

Evaluating the Accuracy and Reliability of Real-World Digital Mobility Outcomes in Older Adults After Hip Fracture: A Cross-Sectional Study

Martin A. Berge, Anisoara Paraschiv-Ionescu, Cameron Kirk, Arne Küderle, Encarna Micó-Amigo, Clemens Becker, Andrea Cereatti, Silvia Del Din, Monika Engdal, Judith Garcia-Aymerich, Karoline B. Grønvik, Clint Hansen, Jeffrey M. Hausdorff, Jorunn L. Helbostad, Carl-Philipp Jansen, Jochen Klenk, Lars Gunnar Johnsen, Sarah Koch, Walter Maetzler, Dimitrios Megaritis, Arne Müller, Lynn Rochester, Lars Schwickert, Kristin Taraldsen, Beatrix Vereijken

Submitted to: JMIR Formative Research
on: October 23, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 28

 Figures 29

 Figure 1..... 30

 Figure 2..... 31

 TOC/Feature image for homepages 32

 TOC/Feature image for homepage 0..... 33

Evaluating the Accuracy and Reliability of Real-World Digital Mobility Outcomes in Older Adults After Hip Fracture: A Cross-Sectional Study

Martin A. Berge¹; Anisoara Paraschiv-Ionescu²; Cameron Kirk³; Arne Küderle⁴; Encarna Micó-Amigo³; Clemens Becker^{5, 6}; Andrea Cereatti⁷; Silvia Del Din^{3, 8}; Monika Engdal¹; Judith Garcia-Aymerich^{9, 10, 11}; Karoline B. Grønvik¹; Clint Hansen¹²; Jeffrey M. Hausdorff^{13, 14, 15}; Jorunn L. Helbostad¹; Carl-Philipp Jansen^{5, 6}; Jochen Klenk^{6, 16, 17}; Lars Gunnar Johnsen^{1, 18}; Sarah Koch^{9, 10, 11, 19}; Walter Maetzler¹²; Dimitrios Megaritis^{3, 20}; Arne Müller²¹; Lynn Rochester^{3, 8}; Lars Schwickert⁶; Kristin Taraldsen²²; Beatrix Vereijken¹

¹ Department of Neuromedicine and Movement Science Norwegian University of Science and Technology Trondheim NO

² Laboratory of Movement Analysis and Measurement Ecole Polytechnique Federale de Lausanne Lausanne CH

³ Translational and Clinical Research Institute Faculty of Medical Sciences Newcastle University Newcastle Upon Tyne GB

⁴ Machine Learning and Data Analytics Lab Department Artificial Intelligence in Biomedical Engineering Friedrich-Alexander-Universität Erlangen-Nürnberg Erlangen DE

⁵ Geriatric Center Medical Faculty Heidelberg Heidelberg University Heidelberg DE

⁶ Department of Geriatrics and Rehabilitation Robert Bosch Hospital Stuttgart DE

⁷ Department of Electronics and Telecommunications Politecnico di Torino Turin IT

⁸ National Institute for Health and Care Research Newcastle Biomedical Research Centre Newcastle University and the Newcastle Upon Tyne Hospitals NHS Foundation Trust Newcastle Upon Tyne GB

⁹ Barcelona Institute for Global Health Barcelona ES

¹⁰ Universitat Pompeu Fabra Barcelona, Catalonia ES

¹¹ CIBER Epidemiología y Salud Pública Madrid ES

¹² Department of Neurology University Hospital Schleswig-Holstein and Kiel University Kiel DE

¹³ Center for the Study of Movement, Cognition, and Mobility Neurological Institute Tel Aviv Medical Center Tel Aviv IL

¹⁴ Department of Physical Therapy Faculty of Medical & Health Sciences and Sagol School of Neuroscience Tel Aviv University Tel Aviv IL

¹⁵ Rush Alzheimer's Disease Center, Rush University Medical Center, and Department of Orthopedic Surgery Rush University Medical Center Chicago US

¹⁶ Institute of Epidemiology and Medical Biometry Ulm University Ulm DE

¹⁷ IB University of Health and Social Sciences Study Centre Stuttgart Stuttgart DE

¹⁸ Department of Orthopaedic Surgery St. Olav's Hospital Trondheim NO

¹⁹ Department of Sport, Exercise, and Health University Basel Basel CH

²⁰ Department of Sport, Exercise and Rehabilitation Northumbria University Newcastle Newcastle Upon Tyne GB

²¹ Novartis Biomedical Research Novartis Pharma AG Basel CH

²² Department of Rehabilitation Science and Health Technology OsloMet Oslo NO

Corresponding Author:

Martin A. Berge

Department of Neuromedicine and Movement Science
Norwegian University of Science and Technology
Helgasetr, 4th Fl.
Vanglunds gate 2
Trondheim
NO

Abstract

Background: Algorithms estimating real-world digital mobility outcomes (DMOs) are increasingly validated in healthy adults and various disease cohorts. However, their accuracy and reliability in older adults after hip fractures, who often walk slowly for short durations, is underexplored.

Objective: The objective of this study was to examine DMO accuracy and reliability in a hip fracture cohort, considering walking bout (WB) duration, physical function, days since surgery, and walking aid usage.

Methods: Participants were monitored in their natural environment for 2.5 hours using a lower back single wearable device and a reference system combining inertial modules, distance sensors, and pressure insoles. DMO estimates from 164 WBs from 11 participants obtained from the single wearable device were assessed against the reference system on a WB and a participant level, including walking speed, cadence, stride length, stride duration, number of steps and walking distance. We stratified by WB duration (All WBs, WBs >10s, WBs 10-30s, WBs >30s), lower versus higher Short Physical Performance Battery (SPPB) scores, and observed whether days since surgery and walking aid usage affected DMO accuracy and reliability.

Results: Walking speed, cadence, stride duration, number of steps, and distance stratified by WB duration showed Intraclass Correlation Coefficients (ICCs) ranging from 0.50-0.99, and mean relative errors (MREs) ranging from -6.9% to 12.8%. Stride length showed poor reliability with ICCs ranging from 0.30-0.49 and MREs from 6.1% to 13.2%. Walking speed and distance ICCs in the higher SPPB score group ranged from 0.85-0.99 and MREs from -10.1% to -1.7%. In the lower SPPB score group, walking speed and distance ICCs ranged from 0.17-0.99 and MREs from 13.5% to 32.6%. We did not observe a discernible effect of time since surgery or walking aid usage on walking speed or distance estimates.

Conclusions: Five accurate and reliable real-world DMOs were identified in older adults after hip fractures: walking speed, cadence, stride duration, number of steps, and distance. Accuracy and reliability of most DMOs improved when excluding WBs <10 seconds and were higher in WBs >30 seconds compared to 10-30 seconds WBs and for those with higher physical function. DMOs can capture daily mobility as early as 1 month after surgery, also in people using walking aids. However, as most WBs in this cohort are of short duration, there is a trade-off between more accurate and reliable walking speed and distance estimates when excluding short-duration WBs, and the loss of a substantial amount of data. These results have important implications for future research aiming to establish clinical validity of DMOs, and enable clinicians to evaluate the effects of interventions on daily life mobility in detail, which in turn facilitates the design of optimal care pathways. Clinical Trial: ISRCTN – 12246987 <https://doi.org/10.1186/ISRCTN12246987>

(JMIR Preprints 23/10/2024:67792)

DOI: <https://doi.org/10.2196/preprints.67792>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in a JMIR journal, my full manuscript will be made available to all users.

No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript



Evaluating the Accuracy and Reliability of Real-World Digital Mobility Outcomes in Older Adults After Hip Fracture: A Cross-Sectional Study

Abstract

Background: Algorithms estimating real-world digital mobility outcomes (DMOs) are increasingly validated in healthy adults and various disease cohorts. However, their accuracy and reliability in older adults after hip fractures, who often walk slowly for short durations, is underexplored.

Objective: Examine DMO accuracy and reliability in a hip fracture cohort, considering walking bout (WB) duration, physical function, days since surgery, and walking aid use.

Methods: Nineteen community-dwelling participants were monitored in their natural environment for 2.5 hours using a lower back single wearable device and a reference system combining inertial modules, distance sensors, and pressure insoles. Six DMO estimates from 164 WBs from 11 participants (age 71-90, assessed 32-390 days after surgery, Short Physical Performance Battery (SPPB) score 3-12, and gait speed range 0.39-1.34 m/s), obtained from a single wearable device, were assessed against the reference system on the WB and the participant level. We stratified by WB duration (All WBs, WBs >10s, WBs 10-30s, WBs >30s) and lower versus higher SPPB scores, and observed whether days since surgery and walking aid use affected DMO accuracy and reliability.

Results: Across the 164 WBs, walking speed and distance ranged from 0.25-1.29 m/s and 1.7-436.5 meters. Estimation of walking speed, cadence, stride duration, number of steps, and distance stratified by WB duration showed Intraclass Correlation Coefficients (ICCs) ranging from 0.50-0.99, and mean relative errors (MREs) from -6.9% to 12.8%. Stride length estimation showed poor reliability with ICCs ranging from 0.30-0.49 and MREs from 6.1% to 13.2%. Walking speed and distance ICCs in the higher SPPB score group ranged from 0.85-0.99 and MREs from -10.1% to -1.7%. In the lower SPPB score group, walking speed and distance ICCs ranged from 0.17-0.99 and MREs from 13.5% to 32.6%. There was no discernible effect of time since surgery or walking aid use on walking speed or distance estimates.

Conclusions: Five accurate and reliable real-world DMOs were identified in older adults after hip fractures: walking speed, cadence, stride duration, number of steps, and distance. Accuracy and reliability of most DMOs improved when excluding WBs <10s and were higher in WBs >30s compared to 10-30s WBs, and for those with higher physical function. DMOs can capture daily gait as early as 1 month after surgery, also in people using walking aids. However, as most WBs in this cohort are short, there is a trade-off between improving accuracy and reliability of walking speed and distance estimates by excluding short WBs, and losing a substantial amount of data. These results have important implications for future research aiming to establish clinical validity of DMOs and enable clinicians to evaluate the effects of interventions on daily life gait in detail, thereby facilitating the design of optimal care pathways.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 12246987; <https://www.isrctn.com/ISRCTN12246987>

Keywords

Gait; Free-living; Digital mobility assessment; Wearable device; Walking speed; Distance; Cadence; Active aging;

Introduction

Hip fracture after a fall is one of the most serious injuries in older adults with a high rate of morbidity and a 1-year mortality of 22% [1, 2]. Hip fracture survivors experience a substantial decline in quality of life and mobility [3]. Mobility, the ability to move or walk around freely and easily, is essential for functioning well, maintaining independence, and ensuring social and emotional well-being [4]. Between 40 and 60 % of people with hip fractures do not recover their pre-fracture level of mobility and ability to perform instrumental activities of daily living [3], and the incidence of nursing home admissions is high [5]. Hip fractures require surgery to restore mobility, but there is large variation in the average length of stay at the hospital across countries, ranging from 4 to 40 days post-surgery [6, 7]. The main goals of care in the first days are typically pain management and early mobilisation, followed by rehabilitation targeting physical function and independent walking to prepare patients for returning to their daily life [8, 9]. However, the different recovery trajectories and the factors determining mobility outcomes for various individuals are poorly understood. The ability to walk and regain mobility is an important indicator of recovery, making it a critical focus in the rehabilitation of patients after a hip fracture. However, getting patients back on their feet requires more than just a step in the right direction. Comprehensive mobility assessment is key to supporting effective recovery, providing valuable information about patients' progress and informing rehabilitation strategies and optimal care pathways.

Assessing mobility requires objective, quantifiable metrics that accurately reflect a patient's ability to move around efficiently and independently. While mobility can be described through broader measures such as upright time and number of sit-to-stand transitions [10, 11], these primarily provide information about activity amount. In contrast, gait-related metrics provide information about the quality and efficiency of movement, reflecting musculoskeletal function, coordination and stability. Additionally, gait impairments are strong predictors of falls, disability and overall health, and inform about the status and progression of different health challenges [12-14], making walking a key aspect of mobility.

Gait after a hip fracture has mostly been assessed using patient-reported outcome measures that are prone to response bias, or via standardised walking tests in clinical settings and laboratory assessments [15]. Traditional supervised gait assessments with sophisticated technology in laboratory settings, such as gold-standard instrumented gait analysis, allow accurate spatiotemporal measurements of gait in a controlled laboratory environment [14, 16]. However, gait assessments under such conditions lack ecological validity for several reasons [17]. Laboratory assessments provide a snapshot in time, are limited by space and infrastructure, are prone to white-coat effects, and often consist of isolated and structured tasks that do not necessarily reflect what people with and without mobility impairment do in their daily lives [17-19]. Importantly, they may also limit the inclusion of people from rural areas, potentially overlooking a key demographic [17]. Additionally, in-lab assessments typically evaluate gait on even, uncluttered ground over short, straight distances only [15]. We therefore need more knowledge regarding gait in people's real-world environments. Additionally, there is a critical need to complement current short-term follow-up and infrequent clinical tests or assessments with more comprehensive and detailed knowledge about daily life gait recovery throughout the first years after a hip fracture [15]. Hence, extended monitoring of gait is vital for understanding the long-term recovery process and optimising intervention strategies [20].

However, assessing accurate gait characteristics in the real world is challenging due to internal and external confounding factors. Real-world gait is highly variable and complex as it entails navigating different walking surfaces, uneven terrain and obstacles, and consists of a wide variety of related activities such as turning, stopping and starting [21]. Methods based on the use of a single inertial measurement unit (IMU) on the lower back are available to assess gait in the real world, through the quantification of digital mobility outcomes (DMOs) [20]. But until recently, algorithms to estimate DMOs lacked comprehensive and systematic validation [22].

The Mobilise-D consortium [23] has made significant strides to change this. Mobilise-D has used a rigorous and comprehensive validation process, including a multisensor wearable set-up for real-world analysis [24] and involving 1 healthy and 5 patient cohorts performing an extensive technical validation protocol [22, 25]. This resulted in the identification and refinement of algorithms that enable robust gait sequence detection and subsequent estimation of key DMOs from a single inertial device worn on the lower back, such as initial foot contact, cadence, stride length, and walking speed. Previous Mobilise-D work showed that these DMOs can be estimated accurately and reliably in the real world across a range of cohorts (Healthy Older Adults, Parkinson's Disease, Multiple Sclerosis, Congestive Heart Failure, Chronic Obstructive Pulmonary Disease, and Hip Fracture), tasks and contextual factors [26, 27]. Analyses of all 6 Mobilise-D cohorts showed that the real-world algorithm performances were valid and accurate but more challenging in the case of short walking bouts (WBs) (<10s) and slower gait speeds (<0.5 m/s) [26, 27], typical hallmarks of gait after a hip fracture. Analyses stratified by cohort indicated that the estimated walking speed and cadence in the hip fracture cohort showed moderate reliability and mean relative error (MRE) of 10% and -0.5%, respectively, and estimated stride length with poor reliability and MRE of 11% [26]. However, these analyses did not consider hip fracture-specific factors that may affect the reliability and accuracy of the DMO estimates, which may be of critical relevance for the hip fracture cohort.

Building upon the earlier Mobilise-D publications [26, 27], this paper investigates additional DMOs and delves deeper into potential factors that may affect the reliability and accuracy of the real-world DMO estimates in the hip fracture cohort. As short WBs are frequent after a hip fracture [15] and walking distance can influence spatiotemporal parameters [28], we investigate whether different WB duration categories affect the estimates. Furthermore, physical functioning is associated with variations in gait parameters and is highly compromised after a hip fracture [13, 29], particularly in the early phases of recovery [30], and often necessitates the use of a walking aid [31].

Hence, this study investigates the following research question: Can the validated Mobilise-D real-world DMOs be accurately and reliably estimated in older adults after hip fracture, considering their level of physical function and potential walking aid usage, across the first year after surgery? To answer, we compared 6 DMOs estimated from a single wearable device against a reference [24, 32] and investigated whether WB duration, physical function, walking aid usage, and days since surgery affect the accuracy and reliability of the DMO estimates.

Given earlier results, we expect that most real-world DMOs can be estimated accurately and reliably in the hip fracture cohort and that results will improve further when excluding WBs with short duration. Additionally, as shorter WBs tend to have higher errors [26, 27], we expect that DMO accuracy and reliability may be lower in older adults with low physical function compared to those with better physical functioning. Moreover, as gait function is especially impaired early after a hip fracture, the accuracy and reliability of DMO estimates may be lower in the earlier stages of recovery compared to later stages.

Methods

Study Design and Participants

This multicentre observational study utilized data from the technical validation study (TVS) of the IMI2-JU-funded Mobilise-D project [22, 23, 25]. Data were collected between July 2020 and March 2022. The hip fracture participants were recruited from the Robert Bosch Foundation for Medical Research (Germany) and the University of Kiel (Germany), and the protocol was approved by the Medical Faculties of the University of Tübingen (647/2019BO2) and the University of Kiel (D540/19). All participants gave their written informed consent.

The participants were recruited within 13 months of surgical treatment (fixation or arthroplasty) for a low-energy fracture of the proximal femur (ICD-10 diagnosis S72.0, S72.1, S72.2), as diagnosed by

X-rays of the hip and pelvis.

Participants had to be 65 years or older to be included. Participants were excluded if they were unable to walk 4 meters independently with or without a walking aid, had a shoe size <36 European size, a Montreal Cognitive Assessment (MoCA) [33] score of ≤ 15 , or had an occurrence of any of the following within 3 months before inclusion: myocardial infarction, hospitalisation for unstable angina, stroke, coronary artery bypass graft, percutaneous coronary intervention, or implantation of a cardiac resynchronisation therapy device [22]. We aimed to recruit 20 participants. Due to challenges posed by COVID-19, recruitment was much slower than anticipated and further hampered by not being allowed to recruit participants during their hospital stay. As a result, we needed to almost double the recruitment period to be able to enrol 19 participants. Of these, 6 participants had to be excluded because of technological issues with the pressure insoles of the reference system. Data from the reference system from 2 additional participants had insufficient signal quality. These 2 participants were relatively young (70 and 66 years old), had a very low SPPB score of 3, and were assessed 28 and 64 days after surgery. Data from the remaining 11 participants were included in the final analysis.

Protocol and Equipment

Participants' activities were monitored for 2.5-hours in a real-world setting within their preferred habitual environment (home, work, or community). The activities were unstructured but to ensure sufficient variability in the collected data, participants were encouraged to complete several activities such as rising from a chair and walking to another room, walking up and down a flight of stairs, walking to the kitchen and making a drink, walking outdoors (if possible, minimum 2 min) and, if walking outside, walk up and down an inclined path [22].

Gait data were collected with a single wearable device (McRoberts DynaPort MM+, 100 Hz sampling frequency, ± 8 g triaxial acceleration range and 1mg resolution, ± 2000 degrees per second triaxial gyroscope range with 70 mdps resolution) worn at the lower back with a Velcro belt. Additionally, participants were equipped with a multisensor reference system named INDIP (INertial module with DIstance sensors and Pressure insoles, sampling frequency of 100 Hz) [22, 24, 32, 34, 35], previously validated with excellent reliability in different cohorts including hip fracture across a complex set of different motor tests including simulated daily activities [24]. Specifically, 2 magneto-IMUs were fixed to the instep using clips. A 3rd IMU was attached to the lower back with Velcro. Asymmetrically positioned distance sensors were fixed above the ankles with Velcro, and 2 pressure insoles were inserted into the shoes. To ensure synchronization, timestamps of the reference system and the single wearable device were aligned (± 10 ms).

Evaluation of Digital Mobility Outcomes (DMOs)

The evaluation of the DMOs was based on previous Mobilise-D work that selected the top-ranked algorithms to detect gait sequences, and to estimate initial contact events, stride length, and cadence within identified gait sequences [27]. The best-performing cadence and stride length algorithms for hip fracture were then used to estimate walking speed [26]. The validation of this processing pipeline, based on a minimum 95% CI ICC threshold for performance metrics (i.e., sensitivity, positive predictive value, accuracy etc.) of at least 0.7, and a relative error of less than 20%, is described in Mico-Amigo et al. [27] and Kirk et al. [26]. All algorithms are available in the mobgap Python library [36].

DMOs were evaluated on a walking bout (WB) level. A WB was defined as a continuous walking sequence comprising a minimum of 2 consecutive strides of both feet [37]. WBs were separated by breaks greater than 3 seconds, and for a stride to be included it required a duration of 0.2-3 seconds and a minimum length of 0.15 m [27]. These criteria were applied to generate WBs for both the

single wearable device and the reference system by initially filtering the identified strides according to the stride level definition and then assembling them into final WBs by identifying breaks in the stride sequence. Final DMOs for both systems were calculated as the average value over all strides within a WB. For a rigorous comparison of DMOs on a WB level, it was essential to focus on WBs concurrently detected by both systems, employing a true positive analysis approach. Accordingly, we considered WBs with a time overlap exceeding 80% of their duration as true positives, as detailed in Kirk et al. [26].

For each WB, 6 gait characteristics were obtained from the single wearable device and the reference system: cadence (steps/min; number of steps taken per minute), stride length (meters; length of 2 consecutive steps), number of steps (n), stride duration (s) [27], walking speed (m/s) [26], and distance (meters). The walked distance was calculated by multiplying 2 validated DMOs [27]: average walking speed * WB duration.

Variables

Participant Characteristic

Participant characteristics were collected for all participants, including age, height, weight, sex, cognitive function (MoCA), fracture type and surgical implant. Pain while walking was assessed using a Visual Analogue Scale (from 0-100, where 0 is no pain, and 100 is worst pain imaginable).

WB Duration

Given the larger single wearable device DMO errors observed in shorter WBs [26, 27], we assessed whether the accuracy and reliability of DMO estimates differed when excluding WB durations shorter than 10 seconds [38, 39]. Also, we divided WBs >10 seconds into 2 subcategories: WBs of 10-30 seconds and WBs >30 seconds. Thus, we analysed the DMOs in 4 categories of WB duration: All WBs combined, WBs > 10 seconds, WBs 10-30 seconds, and WBs >30 seconds. This was performed both on a WB level and a participant level.

Physical Function

The participant's physical function was assessed using the Short Physical Performance Battery (SPPB) +/- 7 days before or after the 2.5-hour activity monitoring. The SPPB consists of a static balance test, a 5-time chair rise test, and a 4-meter walk test (4MWT) at a comfortable gait speed. The total SPPB score ranges from 0-12, where higher scores indicate better mobility capacity and a score of < 8 points indicates impaired Activities of Daily Living functions [40]. Due to the larger DMO errors in cohorts with more impaired gait and slower walking speeds [26, 27], we divided participants into 2 physical function groups based on the 8-point SPPB threshold: a Lower SPPB group consisting of participants with a total score between 0-7, and a Higher SPPB group consisting of participants with a total score between 8-12. This threshold resulted in 2 groups of 5 and 6 participants, respectively.

Time Since Surgery and Walking Aid Usage

Physical function and mobility typically improve gradually in the first year after a hip fracture [15, 41, 42]. We collected the number of days from the surgery to the 2.5-hour free-living testing date, aggregated participant DMOs, and sorted them by number of days since surgery. We also collected information on walking aid usage (yes/no, and type of walking aid).

Statistical Analyses

DMOs were evaluated on a WB level to quantify DMO estimation errors (accuracy) and reliability using the validation metrics detailed in previous Mobilise-D work [26, 27]. For accuracy, absolute agreement was assessed by calculating the mean error, mean absolute error (MAE) and precision

(limits of agreements, LoA) [43] between DMO estimates from the single wearable device (SWD) and the reference system (RS). Mean relative errors (MRE) and mean absolute relative errors (MARE) were calculated by dividing the (absolute) errors per WB by the corresponding estimates from the reference system, expressed in percentage:

- Mean Error = $(1 / n) * \sum \text{from } i=1 \text{ to } n [(DMO_SWD_i - DMO_RS_i)]$
- MRE = $(1 / n) * \sum \text{from } i=1 \text{ to } n [(DMO_SWD_i - DMO_RS_i) / DMO_RS_i] * 100]$
- MAE = $(1 / n) * \sum \text{from } i=1 \text{ to } n [|DMO_SWD_i - DMO_RS_i|]$
- MARE = $(1 / n) * \sum \text{from } i=1 \text{ to } n [|DMO_SWD_i - DMO_RS_i| / DMO_RS_i] * 100]$

For reliability, Intraclass Correlation Coefficients (ICC^(2,1)) [44] were computed to evaluate how closely each of the DMOs of the 2 systems were related. ICC values below 0.5 were considered poor, between 0.5 and 0.75 moderate, between 0.75 and 0.9 good, and above 0.90 excellent [45].

For analyses stratified by the WB duration category, all 6 DMOs were included. In analyses stratified by the 2 SPPB score groups, we included the 2 most relevant DMOs for clinicians and patients: walking speed [15] and distance [46].

Walking speed and distance were also evaluated on a participant level. Histograms, QQ plots, and the Anderson-Darling tests were employed for both variables to assess normality. Given that the DMOs were not normally distributed and the relatively low number of participants, median DMO values were computed across all WBs for each participant for the single wearable device and the reference system. The median DMO values from both systems were visualized in bar graphs with interquartile ranges.

Data preparation and visualisation were conducted in Matlab® R2022a (Mathworks, Natick, MA), while statistical analyses were performed in Stata/MP 18.0 (StataCorp, College Station, TX).

Results

Participant Characteristics

The 11 participants included in the analyses (6 men and 5 women) were community-dwelling, and their characteristics are outlined in Table 1. Notably, the study sample included observations from older adults with hip fractures assessed from 32 days to 390 days after surgery. Across the 164 WBs, walking speed and distance ranged from 0.25-1.29 m/s and 1.7-436.5 meters as measured by the reference system, and 36 (22%) of these WBs were slower than 0.5 m/s. Of the included participants, 9 (82%) had 2 or more WBs slower than 0.5 m/s, and 2 used a walking aid during the observation period, both using a single cane or crutch. Five participants had trochanteric fractures and were treated with intramedullary nails, while 6 participants had cervical fractures, of which 4 received a hemiprosthesis and 2 received a total prosthesis.

Table 1. Characteristics of individual participants and the overall sample.

Participant	1	2	3	4	5	6	7	8	9	10	11	All Median (IQR)
Days since surgery	32	39	60	114	141	179	193	200	244	369	390	179 (60-244)
Age, years	71	76	80	84	79	83	87	72	71	83	90	80 (72-84)
Height, cm	180	182	169	180	158	159	174	174	158	165	174	173.5 (159-180)
Weight, kg	78	85	66	69	54	44	71	95	53	66	96	68.5 (54-85)
MoCA ^a score	26	30	27	28	30	18	25	30	22	21	19	26 (21-30)

Walking pain	5	3	2	4	0	30	19	4	52	2	1	4 (2-19)
SPPB^b score	9	4	3	9	10	7	11	10	6	12	4	9 (4-10)
Gait speed 4MWT^c, m/s	0.79	0.39	0.41	0.89	0.80	0.83	1.34	1.05	0.75	1.08	0.62	0.80 (0.69-0.97)
WBs < 0.5 m/s, n (%)^d	2 (14)	0 (0)	10 (83)	3 (14)	2 (7)	3 (25)	2 (14)	3 (18)	8 (23)	0 (0)	3 (30)	
Gender, M/F^e	M	M	F	M	F	F	M	M	F	F	M	
Fracture, C/T^f	C	C	C	C	T	T	T	C	T	T	C	
Implant, H/N/TP^g	H	TP	TP	H	N	N	N	H	N	N	H	

^aMoCA: Montreal Cognitive Assessment

^bSPPB: Short Physical Performance Battery

^c4MWT: 4-meter walk test

^dNumber of matched walking bouts with a speed below 0.5 m/s as measured by the reference system

^eM: Male; F: Female

^fC: Cervical; T: Trochanteric

^gH: Hemiprosthesis; N: Nail; TP: Total Prosthesis

WB Duration

In total, 164 WBs met our true positive approach [26], 65 (40%) of which were shorter than 10 seconds, 60 (36%) between 10-30 seconds, and 39 (24%) above 30 seconds. Estimations of the number of steps and distance demonstrated good to excellent reliability in all WB duration categories with ICCs exceeding 0.83, MREs ranging from -6.87% to 5.34%, and MAREs ranging from 4.96% to 19.79% (Table 2). Walking speed and stride duration showed moderate reliability in All WBs combined and in WBs >10 seconds, with ICCs ranging from 0.64 to 0.71, MREs ranging from -5.68% to 10.35%, and MAREs ranging from 7.51% to 19.75%. Walking speed showed good reliability and less error in longer WBs >30 seconds (ICC = 0.80, MRE = 5.09%, MARE = 14.50%) compared to shorter 10-30 seconds WBs (ICC = 0.50, MRE = 12.78%, MARE = 20.81%). Similarly, stride duration showed good reliability and less error in WBs >30 seconds (ICC = 0.77, MRE = -4.98%, MARE = 5.30%) compared to 10-30 seconds WBs (ICC = 0.54, MRE = -6.13%, MARE = 8.94%). Cadence showed moderate to excellent reliability in all WB categories, with ICCs ranging from 0.74 to 0.98, low MREs ranging from -0.56% to -0.34%, and MAREs ranging from 1.75% to 5.62%. In contrast, stride length showed poor reliability, with ICCs ranging from 0.30 to 0.49, with an overestimation in MRE varying from 6.05% to 13.23%, and MAREs ranging from 15.46% to 20.61%.

Table 2. DMO estimates from the reference system (RS) and single wearable device (SWD), mean errors and mean relative errors (MRE) with Limits of Agreement (LoA), mean absolute errors (MAE) and mean absolute relative errors (MARE), and Intraclass correlation coefficient (ICC) for All WBs, WBs >10 seconds, the latter also split in WBs 10-30 seconds and WBs >30 seconds.

WB Duration ^a	RS	SWD	Error with LoA	MRE with LoA (%)	MAE	MARE (%)	ICC (2,1) ^b
Walking Speed, m/s							
All ^c	0.65 [0.37,	0.69 [0.45,	0.03 [-0.25,	10.35 [-50.12,	0.11 [0.01,	19.75 [1.80,	0.67 [0.58,

	1.03]	0.96]	0.31]	70.82]	0.33]	66.32]	0.73]
>10s	0.68 [0.39, 1.05]	0.71 [0.44, 1.02]	0.03 [-0.24, 0.30]	9.75 [-53.53, 73.03]	0.10 [0.01, 0.33]	18.32 [1.27, 66.32]	0.71 [0.62, 0.79]
10-30s	0.61 [0.38, 0.98]	0.66 [0.43, 0.89]	0.05 [-0.23, 0.32]	12.78 [-56.61, 82.16]	0.11 [0.01, 0.32]	20.81 [1.74, 65.57]	0.50 [0.32, 0.64]
>30s	0.77 [0.46, 1.19]	0.78 [0.45, 1.10]	0.01 [-0.25, 0.26]	5.09 [-46.99, 57.16]	0.09 [0.01, 0.39]	14.50 [1.11, 79.76]	0.80 [0.69, 0.88]
Cadence, steps/min							
All ^c	89.81 [64.25, 114.21]	88.52 [65.76, 110.95]	-1.29 [-21.79, 19.21]	-0.56 [-19.50, 18.39]	5.47 [0.15, 17.50]	5.62 [0.16, 19.43]	0.74 [0.68, 0.80]
>10s	88.28 [62.32, 111.14]	87.42 [62.90, 104.36]	-0.86 [-14.73, 13.02]	-0.39 [-13.81, 13.04]	3.58 [0.08, 12.36]	3.84 [0.08, 13.61]	0.86 [0.80, 0.89]
10-30s	88.60 [64.83, 110.10]	87.52 [64.83, 102.66]	-1.08 [-18.31, 16.15]	-0.42 [-17.01, 16.17]	4.89 [0.16, 15.30]	5.20 [0.20, 15.89]	0.75 [0.64, 0.83]
>30s	87.77 [58.29, 111.48]	87.26 [59.51, 108.71]	-0.52 [-6.42, 5.39]	-0.34 [-6.45, 5.77]	1.57 [0.02, 10.59]	1.75 [0.03, 10.36]	0.98 [0.96, 0.99]
Stride Length, m							
All ^c	0.88 [0.54, 1.24]	0.93 [0.72, 1.19]	0.05 [-0.33, 0.43]	11.42 [-49.51, 72.35]	0.15 [0.01, 0.41]	20.40 [1.38, 64.86]	0.45 [0.34, 0.55]
>10s	0.91 [0.57, 1.29]	0.96 [0.72, 1.26]	0.05 [-0.33, 0.42]	10.40 [-49.91, 70.71]	0.14 [0.01, 0.43]	18.58 [1.03, 64.86]	0.49 [0.36, 0.61]
10-30s	0.83 [0.56, 1.23]	0.90 [0.71, 1.11]	0.07 [-0.30, 0.43]	13.23 [-47.34, 73.80]	0.14 [0.01, 0.42]	20.61 [1.28, 63.57]	0.41 [0.21, 0.57]
>30s	1.04 [0.60, 1.38]	1.06 [0.89, 1.31]	0.02 [-0.38, 0.42]	6.05 [-53.63, 65.73]	0.13 [0.01, 0.72]	15.46 [0.80, 119.91]	0.30 [0.04, 0.53]
Steps, n							
All	49.77 [6.00, 266.00]	47.00 [5.00, 233.00]	-2.77 [-18.78, 13.23]	-5.62 [-31.00, 19.76]	3.41 [0.00, 10.00]	10.39 [0.00, 30.00]	0.99 [0.99, 1.00]
>10s	76.32 [13.00, 317.00]	72.10 [13.00, 306.00]	-4.22 [-24.21, 15.77]	-5.80 [-25.88, 14.28]	4.87 [0.00, 21.00]	8.61 [0.00, 25.81]	0.99 [0.98, 1.01]
10-30s	22.42 [13.00, 38.50]	20.62 [12.50, 37.00]	-1.80 [-7.52, 3.92]	-6.87 [-30.79, 17.06]	2.50 [0.00, 8.00]	10.98 [0.00, 29.89]	0.90 [0.79, 0.95]
>30s	159.26 [40.00, 479.00]	151.31 [36.00, 423.00]	-7.95 [-37.77, 21.87]	-4.17 [-15.72, 7.39]	8.51 [0.00, 56.00]	4.96 [0.00, 16.86]	0.99 [0.98, 1.00]

Stride Duration, s							
All	1.41 [1.06, 1.88]	1.33 [1.03, 1.65]	-0.08 [-0.40, 0.24]	-4.65 [-25.80, 16.50]	0.12 [0.00, 0.41]	8.15 [0.15, 26.82]	0.70 [0.55, 0.79]
>10s	1.43 [1.09, 1.95]	1.33 [1.08, 1.65]	-0.10 [-0.41, 0.22]	-5.68 [-24.00, 12.64]	0.12 [0.00, 0.42]	7.51 [0.16, 24.25]	0.64 [0.41, 0.77]
10-30s	1.42 [1.10, 1.91]	1.32 [1.09, 1.63]	-0.10 [-0.45, 0.24]	-6.13 [-27.08, 14.82]	0.14 [0.01, 0.42]	8.94 [0.52, 25.02]	0.54 [0.29, 0.70]
>30s	1.43 [1.09, 2.07]	1.34 [1.08, 1.67]	-0.09 [-0.35, 0.18]	-4.98 [-18.43, 8.47]	0.09 [0.00, 0.42]	5.30 [0.04, 20.19]	0.77 [0.54, 0.88]
Distance, m							
All	25.75 [2.36, 125.62]	24.22 [2.44, 119.63]	-1.53 [-22.16, 19.11]	4.91 [-53.08, 62.90]	3.75 [0.06, 15.48]	19.79 [0.92, 65.29]	0.98 [0.97, 0.98]
>10s	40.08 [5.40, 203.72]	37.50 [5.04, 168.91]	-2.58 [-28.93, 23.77]	3.14 [-54.18, 60.46]	5.66 [0.07, 26.48]	17.72 [0.86, 58.11]	0.98 [0.97, 0.98]
10-30s	9.62 [4.84, 19.38]	9.87 [4.69, 19.41]	0.25 [-4.61, 5.12]	5.34 [-56.44, 67.13]	1.71 [0.05, 5.55]	19.53 [0.86, 57.10]	0.83 [0.74, 0.89]
>30s	86.93 [15.61, 253.43]	80.00 [20.80, 233.73]	-6.93 [-47.31, 33.44]	-0.25 [-49.97, 49.47]	11.74 [0.32, 59.36]	14.93 [0.78, 79.93]	0.96 [0.94, 0.98]

^aValues are either provided as mean and [5%, 95%] quantiles or as mean and limit of agreement if indicated by LoA. If not indicated otherwise, values are expressed in DMO units.

^bThe colours indicate ICC level-categories: Dark green = excellent (>0.90), green = good (0.75-0.90), light green = moderate (0.50-0.75), and grey = poor (<0.50).

^cAs reported in Kirk et al. [26].

Physical Function

The reliability of the walking distance estimates was excellent for both SPPB groups and all WB duration categories (ICCs ≥ 0.94), except for WBs 10-30 seconds in the lower SPPB score group (ICC = 0.63) which showed moderate reliability (Table 3). Distance was overestimated in the lower SPPB score group in All WBs (MRE = 20.05%, MARE = 26.83%) and WBs >10 seconds (MRE = 20.34%, MARE = 26.30%), and the error was lower for longer WBs >30 seconds (MRE = 13.52%, MARE = 20.43%) compared to shorter WBs 10-30 seconds (MRE = 24.13%, MARE = 29.56%). Conversely, the distance was underestimated in the higher SPPB score group in All WBs (MRE = -6.37%, MARE = 14.55%) and WBs >10 seconds (MRE = -9.53%, MARE = 11.39%), and the error was lower for WBs >30 seconds (MRE = -8.85%, MARE = 11.49%) compared to WBs 10-30 seconds (MRE = -10.03%, MARE = 11.32%).

In the higher SPPB score group, walking speed reliability was good in all WB duration categories (ICC ≥ 0.85), with a slightly underestimated walking speed, MRE ranging from -1.73% to -3.59% (MARE reaching from 8.62% to 11.15%). Conversely, the lower SPPB score group showed poor walking speed reliability and higher errors in all WBs combined, WBs >10 seconds, and WBs 10-30s, with ICCs ranging from 0.17 to 0.44, MREs reaching from 26.58% to 32.56% and MAREs reaching from 31.22% to 35.71%. Notably, the lower SPPB score group had a moderate walking

speed reliability in WBs >30 seconds (ICC = 0.66), with a MRE of 18.96% and MARE of 23.14%.

Table 3. DMO estimates for the higher (n=6) and the lower (n=5) SPPB groups from the reference system (RS) and single wearable device (SWD), mean errors and mean relative errors (MRE) with Limits of Agreement (LoA), mean absolute errors (MAE) and mean absolute relative errors (MARE), and Intraclass correlation coefficient (ICC).

SPPB ^a Group (score)	WB Duration ^b	RS	SWD	Error with LoA	MRE with LoA (%)	MAE	MARE (%)	ICC (2,1) ^c
Walking Speed, m/s								
Lower (0-7)	All	0.59 [0.34, 0.91]	0.70 [0.43, 0.94]	0.12 [-0.19 , 0.42]	26.58 [-49.26 , 102.41]	0.15 [0.02, 0.39]	31.30 [2.76, 95.98]	0.44 [0.14, 0.63]
	>10s	0.56 [0.34, 0.91]	0.68 [0.43, 0.92]	0.12 [-0.18 , 0.41]	27.70 [-54.84 , 110.25]	0.14 [0.02, 0.39]	31.22 [2.76, 95.98]	0.42 [0.01, 0.64]
	10-30s	0.53 [0.32, 0.66]	0.67 [0.42, 0.87]	0.14 [-0.16 , 0.43]	32.56 [-54.61 , 119.72]	0.16 [0.02, 0.35]	35.71 [4.02, 95.98]	0.17 [0.00, 0.42]
	>30s	0.63 [0.34, 1.04]	0.71 [0.44, 1.05]	0.08 [-0.21 , 0.38]	18.96 [-54.21 , 92.13]	0.11 [0.01, 0.44]	23.14 [0.77, 127.33]	0.66 [0.33, 0.85]
Higher (8-12)	All	0.70 [0.39, 1.05]	0.68 [0.48, 0.97]	-0.03 [-0.21, 0.15]	-1.73 [-30.26 , 26.80]	0.08 [0.01, 0.19]	11.15 [1.24, 29.51]	0.86 [0.80, 0.90]
	>10s	0.76 [0.49, 1.19]	0.72 [0.47, 1.08]	-0.04 [-0.20, 0.12]	-3.49 [-23.62 , 16.65]	0.07 [0.01, 0.19]	8.82 [1.24, 22.25]	0.89 [0.80, 0.93]
	10-30s	0.68 [0.44, 1.00]	0.65 [0.46, 0.89]	-0.03 [-0.18, 0.12]	-3.41 [-23.29 , 16.47]	0.06 [0.01, 0.16]	8.62 [1.27, 22.25]	0.85 [0.73, 0.91]
	>30s	0.86 [0.63, 1.19]	0.81 [0.65, 1.10]	-0.05 [-0.22, 0.13]	-3.59 [-24.49 , 17.32]	0.08 [0.01, 0.21]	9.09 [1.24, 18.61]	0.87 [0.72, 0.93]
Distance, m								
Lower (0-7)	All	22.80 [1.98, 125.62]	24.31 [2.30, 130.75]	1.50 [-9.72 ,	20.05 [-49.57 ,	3.10 [0.07, 15.48]	26.83 [0.93, 93.85]	0.99 [0.99, 0.99]

		]	]	12.72]	89.67]			
	>10s	35.68 [5.55, 125.63]	37.85 [5.75, 145.84]	2.16 [-12.1 4, 16.46]	20.34 [-52.06 , 92.75]	4.57 [0.15, 20.23]	26.30 [0.86, 90.93]	0.99 [0.98, 0.99]
	10-30s	8.79 [4.31, 16.80]	10.66 [5.04, 16.80]	1.87 [-3.29 , 7.02]	24.13 [-50.53 , 98.80]	2.31 [0.07, 6.03]	29.56 [0.86, 90.93]	0.63 [0.29, 0.81]
	>30s	84.09 [14.35, 237.84]	86.79 [27.33, 233.73]	2.70 [-20.7 0, 26.10]	13.52 [-55.05 , 82.09]	8.65 [0.32, 26.48]	20.43 [0.33, 107.91]	0.99 [0.96, 0.99]
Higher (8-12)	All	27.94 [2.56, 139.68]	24.16 [2.49, 119.63]	-3.78 [-28.41, 20.85]	-6.37 [-40.14 , 27.41]	4.24 [0.04, 20.05]	14.55 [0.87, 36.45]	0.97 [0.96, 0.98]
	>10s	43.32 [5.40, 203.72]	37.25 [4.92, 168.91]	-6.07 [-36.94, 24.80]	-9.53 [-30.31 , 11.24]	6.46 [0.06, 34.81]	11.39 [0.87, 25.97]	0.97 [0.95, 0.98]
	10-30s	10.30 [5.29, 21.23]	9.23 [4.46, 22.02]	-1.07 [-3.66, 1.53]	-10.03 [-30.7 5, 10.69]	1.21 [0.04, 4.03]	11.32 [0.87, 28.68]	0.94 [0.80, 0.97]
	>30s	88.71 [21.79, 253.43]	75.76 [20.80, 194.07]	-12.95 [-57.4 8, 31.58]	-8.85 [-30.07 , 12.37]	13.67 [0.56, 59.36]	11.49 [0.92, 23.42]	0.96 [0.89, 0.98]

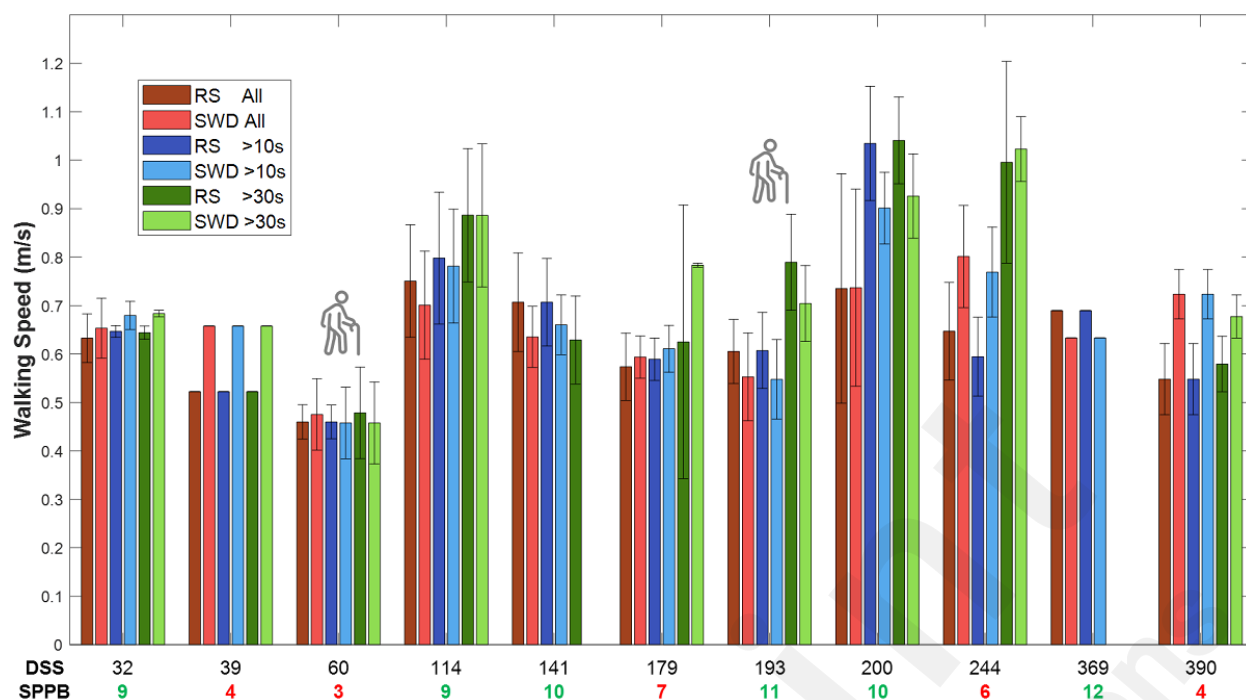
^aShort Physical Performance Battery

^bValues are either provided as mean and [5%, 95%] quantiles or as mean and limit of agreement if indicated by LoA. If not indicated otherwise, values are expressed in DMO units.

^cThe colours indicate ICC level-categories: Dark green = excellent (>0.90), green = good (0.75-0.90), light green = moderate (0.50-0.75), and grey = poor (<0.50).

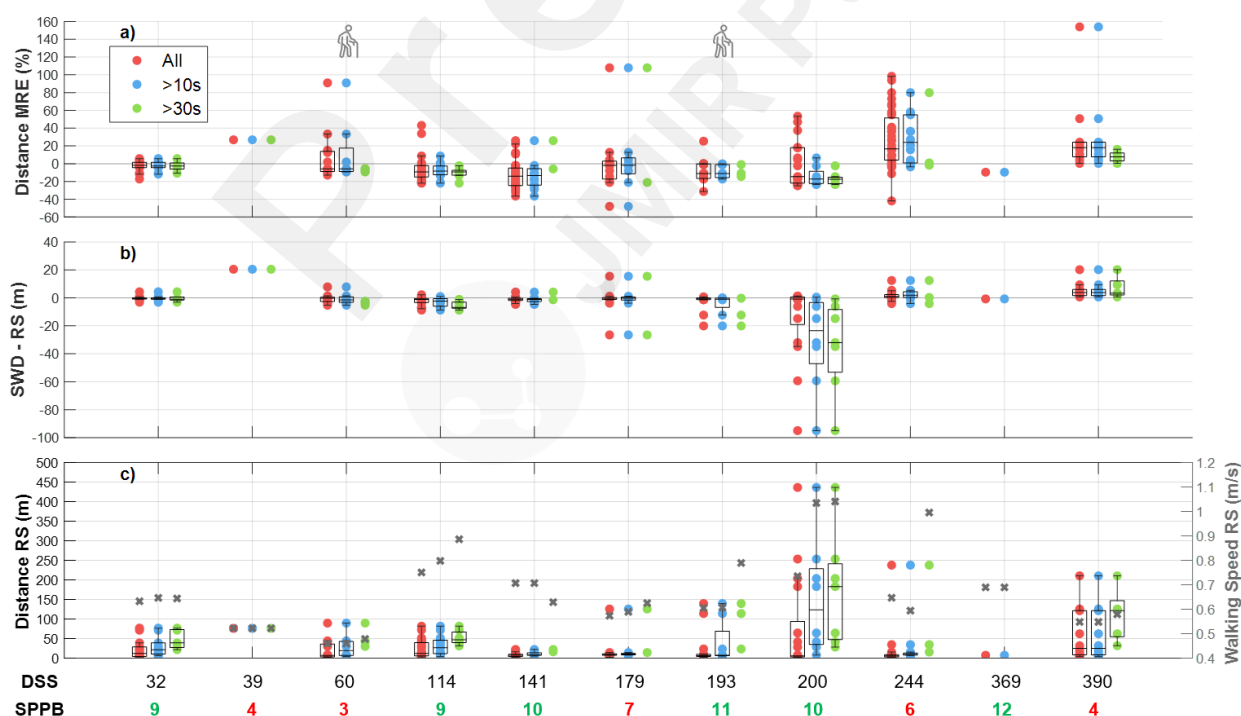
Time Since Surgery and Walking Aid Usage

To explore whether time since surgery influenced DMO accuracy, we plotted the participant's walking speed and distance estimates from both systems, sorted by the number of days since surgery which ranged from 32 to 390 days. As seen in Figures 1 and 2, there is no evident pattern in DMO accuracy across the number of days since surgery for any of the WB duration categories. Similarly, no clear deviation was found for the 2 participants who used a walking aid.



Figure

1. Walking speed estimates from the reference system (RS) and single wearable device (SWD) for each participant for All WBs combined (red bars), WBs above 10 seconds (blue bars), and WBs above 30 seconds (green bars). See Table 4 below for the number of WBs per category for each participant. The bars indicate the median and the error bars the interquartile range. Participants are sorted by days since surgery (DSS). The total Short Physical Performance Battery (SPPB) scores for each participant are listed below the x-axis, scores in green are associated with medium/high functioning (8-12), and scores in red with impaired activities of daily living functions (0-7). The 2 walking aid users are marked with an icon.



Figure

2. Mean relative error (a) and error (b) of the single wearable device (SWD) distance estimates and distance and median walking speed as measured by the reference system (RS) (c) for each participant for All WBs combined (red markers), WBs above 10 seconds (blue markers), and WBs above 30

seconds (green markers). Boxplots are shown where there are ≥ 5 data points, and indicate the median, interquartile range, and whiskers showing the minimum and maximum values within 1.5 times the interquartile range. Participants are sorted by days since surgery (DSS). The total Short Physical Performance Battery (SPPB) scores for each participant are listed. Scores in green indicate medium/high functioning (8-12), while scores in red indicate the onset of impaired activities of daily living functions (0-7). Walking aid users are marked with an icon.

Table 4. The number of true positive Walking Bouts (WBs) in each WB duration category, for all participants in total, per SPPB score group, and per participant sorted by days since surgery.

WB Duration ^a	WB n	SPPB ^b		Days since surgery										
		L ^c	H ^d	32	39	60	114	141	179	193	200	244	369	390
All	164	70	94	14	1	12	21	27	12	14	17	35	1	10
>10s	99	42	57	10	1	8	15	15	9	8	8	14	1	10
10-30s	60	27	33	5	0	4	8	13	7	5	1	11	1	5
>30s	39	15	24	5	1	4	7	2	2	3	7	3	0	5

^aAs described in the methods, WBs shorter than 10 seconds (n=65) are included in the “All WBs” category and are not reported separately.

^bSPPB: Short Physical Performance Battery

^cL: Lower

^dH: Higher

Discussion

This paper is the first to investigate critical hip fracture-specific factors that potentially influence the accuracy and reliability of validated DMOs, building on the rigorous methods and data from the Mobilise-D TVS [22, 26, 27]. We compared 6 real-world DMO estimates from a single wearable device against a validated multisensor reference system [24], and investigated factors that may affect the accuracy and reliability of validated DMOs in older adults after a hip fracture: WB duration, physical function, time since surgery, and walking aid usage. Overall, 5 of the DMOs (walking speed, cadence, stride duration, number of steps, and distance) showed moderate to excellent accuracy and reliability across WB durations. The 6th DMO (stride length) showed poor accuracy and reliability in this hip fracture cohort. Furthermore, walking speed and distance were more accurate and reliable in participants with better physical function. We did not observe a discernible effect of time since surgery or walking aid usage on the walking speed and distance estimates. These results indicate that real-world gait assessment is feasible in a hip fracture cohort, and that gait characteristics can be estimated accurately and reliably from as early as 1 month after hip fracture surgery.

Impact of WB Duration

Previous Mobilise-D studies that validated DMOs across 6 cohorts showed generally lower algorithm performances when including very short WBs (less than 10s) and for cohorts walking slowly and relying more on walking aids, such as many older adults after hip fracture [26, 27]. As these previous studies did not perform cohort-specific analyses stratified by WB duration, this study further examined the hip fracture cohort, following the true-positive evaluation of DMO accuracy and reliability on a WB level as outlined in Kirk et al. [26] and Micó-Amigo et al. [27]. As expected, excluding shorter WBs increased the accuracy and reliability of walking speed, cadence, steps, stride duration and distance, with WBs longer than 30 seconds having the highest accuracy and reliability. These lower errors are likely due to longer WBs capturing more continuous, consistent gait [38]. In

addition, longer WB durations likely correspond to outdoor walks, which are generally not as slow and intermittent as indoor walking [38, 39]. For short-duration WBs, contextual factors and turns might be more prominent [21], likely leading to higher DMO errors. However, as patients after a hip fracture typically walk in short bouts, excluding these would result in the loss of a potentially large amount of data, jeopardizing the representativeness of the data.

In addition to slow gait in short bouts, the complexity and variability of real-world environments further challenge the performance of algorithms. Factors such as deviations from a straight path, turns and obstacles, limited visibility, crowded areas, and other mobility tasks such as navigating stairs or slopes, may have influenced the DMO estimates. These conditions make stride length estimation particularly challenging, as this DMO is sensitive to intermittent gait [27]. Stride length exhibited larger errors and poor reliability in all WB durations, contributing to the errors in walking speed and distance. This lower performance of stride length could be related to the algorithms assuming an inverted pendulum, which might not properly fit the hip fracture cohort due to their asymmetrical gait [13]. Further work is necessary to improve stride length estimation in older adults after hip fracture.

Impact of Physical Function

Our results show that walking speed can be estimated with good reliability in older adults with higher SPPB scores in all WB duration categories, with errors ranging from -3.59% to -1.73%. Conversely, the lower SPPB score group exhibited larger errors, ranging from 26.58% to 32.56%, and demonstrated poor reliability for all WB duration categories, except for WBs lasting longer than 30 seconds, which showed moderate reliability and lower errors (18.96%). These group differences are likely due to the differences in their real-world walking speed. As reported in our previous work, the performance of all algorithms decreased for speeds below 0.5 m/s [27], a threshold distinguishing between slow and medium-speed walkers [47]. Those with lower SPPB scores walked slower on average than those with higher SPPB scores, 0.59 m/s versus 0.70 m/s, respectively, as measured by the reference system. The lower SPPB score group walked slower and had higher errors in 10-30 seconds WBs. Although 22% of all WBs across participants were slower than 0.5 m/s, only one participant had a median speed slower than this (see Figure 1 and Table 1, participant 3). Notably, this participant was among those with the lowest errors in walking speed and distance estimates, despite having a SPPB score of only 3. In general, errors in slow walkers may be due to lower amplitude in the acceleration signals, inconsistent gait cycles [48, 49], irregular gait patterns [50, 51], short step length, shuffling gait [8], and less symmetrical gait [13]. Nevertheless, DMOs in the lower SPPB score group demonstrated better accuracy and moderate reliability in WBs >30 seconds, indicating that accuracy improves during longer WBs irrespective of physical function.

Patients have individual preferences regarding whether walking faster or further is more important [46], and this often depends on the context. For instance, walking speed is essential for tasks such as crossing the road, while walking distance is more relevant for engaging in activities such as going to the grocery store or participating in the community. As the majority of patients report walking further as more important [46], a reliable distance estimate would provide clinicians with valuable information that is meaningful to their patients. Overall, distance estimates showed good to excellent reliability and low errors, especially for the longest WBs. When stratified by SPPB score, distance displayed the same error pattern as walking speed, with errors being larger in shorter WBs. Although longer WBs could be more informative as these likely represent more consistent outdoor walking, for many patients, life-space after a hip fracture may be confined primarily to their home environment, where furniture, walls and doors are likely to result in significantly shorter WBs.

Distance was overestimated in older adults with lower SPPB scores and underestimated in those with higher SPPB scores, showing large variations in errors despite excellent reliability. This could be explained by the combination of overestimated average walking speed and underestimated WB duration [27], the multiplication of which formed the distance DMO. Furthermore, high errors in the

lower SPPB score group could be explained by an inclusive WB definition that allowed for brief pauses of up to 3s [37]. Such brief pauses are included in the total duration of a WB but are not taken into account in the calculation of mean walking speed, leading to a potential overestimation of distance. Older adults with lower physical function after a hip fracture may be more likely to have brief pauses and slower turns within a WB than those with higher physical function. Therefore, when interpreting walked distance, researchers and clinicians should focus on longer and more continuous WBs if possible.

Impact of Time Since Surgery and Walking Aids

Promisingly, our results showed no discernible change in the accuracy of walking speed and distance estimates across 32-390 days after hip fracture surgery, suggesting that these DMOs can be used to capture daily life gait from as early as 1 month after surgery. As this study assessed DMO errors in the real world, we only included participants who were mobile in their home environment, which may have contributed to not having included patients within the first month after surgery. Therefore, we cannot conclude on the accuracy and reliability of DMOs in the acute phase after hip fracture. However, in the initial days and weeks following hip fracture surgery, when many patients are still in the hospital or rehabilitation centre [6, 7], clinical assessments may be more appropriate and informative than real-world gait monitoring. The primary focus during this period is typically first on pain management and early mobilisation, subsequently followed by rehabilitation aimed at improving physical function and enabling independent walking to prepare patients for discharge to home environments [8, 9]. Importantly, physical function usually improves the most during the first 3 months after hip fracture surgery [41], with DMO accuracy for participants assessed within this period showing no discernible deviation from those assessed after the first 3 months. Similarly, DMO accuracy for the 2 participants who used walking aids did not deviate visibly from those without. Although we found no evidence that days since surgery or the use of walking aids systematically affect DMO accuracy and reliability, these results in our small sample size should be confirmed in a larger sample.

General Discussion

In hip fracture patients with reduced walking ability, shorter walking duration or slower walking speed, additional mobility outcomes may provide relevant information about their movement characteristics. The core DMOs from the Mobilise-D TVS allow for a broader assessment of additional DMOs, such as walking duration and number of WBs exceeding different duration thresholds. Other relevant metrics in this population may include the number of turns, time spent in an upright position, and the frequency of posture transitions (e.g., sit-to-stand). In particular, the number of sit-to-stand transitions may be an important mobility outcome for hip fracture patients, especially in the acute phase, and ongoing research is exploring this further.

While the lower back remains a viable placement for gait monitoring in hip fracture patients' daily lives, it may not be optimal for long-term recordings. Extended monitoring periods require sufficient battery life and may cause discomfort, particularly when sitting or lying. Additionally, adhesive attachments may irritate the skin, while belts used for fixation can introduce movement artifacts that compromise data quality. Other wearable device locations may help overcome some of these challenges. Although wrist-worn devices could be an option, wrist-based data is currently less accurate than lower-back data [52] and has shown lower accuracy in patients recovering from hip fractures [53]. In frail patients with slow and cautious gait, a more suitable location for a single wearable device may be the foot or shank, which is better suited for detecting initial contact events than the lower back [54]. Furthermore, promising alternative approaches for gait monitoring in hip fracture patients include instrumenting walking aids [55] and even the prostheses themselves [56], a recent area of innovative development.

Recent comprehensive validation studies represent a significant step toward advancing real-world gait monitoring in multiple cohorts by providing robust reliability and accuracy data for DMOs [26, 27]. Currently, clinically relevant accuracy thresholds have not yet been established for these DMOs in the hip fracture patient cohort, and they may also differ for different DMOs, making it challenging to define their appropriateness for clinical use. Our comprehensive and transparent reporting of accuracy and reliability measures allows readers to interpret the results in the context of their own research or clinical practice. Furthermore, based on the DMOs in this study, ongoing research is investigating the minimal important difference using the Mobilise-D clinical validation study dataset [57]. Further research is needed to establish clinically relevant thresholds to strengthen clinical applicability.

Clinical Implications

As most walking activity happens in daily life, outside a lab or clinic, understanding gait recovery after hip fracture surgery should not only rely on snapshot tests in the clinic or assessments in the lab, but be complemented with accurate and reliable real-world monitoring. This study shows that real-world gait monitoring can provide much-needed information about a patient's functional recovery. Currently, the Mobilise-D consortium is undertaking a comprehensive clinical validation of all validated DMOs [57], including construct validity, predictive capacity, and the ability to detect change. Together with the encouraging results of this study, this opens up studying real-world trajectories of gait recovery after a hip fracture and investigating what characteristics of the patient, treatment, or rehabilitation may predict these trajectories.

This study shows that for the hip fracture cohort, the number of steps and cadence DMOs are the most accurate and reliable, as well as suitable for both short and longer walking periods. These DMOs can also account for participants with functional constraints preventing them from performing outdoor walking. Furthermore, we found that walking speed, stride duration, and walking distance can be used to provide additional information beyond the most frequently studied DMOs such as steps, cadence and walking time [58-60]. In terms of walking speed, we recommend excluding the shortest WBs <10 seconds, to ensure accurate and reliable estimates without disregarding a disproportionate amount of data. Furthermore, clinicians should consider the patient's physical function when interpreting walking speed due to the overestimation in people with lower SPPB scores and underestimation in people with higher SPPB scores. Distance and stride duration showed promising results across all WB durations. However, in subgroups with lower SPPB scores (< 8), WBs >30 seconds are required for obtaining more accurate distance estimates and this may be challenging for the frailest patients. We showed that the distance and walking speed DMOs can be used as early as 1 month after surgery, also in people using a single cane or crutch, but more data on walking aid use is needed from a larger sample.

Strengths and Limitations

To the best of our knowledge, this is the first study to rigorously investigate the accuracy and reliability of real-world DMOs while considering critical hip fracture-related factors. A key strength is the sufficient number of 164 WBs [25], which improves the statistical power. However, the number of available WBs decreased substantially when excluding short WBs. Although this improved the accuracy and reliability of several DMOs, it also excluded a large part of participants' data, potentially jeopardizing their representativeness of daily life activities. Moreover, some participants contributed more WBs than others, especially those with better physical function. This imbalance may have biased the analyses by overrepresenting the walking patterns of more mobile participants. Future studies could address this by accounting for within-subject variability and ensuring more balanced data collection across participants.

While our study sample included participants with a wide range of physical function and walking

speeds, with SPPB scores ranging from 3 to 12 and 22% of the WBs below 0.5 m/s, our hip fracture sample was relatively well-functioning compared to the general hip fracture population, which may have SPPB scores ranging from 1 to 5 in the first year after fracture [30]. Additionally, while we had enough WBs to compare sample by sample in the true positive analysis [25], our number of participants included in the analyses was small. As a result, formal analyses of the potential effect of days since surgery and walking aid usage on DMO accuracy and reliability were not feasible.

Several factors might explain the specifics of our study sample. First, we only included participants who were mobile in their home environment. Second, as we recruited during the pandemic, we were prohibited from recruiting participants during their hospital stay, which consequently excluded those in the acute phase. Third, the complexity and thoroughness of the protocol likely also limited recruitment of the frailest patients. Fourth, while we initially recruited 19 participants from a target sample size of 20, the sample size was reduced to 11 in the final analysis, primarily due to technical errors with the reference system. However, 2 participants were excluded due to insufficient signal quality, which may have been related to their low physical function (SPPB scores of 3), potentially accompanied by shuffling and/or a very cautious gait, making it challenging to detect gait events. As a result, our findings may not fully generalize to more severely impaired patients, and a larger sample is needed to provide stronger evidence. These limitations constrain our findings but offer a promising foundation for future research to improve the robustness and generalizability of DMOs in frail patient populations.

Conclusions

Considering specific factors critical for older adults after a hip fracture, our study identified 5 accurate and reliable real-world DMO estimates from a single wearable device worn on the lower back: walking speed, cadence, stride duration, number of steps, and distance. The accuracy and reliability of most DMOs improved when excluding WBs less than 10 seconds and were higher in WBs >30 seconds compared to 10-30 seconds WBs, and for those with higher physical function. DMOs can capture daily gait as early as 1 month after surgery, also in people using walking aids. However, as the majority of patients after hip fracture perform WBs of short duration, there is a trade-off between more accurate and reliable walking speed and distance estimates and the disregard of substantial amounts of data. Our results add more granularity to real-world gait assessments in populations with severe gait impairments. They have important implications for future research since they can provide a significant contribution to clinical validation studies, randomized controlled trials, and descriptive studies on gait recovery after hip fracture that is meaningful for clinicians and patients alike. Finally, these results support using these DMOs to assess intervention effects on real-world gait in detail, thereby aiding in the design of optimal care pathways.

Acknowledgements

This work was supported by the Mobilise-D project that has received funding from the Innovative Medicines Initiative (IMI) 2 Joint Undertaking under grant 820820. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

The authors thank the Mobilise-D TVS team of Work Package 2 and all participants who devoted their time and energy to this taxing protocol in times of COVID-19. A special thanks to Miriam Dinger at the Robert Bosch Foundation for Medical Research and Dr Tim Klüter at the University Hospital Schleswig-Holstein for their indispensable help in recruiting patients. MAB acknowledges ChatGPT, an AI model by OpenAI, for suggestions that helped improve the clarity of some sentences in this paper.

SDD and LR were also supported by the IDEA-FAST project that has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 853981. LR is a

National Institute for Health and Care Research (NIHR) Senior Investigator. Both SDD and LR were supported by the 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. SDD and LR were also supported by the NIHR/Wellcome Trust Clinical Research Facility (CRF) infrastructure at Newcastle upon Tyne Hospitals NHS Foundation Trust. SDD was supported by the UK Research and Innovation (UKRI) Engineering and Physical Sciences Research Council (EPSRC) (Grant Ref: EP/X031012/1 and Grant Ref: EP/X036146/1). Barcelona Institute of Global Health (ISGlobal) acknowledges support from the grant CEX2018-000806-S funded by Ministerio de Ciencia, Innovación y Universidades (MCIN)/ Agencia Estatal de Investigación (AEI) MCIN/AEI/10.13039/501100011033, and support from the Generalitat de Catalunya through the Centres de Recerca de Catalunya (CERCA) Program. The views expressed in this publication are those of the author(s) and do not necessarily reflect those of IMI, the European Union, EFPIA, any Associated Partners, NIHR, or the Department of Health and Social Care, who are not responsible for any use that may be made of the information contained herein.

Conflict of Interest

S. Del Din reports consultancy activity with Hoffmann-La Roche Ltd. outside of this study.

Abbreviations

DMO = Digital Mobility Outcome
DSS = Days Since Surgery
ICC = Intraclass Correlation Coefficient
IMU = Inertial Measurement Unit
LoA = Limits of Agreement
MoCA = Montreal Cognitive Assessment
MAE = Mean Absolute Error
MARE = Mean Absolute Relative Error
MRE = Mean Relative Error
RS = Reference System
SPPB = Short Physical Performance Battery
SWD = Single Wearable Device
TVS = Technical Validation Study
WB = Walking Bout
4MWT = 4 Minute Walk Test

Data Availability

The complete dataset of the Mobilise-D TVS can be found on Zenodo [61].

References

1. Emmerson BR, Varacallo M, Inman D. Hip Fracture Overview. Treasure Island (FL): StatPearls Publishing; 2023 [2024-10-05]; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK557514/>.
2. Downey C, Kelly M, Quinlan JF. Changing trends in the mortality rate at 1-year post hip fracture - a systematic review. *World J Orthop.* 2019;10(3):166-75. PMID: 30918799. doi: 10.5312/wjo.v10.i3.166.
3. Dyer SM, Crotty M, Fairhall N, Magaziner J, Beaupre LA, Cameron ID, et al. A critical review of the long-term disability outcomes following hip fracture. *BMC Geriatrics.* 2016;16(1):158. doi: 10.1186/s12877-016-0332-0.

4. Delgado-Ortiz L, Polhemus A, Keogh A, Sutton N, Remmele W, Hansen C, et al. Listening to the patients' voice: a conceptual framework of the walking experience. *Age Ageing*. 2023;52(1):afac233. doi: 10.1093/ageing/afac233.
5. Wahlsten LR, Smedegaard L, Brorson S, Gislason G, Palm H. Living settings and cognitive impairment are stronger predictors of nursing home admission after hip fracture surgery than physical comorbidities A nationwide Danish cohort study. *Injury*. 2020;51(10):2289-94. doi: 10.1016/j.injury.2020.06.041.
6. Grigoryan KV, Javedan H, Rudolph JL. Orthogeriatric Care Models and Outcomes in Hip Fracture Patients: A Systematic Review and Meta-Analysis. *J Orthop Trauma*. 2014;28(3). doi: 10.1097/BOT.0b013e3182a5a045.
7. Werner M, Macke C, Gogol M, Krettek C, Lioudakis E. Differences in hip fracture care in Europe: a systematic review of recent annual reports of hip fracture registries. *Eur J Trauma Emerg Surg*. 2022 Jun 1;48(3):1625-38. doi: 10.1007/s00068-021-01797-8.
8. Nijdam TMP, Laane DWPM, Schiepers TEE, Smeeing DPJ, Kempen DHR, Willems HC, et al. The goals of care in acute setting for geriatric patients in case of a hip fracture. *Eur J Trauma Emerg Surg*. 2023;49(4):1835-44. doi: 10.1007/s00068-023-02258-0.
9. Reyes BJ, Mendelson DA, Mujahid N, Mears SC, Gleason L, Mangione KK, et al. Postacute Management of Older Adults Suffering an Osteoporotic Hip Fracture: A Consensus Statement From the International Geriatric Fracture Society. *Geriatr Orthop Surg Rehabil*. 2020;11. PMID: 32728485. doi: 10.1177/2151459320935100.
10. Taraldsen K, Thingstad P, Sletvold O, Saltvedt I, Lydersen S, Granat MH, et al. The long-term effect of being treated in a geriatric ward compared to an orthopaedic ward on six measures of free-living physical behavior 4 and 12 months after a hip fracture - a randomised controlled trial. *BMC Geriatr*. 2015 Dec 4;15:160. PMID: 26637222. doi: 10.1186/s12877-015-0153-6.
11. Taylor NF, Peiris CL, Thompson AL, Prendergast LA, Harding KE, Hau R, et al. Association between physical activity and short-term physical function changes after hip fracture: An observational study. *Physiother Res Int*. 2021;26(1):e1876. PMID: 32918389. doi: 10.1002/pri.1876.
12. Dargent-Molina P, Favier F, Grandjean H, Baudoin C, Schott AM, Hausherr E, et al. Fall-related factors and risk of hip fracture: the EPIDOS prospective study. *Lancet*. 1996 Jul 20;348(9021):145-9. PMID: 8684153. doi: 10.1016/s0140-6736(96)01440-7.
13. Thingstad P, Egerton T, Ihlen EF, Taraldsen K, Moe-Nilssen R, Helbostad JL. Identification of gait domains and key gait variables following hip fracture. *BMC Geriatrics*. 2015;15(1):150. doi: 10.1186/s12877-015-0147-4.
14. Hulleck AA, Menoth Mohan D, Abdallah N, El Rich M, Khalaf K. Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. *Front Med Technol*. 2022;4:901331. PMID: 36590154. doi: 10.3389/fmedt.2022.901331.
15. Taraldsen K, Polhemus A, Engdal M, Jansen CP, Becker C, Brenner N, et al. Evaluation of mobility recovery after hip fracture: a scoping review of randomized controlled studies. *Osteoporos Int*. 2024;35(2):203-15. doi: 10.1007/s00198-023-06922-4.
16. Cappozzo A. Gait analysis methodology. *Human Movement Science*. 1984;3(1):27-50. doi: 10.1016/0167-9457(84)90004-6.
17. Warmerdam E, Hausdorff JM, Atrsaei A, Zhou Y, Mirelman A, Aminian K, et al. Long-term unsupervised mobility assessment in movement disorders. *The Lancet Neurology*. 2020;19(5):462-70. doi: 10.1016/S1474-4422(19)30397-7.
18. Van Ancum JM, van Schooten KS, Jonkman NH, Huijben B, van Lummel RC, Meskers CGM, et al. Gait speed assessed by a 4-m walk test is not representative of daily-life gait speed in community-dwelling adults. *Maturitas*. 2019;121:28-34. doi: 10.1016/j.maturitas.2018.12.008.

19. Hillel I, Gazit E, Nieuwboer A, Avanzino L, Rochester L, Cereatti A, et al. Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring. *Eur Rev Aging Phys Act.* 2019;16:6. PMID: 31073340. doi: 10.1186/s11556-019-0214-5.
20. Rochester L, Mazzà C, Mueller A, Caulfield B, McCarthy M, Becker C, et al. A Roadmap to Inform Development, Validation and Approval of Digital Mobility Outcomes: The Mobilise-D Approach. *Digital Biomarkers.* 2020;4(1):13-27. doi: 10.1159/000512513.
21. Rehman RZ, Klocke P, Hryniv S, Galna B, Rochester L, Del Din S, et al. Turning Detection During Gait: Algorithm Validation and Influence of Sensor Location and Turning Characteristics in the Classification of Parkinson's Disease. *Sensors.* 2020;20(18):5377. doi: 10.3390/s20185377.
22. Mazzà C, Alcock L, Aminian K, Becker C, Bertuletti S, Bonci T, et al. Technical validation of real-world monitoring of gait: a multicentric observational study. *BMJ Open.* 2021;11(12):e050785. doi: 10.1136/bmjopen-2021-050785.
23. Mobilise-D. Connecting digital mobility assessment to clinical outcomes for regulatory and clinical endorsement. 2019 [2024-08-23]; Available from: <https://mobilise-d.eu/>.
24. Salis F, Bertuletti S, Bonci T, Caruso M, Scott K, Alcock L, et al. A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. *Frontiers in Bioengineering and Biotechnology.* 2023;11. doi: 10.3389/fbioe.2023.1143248.
25. Scott K, Bonci T, Salis F, Alcock L, Buckley E, Gazit E, et al. Design and validation of a multi-task, multi-context protocol for real-world gait simulation. *J Neuroeng Rehabil.* 2022;19(1):141. doi: 10.1186/s12984-022-01116-1.
26. Kirk C, Küderle A, Micó-Amigo ME, Bonci T, Paraschiv-Ionescu A, Ullrich M, et al. Mobilise-D insights to estimate real-world walking speed in multiple conditions with a wearable device. *Sci Rep.* 2024;14(1):1754. doi: 10.1038/s41598-024-51766-5.
27. Micó-Amigo ME, Bonci T, Paraschiv-Ionescu A, Ullrich M, Kirk C, Soltani A, et al. Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. *J Neuroeng Rehabil.* 2023;20(1):78. doi: 10.1186/s12984-023-01198-5.
28. Najafi B, Helbostad JL, Moe-Nilssen R, Zijlstra W, Aminian K. Does walking strategy in older people change as a function of walking distance? *Gait & Posture.* 2009;29(2):261-6. doi: 10.1016/j.gaitpost.2008.09.002.
29. Kline Mangione K, Craik RL, Lopopolo R, Tomlinson JD, Brenneman SK. Predictors of gait speed in patients after hip fracture. *Physiother Can.* 2008;60(1):10-8. PMID: 20145738. doi: 10.3138/physio/60/1/10.
30. Prestmo A, Hagen G, Sletvold O, Helbostad JL, Thingstad P, Taraldsen K, et al. Comprehensive geriatric care for patients with hip fractures: a prospective, randomised, controlled trial. *The Lancet.* 2015;385(9978):1623-33. doi: 10.1016/S0140-6736(14)62409-0.
31. Thomas S, Halbert J, Mackintosh S, Cameron ID, Kurrle S, Whitehead C, et al. Walking aid use after discharge following hip fracture is rarely reviewed and often inappropriate: an observational study. *J Physiother.* 2010;56(4):267-72. doi: 10.1016/S1836-9553(10)70010-2.
32. Bertuletti S, Della Croce U, Cereatti A. A wearable solution for accurate step detection based on the direct measurement of the inter-foot distance. *J Biomech.* 2019;84:274-7. doi: 10.1016/j.jbiomech.2018.12.039.
33. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc.* 2005;53(4):695-9. PMID: 15817019. doi: 10.1111/j.1532-5415.2005.53221.x.
34. Salis F, Bertuletti S, Bonci T, Della Croce U, Mazzà C, Cereatti A. A method for gait events detection based on low spatial resolution pressure insoles data. *J Biomech.* 2021;127:110687. doi:

10.1016/j.jbiomech.2021.110687.

35. Caruso M, Sabatini AM, Knaflitz M, Gazzoni M, Croce UD, Cereatti A. Orientation Estimation Through Magneto-Inertial Sensor Fusion: A Heuristic Approach for Suboptimal Parameters Tuning. *IEEE Sensors Journal*. 2020;21(3):3408-19. doi: 10.5281/zenodo.4457500.

36. Mobilise-D. MobGap - The Mobilise-D algorithm toolbox. 2024 [2024-08-28]; Available from: <https://github.com/mobilise-d/mobgap>.

37. Kluge F, Del Din S, Cereatti A, Gaßner H, Hansen C, Helbostad JL, et al. Consensus based framework for digital mobility monitoring. *PLoS One*. 2021;16(8):e0256541. PMID: 34415959. doi: 10.1371/journal.pone.0256541.

38. Baroudi L, Barton K, Cain SM, Shorter KA. Classification of human walking context using a single-point accelerometer. *Sci Rep*. 2024;14(1):3039. doi: 10.1038/s41598-024-53143-8.

39. Schmitt AC, Baudendistel ST, Lipat AL, White TA, Raffegeau TE, Hass CJ. Walking indoors, outdoors, and on a treadmill: Gait differences in healthy young and older adults. *Gait & Posture*. 2021;90:468-74. doi: 10.1016/j.gaitpost.2021.09.197.

40. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol*. 1994;49(2):M85-94. PMID: 8126356. doi: 10.1093/geronj/49.2.m85.

41. Beckmann M, Bruun-Olsen V, Pripp AH, Bergland A, Smith T, Heiberg KE. Recovery and prediction of physical function 1 year following hip fracture. *Physiother Res Int*. 2022;27(3):1947. doi: 10.1002/pri.1947.

42. Keppler AM, Holzschuh J, Pfeufer D, Gleich J, Neuerburg C, Kammerlander C, et al. Mobility improvement in the first 6 postoperative weeks in orthogeriatric fracture patients. *Eur J Trauma Emerg Surg*. 2022;48(4):2867-72. PMID: 34932124. doi: 10.1007/s00068-021-01856-0.

43. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015;25(2):141-51. PMID: 26110027. doi: 10.11613/bm.2015.015.

44. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30-46. doi: 10.1037/1082-989X.1.1.30.

45. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 2016;15(2):155-63. PMID: 27330520. doi: 10.1016/j.jcm.2016.02.012.

46. Combs SA, Van Puymbroeck M, Altenburger PA, Miller KK, Dierks TA, Schmid AA. Is walking faster or walking farther more important to persons with chronic stroke? *Disabil Rehabil*. 2013;35(10):860-7. PMID: 23035811. doi: 10.3109/09638288.2012.717575.

47. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait speed and survival in older adults. *JAMA*. 2011;305(1):50-8. PMID: 21205966. doi: 10.1001/jama.2010.1923.

48. Hebenstreit F, Leibold A, Krinner S, Welsch G, Lochmann M, Eskofier BM. Effect of walking speed on gait sub phase durations. *Hum Mov Sci*. 2015;43:118-24. PMID: 26256534. doi: 10.1016/j.humov.2015.07.009.

49. Quintero D, Lambert DJ, Villarreal DJ, Gregg RD. Real-Time Continuous Gait Phase and Speed Estimation from a Single Sensor. *Control Technol Appl*. 2017;2017:847-52. PMID: 30148285. doi: 10.1109/ccta.2017.8062565.

50. Soltani A, Aminian K, Mazza C, Cereatti A, Palmerini L, Bonci T, et al. Algorithms for Walking Speed Estimation Using a Lower-Back-Worn Inertial Sensor: A Cross-Validation on Speed Ranges. *IEEE Trans Neural Syst Rehabil Eng*. 2021;29:1955-64. doi: 10.1109/TNSRE.2021.3111681.

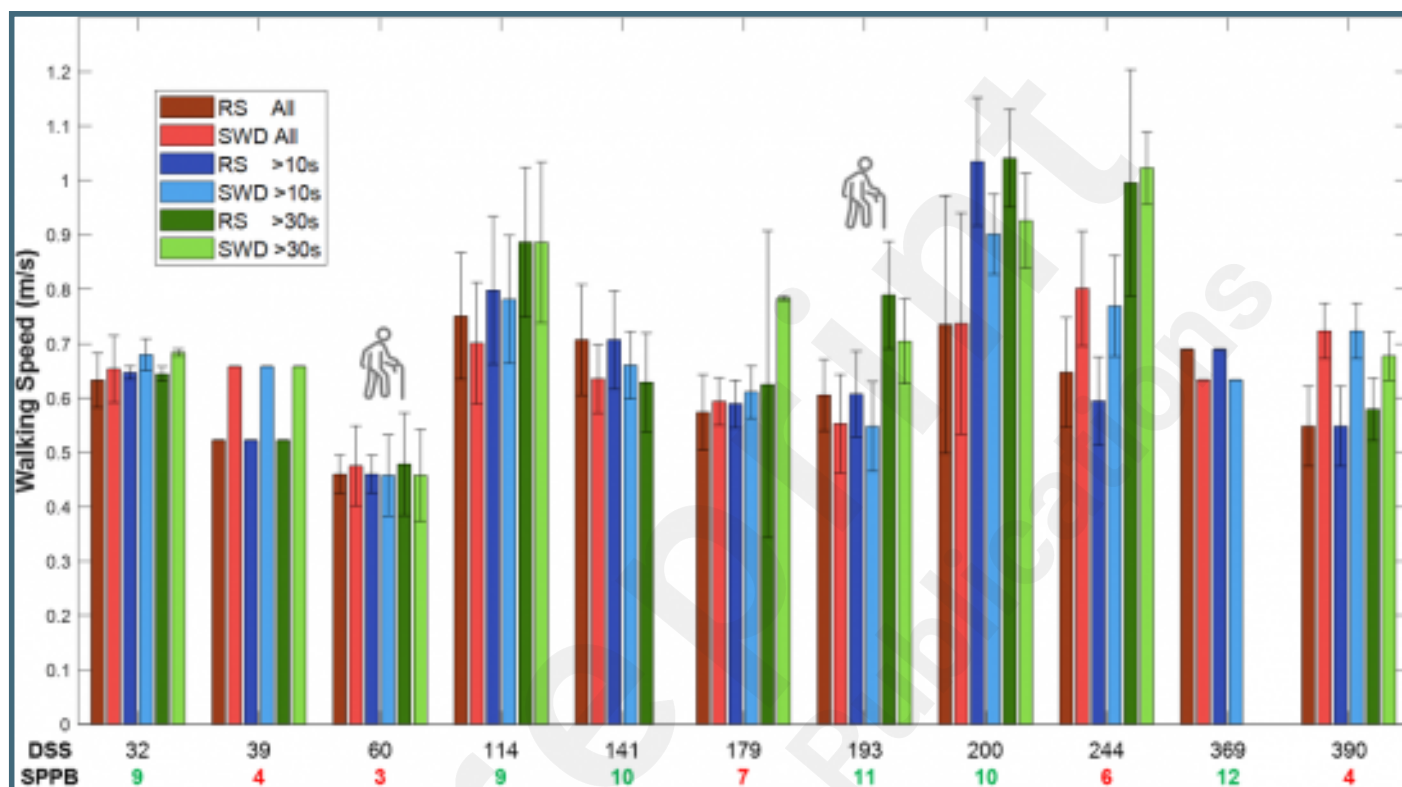
51. Taraldsen K, Thingstad P, Døhl Ø, Follestad T, Helbostad JL, Lamb SE, et al. Short and long-term clinical effectiveness and cost-effectiveness of a late-phase community-based balance and gait

- exercise program following hip fracture. The EVA-Hip Randomised Controlled Trial. *PLoS One*. 2019;14(11):e0224971. PMID: 31738792. doi: 10.1371/journal.pone.0224971.
52. Kim DW, Hassett LM, Nguy V, Allen NE. A Comparison of Activity Monitor Data from Devices Worn on the Wrist and the Waist in People with Parkinson's Disease. *Mov Disord Clin Pract*. 2019 Nov;6(8):693-9. PMID: 31745480. doi: 10.1002/mdc3.12850.
53. Kluge F, Brand YE, Micó-Amigo ME, Bertuletti S, D'Ascanio I, Gazit E, et al. Real-World Gait Detection Using a Wrist-Worn Inertial Sensor: Validation Study. *JMIR Form Res*. 2024 2024/5/1;8:e50035. doi: 10.2196/50035.
54. Pacini Panebianco G, Bisi MC, Stagni R, Fantozzi S. Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements. *Gait & Posture*. 2018 2018/10/01/;66:76-82. doi: 10.1016/j.gaitpost.2018.08.025.
55. Fernandez-Carmona M, Ballesteros J, Díaz-Boladeras M, Parra-Llanas X, Urdiales C, Gómez-de-Gabriel JM. Walk-IT: An Open-Source Modular Low-Cost Smart Rollator. *Sensors*. 2022;22(6). doi: 10.3390/s22062086.
56. Merle G, Parent-Harvey A, Harvey EJ. Sensors and digital medicine in orthopaedic surgery. *OTA Int*. 2022 Apr;5(2 Suppl):e189. PMID: 35949264. doi: 10.1097/oi9.0000000000000189.
57. Mikolaizak AS, Rochester L, Maetzler W, Sharrack B, Demeyer H, Mazzà C, et al. Connecting real-world digital mobility assessment to clinical outcomes for regulatory and clinical endorsement-the Mobilise-D study protocol. *PLoS One*. 2022;17(10):e0269615. PMID: 36201476. doi: 10.1371/journal.pone.0269615.
58. O'Halloran PD, Shields N, Blackstock F, Wintle E, Taylor NF. Motivational interviewing increases physical activity and self-efficacy in people living in the community after hip fracture: a randomized controlled trial. *Clin Rehabil*. 2016;30(11):1108-19. PMID: 26603892. doi: 10.1177/0269215515617814.
59. Senserrick C, Lawler K, Scroggie GD, Williams K, Taylor NF. Three short sessions of physiotherapy during rehabilitation after hip fracture were no more effective in improving mobility than a single longer session: a randomised controlled trial. *Physiotherapy*. 2021;112:87-95. PMID: 34052569. doi: 10.1016/j.physio.2020.12.002.
60. Zusman EZ, Dawes M, Fleig L, McAllister MM, Cook WL, Guy P, et al. Older Adults' Sedentary Behavior and Physical Activity After Hip Fracture: Results From an Outpatient Rehabilitation Randomized Controlled Trial. *J Geriatr Phys Ther*. 2019;42(2):E32-e8. PMID: 30028352. doi: 10.1519/jpt.0000000000000193.
61. Küderle A. Mobilise-D Technical Validation Study (TVS) dataset (1.0.0). 2024. doi: 10.5281/zenodo.13899386.

Supplementary Files

Figures

Walking speed estimates from the reference system (RS) and single wearable device (SWD) for each participant for All WBs combined (red bars), WBs above 10 seconds (blue bars), and WBs above 30 seconds (green bars). See Table 4 below for the number of WBs per category for each participant. The bars indicate the median and the error bars the interquartile range. Participants are sorted by days since surgery (DSS). The total Short Physical Performance Battery (SPPB) scores for each participant are listed below the x-axis, scores in green are associated with medium/high functioning (8-12), and scores in red with impaired activities of daily living functions (0-7). The 2 walking aid users are marked with an icon.



Mean relative error (a) and error (b) of the single wearable device (SWD) distance estimates and distance and median walking speed as measured by the reference system (RS) (c) for each participant for All WBs combined (red markers), WBs above 10 seconds (blue markers), and WBs above 30 seconds (green markers). Boxplots are shown where there are ≥ 5 data points, and indicate the median, interquartile range, and whiskers showing the minimum and maximum values within 1.5 times the interquartile range. Participants are sorted by days since surgery (DSS). The total Short Physical Performance Battery (SPPB) scores for each participant are listed. Scores in green indicate medium/high functioning (8-12), while scores in red indicate the onset of impaired activities of daily living functions (0-7). Walking aid users are marked with an icon.



TOC/Feature image for homepages

Person walking upstairs in the real-world with a McRoberts wearable device attached to the lower back.

