, L., Cereatti, A., Mazzà, C., for the Mobilise-D consortium, 2022. Design and validation of a multi-task, multi-context protocol for real-world gait simulation. Journal of NeuroEngineering and Rehabilitation 19, 141. https://doi.org/10.1186/s12984-022-01116-1

Soltani, A., Dejnabadi, H., Savary, M., Aminian, K., 2020. Real-World Gait Speed Estimation Using Wrist Sensor: A Personalized Approach. IEEE Journal of Biomedical and Health Informatics 24, 658–668. https://doi.org/10.1109/JBHI.2019.2914940

14. Appendix

Gait Sequences	Describes algorithms that can detect gait in a continuous sensor signal (no restriction on sensor modality). The prime purpose is to pre-select regions of interest in the signal to aid a subsequent step/stride detection algorithm. In the case of single sensor IMU signals a Gait Sequence Detection algorithm can, for example, be based on a signal threshold (e.g. Hickey, Physio Meas 2017) or on frequency analysis (e.g. Ullrich, BHI 2019) of the raw signal. These algorithms are implemented as part of Task 2.2.1 (but are called WB detection in the project description). The output of the gait sequence detection. They describe regions that potentially contain gait. However, no general rules or requirements of the nature of the contained gait exist. This depends on the algorithm that is used to calculate the Gait Sequence. Therefore, gait sequences are not usable to calculate primary DMOs and further processing steps should be taken to generate Walking Bouts based on the identified gait sequences.
Walking bout detection	Originally used to describe algorithms that are able to detect gait in continuous sensor signals. This is no longer the case, as it cannot be guaranteed that these type of algorithms are able to detect sequences of gait that already fulfil the minimal technical requirements of WBs (provided by the consensus process). Therefore, these kind of algorithms are renamed to Gait Sequence Detection. In the workflow proposed in this document the term WB Detection might be used to refer to one of the steps in the WB Assembly. However, the use of the term WB Detection is discouraged (See the WB Assembly definition for details).

Appendix 1.
Detailed descriptions of operational definitions of Mobilise-D (Kluge et al., 2021)

Stride Selection	The process of filtering a list of detected strides using simple thresholds for certain spatial temporal parameters. This step is introduced to avoid the inclusion of shuffling, stair walking and similar “gait like activities” into the final analysis. Such activities are often misclassified as gait by common IMU based algorithms, but can be identified once the full set of spatial-temporal parameter is calculated. As an example, all strides below a certain stride length should not be considered a stride, but rather a misclassified shuffling (or similar movement). The exact value of the applied thresholds might change based on patient cohort and needs to be determined in agreement with medical professionals
WB Assembly	Large block in the algorithm pipeline situated after the detection and selection of single strides and the calculation of their parameters. It fulfills three main purposes: (1) formalize the process of grouping the detected strides into sets of consecutive strides, (2) identify WBs within these sets of consecutive strides based on a set of minimal criteria (Minimal WB requirements) and (3) to filter out WBs that do not fulfill certain additional requirements (Additional WB Requirements). The additional requirements are meant to focus a subsequent analysis only on clinically relevant WBs. The exact requirements for clinically relevant WBs are unknown at this point in the project. The specific thresholds can be modified later to compare various analysis approaches. Steps 2 and 3 performed by the WB Assembly could be called WB Detection and WB Selection, respectively. However, because they are closely connected and implemented as a single block/ function, the term WB Assembly to refer to the whole process is preferred.

Level walking bouts and walking bouts	Two separate types of WBs that are considered for analysis. Both types have their own set of Minimal WB Requirements and are generated by the WB Assembly using their respective rule set. Micro WBs are designed to be used to calculate the primary DMOs (average gait speed over a WB). Therefore, they only contain flat walking. without harsh turns. The average/std values over these WBs are expected to reflect a participant's ability to walk under ideal conditions. In contrast, a Macro WB has less strict conditions and is designed to capture long uninterrupted periods of gait including stairs, incline/decline walking and harsh turns. Hence, all strides of a Micro WB are also included in a Macro WB. Macro WBs are intended to be used to analyze secondary outcomes like the maximum walking capacity of a patient.
Incline/Decline walking bouts	WBs that only contain incline/decline walking or stairs. Otherwise, they use a very similar rule set to the Micro WBs. Incline/decline WBs can be used to specifically analyze the performance of participants under these more difficult conditions. All strides that are included in an incline/decline WB will also be included in a Macro WBs.
Continuous Walking Period	Long uninterrupted walking periods that are calculated by the INDIP system. They are very similar to Macro WBs and should be considered identical for the context of comparison and evaluation. However, to facilitate that difference can still exist, a separate term is used. This term should not be used to describe the output of the single sensor system and should also not be confused with a Gait Sequence.
Minimal walking bout Requirements	Strict minimal set of thresholds and rules every WB needs to fulfill. Every sequence of strides that does not pass one (or more) of thresholds is not considered a WB. The Minimal WB Requirements are applied as part of the WB Assembly. Running the WB Assembly with its "default" parameters is equivalent to only applying the Minimal WB Requirements to find WBs. **A walking bout must consist of two strides of each limb, at a minimum.**
Additional walking bout requirements	Set of requirements that allow to filter the list of all WBs to focus the analysis on a specific subset of WBs. The Additional WB Requirements can work on the same parameters as the Minimal WB Requirements, but must be stricter than those. For example, if one of the Minimal WB Requirements defines max. break between two strides to be X seconds, one of the Additional WB Requirements could redefine this parameter to a value <X seconds (= less inclusive rule), but not to a value >X seconds (= more inclusive rule). Additional WB Requirements are applied as part of the WB Assembly together with the Minimal WB Requirements. If Additional Requirements are used in the WB Assembly, the final output should not be simply called WB, but Selected WB to indicate the use of a stricter rule set.
WB Termination Criteria	All WB requirements that describe an event that ends a WB. For example, the max. break time between two strides or the max. allowed turning angle are Termination Criteria, as both describe an event (a break with a specific length or a turn of a specific degree value) that would mark the end of a WB, if it occurs. Termination Criteria vs. Inclusion Criteria (see below) is an additional grouping of requirements that is relevant for the technical implementation of the WB Assembly. Every rule used as Minimal/Additional WB requirement is either a Termination or an Inclusion criteria. WB is specifically terminated by any resting period of greater than three seconds between detected strides.
WB Inclusion Criteria	WB requirements that operate on summary parameters of a WB. For example, the number of strides or the mean gait velocity are Inclusion Criteria. This means they can only be evaluated once a WB candidate (sequence of strides that might be a WB) is identified.
Selected WB	Final output of the WB Assembly. The use of the word Selected facilitates that the WB Assembly applies Additional WB Requirements on top of to the Minimal WB Requirements. If the WB Assembly is used without applying Additional WB Requirements, its output might simply be called WB, without any addition prefix.

		Testing condition											
		x-up		y-up		z-up		x-down		y-down		z-down	
	Device	Norm [g]	Noise [mg]	Norm [g]	Noise [mg]	Norm [g]	Noise [mg]	Norm [g]	Noise [mg]	Norm [g]	Noise [mg]	Norm [g]	Noise [mg]
a	DynaPort MM+	0.99	1.88	0.99	1.88	0.98	2.31	1.02	1.81	1.01	1.81	1.00	2.32
	Axivity AX6	0.99	1.43	0.98	1.56	1.02	1.42	1.01	1.49	1.02	1.40	1.00	1.52
	INDIP	1.00	1.63	0.99	2.52	1.00	3.40	1.00	1.73	1.00	2.77	0.99	3.41

	Device	Mean [deg]	STD [deg]	Mean [deg]	STD [deg]	Mean [deg]	STD [deg]	Mean [deg]	STD [deg]	Mean [deg]	STD [deg]	Mean [deg]	STD [deg]
ϑ	DynaPort MM+	1.0	0.4	1.8	0.6	1.2	0.5	1.3	0.4	0.9	0.4	1.2	0.6
	Axivity AX6	2.8	1.3	1.9	1.1	2.5	1.1	2.2	1.0	2.3	1.2	3.1	1.3
	INDIP	0.6	0.4	0.4	0.2	0.4	0.2	0.5	0.3	0.4	0.2	0.5	0.2

	Device	Bias [deg/s]	Noise [deg/s]	Bias [deg/s]	Noise [deg/s]	Bias [deg/s]	Noise [deg/s]	Bias [deg/s]	Noise [deg/s]	Bias [deg/s]	Noise [deg/s]	Bias [deg/s]	Noise [deg/s]
ω	DynaPort MM+	0.90	0.16	0.90	0.16	0.88	0.16	0.89	0.16	0.89	0.16	0.90	0.16
	Axivity AX6	0.73	0.05	0.72	0.05	0.72	0.05	0.72	0.05	0.72	0.05	0.72	0.05
	INDIP	n.a.	0.06	n.a.	0.07	n.a.	0.05	n.a.	0.06	n.a.	0.07	n.a.	0.05

Appendix 2.

Norm and Noise for the acceleration (a), mean and standard deviation of the deviation from gravity (ϑ, degrees), and bias and noise of the angular velocity (ω) measured from the devices during the short static acquisitions in the six different positions.

	Device	n	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Root Allan Variance characteristic BI [dps]	DynaPort	35	16.28	14.35	14.53	14.66	15.81	14.71	14.68
	Axivity	34	11.74	10.49	10.07	9.63	9.91	11.14	9.49
	INDIP	100	11.62						

Appendix 3.

Analysis of the signal stability (bias instability (Root Allan Variance), BI) for the different devices as assessed during each day of the long static acquisitions.

	Device	x	y	z
Mean Error [%]	DynaPort MM+	1.9 ± 1.2	1.6 ± 1.3	1.6 ± 1.0*
	Axivity AX6	1.4 ± 1.0	1.7 ± 0.9	1.1 ± 0.8
	INDIP	0.2 ± 0.5	0.3 ± 0.7	0.3 ± 0.6

Appendix 4.

Mean and standard deviation of the angular velocity errors measured for the different devices during the short dynamic acquisitions (*p<0.005 when comparing Axivity AX6 and DynaPort MM+).

Participant interviews

Aim of the interview

The aims of the semi-structured interview are to:

1) Ensure that the McRoberts Dynaport device is comfortable and acceptable to participants.
2) Explore the acceptability of wearable devices, for the purposes of healthcare monitoring, to participants in general.

Specifically, for the Mobilise-D study, it is critical to understand whether the Dynaport device is acceptable for them to wear for a week. These questions are therefore being asked first. Depending on the time remaining in the interview, the general questions regarding the use of wearables as monitoring devices should be left until you have learned everything you wish to about the Dynaport device.

Note: Questions in this topic guide are included to answer the above aims. If the participant is open and talks freely, they may answer some of the questions without being asked. Therefore, depending on the person, not all of these questions need to be asked. If they begin to talk about topics that may be interesting or relevant to the above aims, please feel free to continue to explore these, even if there is no specific question linked to it. In contrast, if participants are not very open, some potential prompts have been included with the questions below. These prompts are there as an optional guide and do not need to be used.

Appendix 5.

Interview structure for assessing comfortability and acceptability of measurement device (Keogh et al., 2023)

General points to remember regarding the interview process

Questions for each aim have been listed accordingly:

- Main interview questions
  - These focus on the primary aim
  - They should form the basis of the interview
- Planned follow-up questions
  - These will help make the interview more specific if required
  - They may be useful to direct non-talkative participants

Spontaneous questions that result from the responses of participants obviously cannot be included in this guide, but are encouraged.

When completing the interviews please remember the following:

- Start with broad questions to generate discussion.
- Try to avoid questions with multiple options as this will lead them towards a specific answer
- If you need to confirm specific details, then you can use questions which require short responses.
- Try to find a quiet area to complete the interview if possible. If possible, please also try to be alone with the participant. However, if they request or need to have another person present (e.g. carer, spouse) then do not refuse them this.
- Bring a notebook or something to take notes as you go. This will help you to remember key phrases. Participants may also tell you stories while they complete the questionnaires, you may write down these words at the time and ask about them in the interview.
- Please be aware that participants may become upset during the interview. If this happens, let them continue if they can.
- When participants provide you with responses do not respond with subjective phrases such as ‘great’ or ‘that’s surprising’ as this may alter their future responses. Please only respond with unemotional responses such as ‘ok’ or ‘that’s interesting’.
- If participants remark that they are uncertain about something, please do not educate them during the interview. For example, if they remark that they do not know what the Dynaport was measuring, wait until the interview is complete before providing them with this information.

Interview guide

Dynaport questions

Aim: Ensure that the McRoberts Dynaport device is comfortable and acceptable to participants

Main questions

1. Can you describe your experience of using the Dynaport sensor in the last week?
2. Can you tell me what you liked about the device? Disliked?
   Prompts if needed
   - Size/weight
   - Attachment to body
   - Ease of use
   - Comfort

3. How did the device make you feel? / Can you describe what it felt like to wear the device?
   Prompts if needed
   - In social environments/at home
   - Interaction with daily activities
   - Emotions associated with being monitored

4. What you change about the device if you could?

Follow-up questions: "If you don't mind, I'd like to ask you some specific details about the device."

1 How did you find the process of putting it on and taking it off?

2 How did the device influence your daily activities?
Prompts if needed
- How were they impacted?
- How did this make them feel?

3 Can you tell me about any difficulties that you had with the device?

4 How did you feel about wearing the device for a week?
Prompts if needed
- How would they feel if it was longer?
- Any concerns for the week?

Closing question

1 Is there anything else you would like me to know about the Dynaport?

The use of wearable devices in healthcare questions

Aim: Explore the acceptability of wearable devices, for the purposes of healthcare monitoring, to participants in general.

Main questions

1 Can you tell me about your experience of your health condition?
Note: Condition specific symptoms are listed at the end of this document.

2 Can you tell me about your experience of the care you've received for your condition?
Prompts if needed
- How do they feel about it?

3 Can you tell me about what sort of technology you currently use in your everyday life?
Prompts if needed
- How do you feel about using technology?
- What would make you use technology more? Less?
- Emotions associated with being monitored

4 What are your opinions on the use of technology in healthcare?
Prompts if needed
- What do you think it can be used for?

5 How would you feel about using technology to generate health information about yourself? (e.g. condition related smartphone app, self-reported outcomes platform, fitness tracker etc.),
Prompts if needed
- Why?
- What would make you use it?
- What would stop you from using it?
- What would need to change for you to use it?

Follow-up questions:

1. How do you feel about capturing health information in your daily life, using a wearable remote monitoring device?

2. How do you think digital technology used in your daily life would influence how you manage your condition?
   Prompts if needed
   - How would it impact their relationship with their health care provider?
   - Integration into activities of daily living

3. How would you feel about sharing this data with your health care provider? What about researchers?
   Prompts if needed
   - Usefulness
   - Impact of this

Closing question

1 Is there anything else you would like me to know about using technology in healthcare?

Note for researchers

Some participants may mention or allude to symptoms that are common with their respective conditions. These may include, but are not limited to:

Parkinson's Disease: risk of falls, episodes of 'freezing gait' where they feel like they cannot move forward, a tremor, reduced balance, slow walking speed, slowness of movement, muscular stiffness.

PFF: weakness, pain, reduced mobility

CHD: shortness of breath, fatigue, weakness, swelling in legs and ankles, irregular heartbeat, reduced ability to exercise, fluid retention, persistent cough.

COPD: shortness of breath, the need for home oxygen, wheezing, tightness in their chest, a chronic cough, increased mucus production, frequent infections, lack of energy, swelling, weight loss.

MS: muscle spasms or weakness, numbness and tingling, reduced mobility, fatigue, bladder and bowel issues, pain, depression.

Appendix 6.

Rabinovich questionnaire for usability assessment (Rabinovich et al., 2013)
Please select a response for each of these questions that best reflects your experiences and opinions of the wearable device.

Section A

How much trouble did you have getting started with the wearable device?.

NOTE: 'getting started with' refers to the first few hours of wear/use in the week. It is how the participant felt while wearing it initially at the start of the week.

- ☐ No trouble to start up
- ☐ Sometimes trouble to start up
- ☐ Regularly caused trouble to start up
- ☐ Always trouble to start up
- ☐ I had to call the centre to get help in starting-up

The wearable device was easy to put on/take off.

- ☐ Yes this was very easy
- ☐ This worked just fine
- ☐ I found it somewhat difficult
- ☐ I found it difficult
- ☐ I was unable to manage this on my own

I experienced technical problems with the wearable device.

- ☐ All the time
- ☐ Frequently
- ☐ Sometimes
- ☐ Seldom
- ☐ Never

The wearable device interfered with my normal activities.

- ☐ All the time
- ☐ Frequently
- ☐ Sometimes
- ☐ Occasionally
- ☐ Never

I felt comfortable wearing the wearable device.

- ☐ All the time
- ☐ Frequently
- ☐ Sometimes
- ☐ Occasionally
- ☐ Never

I felt embarrassed wearing the wearable device

- ☐ All the time
- ☐ Frequently
- ☐ Sometimes
- ☐ Occasionally
- ☐ Never

The instructions on how to use the wearable device were clear.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

Using the Wearable device on a daily basis was easy.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

The wearable device was bulky/heavy.

- ☐ Yes , very much so
- ☐ Yes much
- ☐ Not particularly
- ☐ Not at all
- ☐ No opinion

The wearable device bothered me in the bed.

- ☐ Yes , very much so
- ☐ Yes much
- ☐ Not particularly
- ☐ Not at all
- ☐ No opinion, I did not wear the monitor at night

I felt my privacy was invaded by the wearable device.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

If my doctor would like to use the wearable device to assess my physical activity, I would be willing to wear the monitor for.

- ☐ Less than 1 day
- ☐ 2-4 days
- ☐ 1week
- ☐ Longer than 1 week.
- ☐ I would not mind wearing the monitor continuously (longer than 1 month)

Section B

We want to ask you to provide a final score for the wearable device.

All things considered can you give the device a score from **0% to 100%**, where 0 means the worst possible monitor, and 100% means the ideal monitor in your opinion.

My score for the wearable device: ...../100

In the section below we want to give you the opportunity to give other comments on the wearable device.

I experienced the following problems with the wearable device:

I liked these features of the monitor in particular:

The Comfort Rating Scale

Please read each of the following questions, and mark which number best represents your agreement with this statement from a scale of '0 – low agreement' to '20 – high agreement'.

NOTE: Both sentences should be read. Please consider both statements and how they apply to how you felt during the week. Some may not apply. If this happens, please consider the statement in the pair that best matches your experience.

1) Emotion (how you feel while wearing the device)

I am worried about how I look when I wear this device. I feel tense or on edge because I am wearing it.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

2) Attachment (how the device feels against your body)

I can feel the device on my body. I can feel the device moving.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

3) Harm (whether you are concerned that the device is painful or harming you in some manner)

The device is causing me some harm. The device is painful to wear.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

4) Perceived change (whether the device makes you feel different to normal)

Wearing the device makes me feel physically different. I feel strange wearing the device.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

5) Movement (whether the device made you change the way you move during the day or night)

The device affects the way I move. The device inhibits or restricts my movement.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

6) Anxiety (whether the device makes you feel concerned about any aspect of your safety)

I do not feel secure or safe while wearing the device.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Appendix 7. List of primary and secondary DMOs for validation analysis

Outcome block	DMOs	Definition	Units	Minimum granularity (level of aggregation/ analysis)
Primary DMOs				
	Walking Speed	Ratio between the displacement covered within a WB and the time to cover it, calculated as the mean of stride speed values over the WB, as follows: $$\boldsymbol{WalkingSpeed_i} = \frac{\sum_{k=1}^{n\_STRIDE_i} StrideSpeed_k}{n\_STRIDE_i}$$ Where $Walking\_speed_i$ is the walking speed of the i^{th} WB, n_STRIDE_i is the number of qualified strides identified in the relevant i^{th} WB and $StrideSpeed_k$ is the stride speed of the k^{th} stride of the relevant i^{th} WB. Stride speed is defined as the ratio between the displacement covered within a gait cycle and the time to cover it, calculated as the product between cadence and stride length, as follows: $$StrideSpeed_k = Cadence_k \cdot Stride\_l_k$$ Where $StrideSpeed_k$ is the walking speed of the k^{th} stride, $Cadence_k$ is the cadence value of the k^{th} stride, $StrideLength_k$ is the stride length value of the k^{th} stride.	Meters/ second	Walking Bout
Secondary DMOs				
Walking Bout	Number of Walking Bouts	Based on the identification of gait as an activity (yes/no) across each sample of 0.1 seconds: n	Count	Lab Test/Real-world 2.5 hours assessment
	Walking Bout Duration	The time between the start (S_{WB_i}) and the end (E_{WB_i}) of a walking bout: $$\boldsymbol{D_{WB_i}} = S_{WB_i} - E_{WB_i}$$ $i = 1, \ldots, n$	Seconds	Walking Bout

Outcome block	DMOs	Definition	Units	Minimum granularity (level of aggregation/ analysis)
Step Duration	Initial Contact Events – Number of Events (steps)	Based on the qualified identification of Initial contact events (j): n_IC_{ij} $i = 1, \ldots, n;\ j = 1, \ldots, m$	Counts	Step, Walking Bout
	Step Duration	Duration between two consecutive initial contact events: $$\boldsymbol{D_{IC_{ij}\_IC_{ij+1}}} = IC_{ij+1} - IC_{ij}$$ $i = 1, \ldots, n;\ j = 1, .., m-1$	Seconds	Step, Walking Bout
Cadence	Cadence	Defined as step frequency within a minute, evaluated as double the stride frequency: $$\boldsymbol{Cadence_i} = StrideFrequency_i * 2$$ $i = 1, \ldots, n$ Stride Frequency is calculated as follows: $$Stride\ Frequency_i = \frac{\sum_{k=1}^{n\_STRIDE_i} \left(60/_{STRIDE\_d_k}\right)}{n\_STRIDE_i}$$ where $i = 1, \ldots, n$ are the different WBs, n_STRIDE_i is the number of right and left strides in the relevant i^{th} WB. $STRIDE\_d_k$ is the duration [seconds] of the kth stride in the relevant i^{th} WB.	Steps/ minute	Walking Bout
Stride Length	Stride Length	Average stride length within a walking bout i.e. the average value of all stride lengths within the walking bout: $$\boldsymbol{Stride\ Length_i} = \frac{\sum_{k=1}^{n\_STRIDE_i} (STRIDE\_l_{ik})}{n\_STRIDE_i}$$ where $i = 1, \ldots, n$ are the different WBs, n_STRIDE_i is the number of right and left strides in the relevant i^{th}WB $STRIDE\_l_{ik}$ is the length [meters] of the k^{th} stride in the relevant i^{th}WB	Meters	Walking Bout

Outcome block	DMOs	Definition	Units	Minimum granularity (level of aggregation/ analysis)
Turning	Number of Turns	Based on the identification of turns (yes/no) to a sample level of 0.1 seconds: p	Counts	Lab Test/Real-world 2.5 hour assessment
	Turn Duration	Time between the start and the end of the turns within the walking bout: $\boldsymbol{DT_{WB_i}} = ST_{WB_i} - ET_{WB_i}$ $i = 1,\ldots,n;\ w = 1,\ldots,p$	Seconds	Turn, Walking Bout
Other Secondary Outcomes	Swing Phase Duration	The time between the Final Contact of a foot and the following Initial Contact of the same foot $\boldsymbol{Swing\_duration_{ij}} = IC_{ij+1} - FC_{ij}$ $i = 1,\ldots,n;\ j = 1,\ldots,m-1$	Seconds	Stride, Walking Bout
	Stance Phase Duration	The time between Initial Contact and the Final Contact of the same foot $\boldsymbol{Stance\_duration_{ij}} = FC_{ij+1} - IC_{ij}$ $i = 1,\ldots n;\ j = 1,..m-1$	Seconds	Stride, Walking Bout

		Performance Metrics				
Block	**DGOs**	Sensitivity 95% CI	Positive Predictive Value 95% CI	Accuracy 95% CI	Specificity 95% CI	F1-score 95% CI
	Acceptable values	>0.7	>0.7	>0.7	>0.7	>0.7
Gait Sequence Detection/ Walking Bout	Number Identification of Walking Bouts					
	Start of Walking Bouts					
	End of Walking Bouts					
	Duration of Walking Bouts					
Initial Contact Detection & Left-Right identification	Initial Contact Events Time					
	Step and Stride Duration					
	Laterality					
	Sequence Order					
Cadence	Cadence steps/min					
Stride Length	Stride Length					
Real Walking Speed	Walking Speed					
Turning	Number Identification of Turns					
	Start of Turns					
	End of Turns					
	Duration of Turns					
	Maximal Angle of Turns					
Other DGOs	Final Contact Events Time					
	Stance Phase Duration					
	Swing Phase Duration					

Criterion Validity Metrics						Minimum Aggregation Levels			
Absolute Errors of mean & SD	Relative Errors of mean & SD	Maximum and RMS of Absolute Errors; Maximum Relative Error	Cohen's Kappa	Concurrent validity, ICC 95% CI	Statistical Significance p - value	Steps/Stride level	WB level	Participant level test assessment	Cohort/ subgroup level
	>20%		>0.7	>0.7	p>0.05				
								✓	✓
							✓	✓	✓
							✓	✓	✓
							✓	✓	✓
						✓	✓	✓	✓
							✓	✓	✓
							✓	✓	✓
							✓	✓	✓
							✓	✓	✓
							✓	✓	✓
							✓	✓	✓
							✓	✓	✓
						✓	✓	✓	✓
							✓	✓	✓
							✓	✓	✓

Abbreviations: CI: Confidence Intervals, ICC (2,1): Intra-Class Correlation

Appendix 8.
Framework of performance and validity metrics for technical validation of DMOs compared to reference systems. List of DMOs and corresponding performance/validity metrics with acceptable ranges and minimum aggregation level are provided. Similar types of DMOs for which the same set of statistical analyses and performance metrics must be applied are identified by colour.


innovative
medicines
initiative
efpia