THE PREDICTION OF BIOPSYCHOSOCIAL FUNCTIONING USING HOME-BASED ACCELEROMETRY

NADINE DE JONG

THE PREDICTION OF BIOPSYCHOSOCIAL FUNCTIONING USING HOME-BASED ACCELEROMETRY

- MSc thesis -

Nadine de Jong

Student number: 4542045

10 - 01 - 2024

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University; Delft University of Technology; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)

Dept. of Rehabilitation Medicine, LUMC

April 2023 – January 2024

Supervisor(s):

Drs. P.M. Dekker

Dr. ir. J.H. de Groot

Thesis committee members:

Dr. ir. Winfred Mugge, TU Delft (chair)

Drs. P.M. Dekker, LUMC

Dr. ir. J.H. de Groot, LUMC

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Cover: Image obtained from the introduction video of McRoberts - MoveMonitor, made by Motoko. Available at https://vimeo.com/248286364.

Summary

Preventive, personalized treatment of biopsychosocial (BPS) functional decline or stagnation in the rehabilitation process requires early prediction of those at risk. Ambulant monitoring of objective physical activity using wearable accelerometers could be a quantitative, noninvasive, and affordable method to gain insight into current and future biopsychosocial functioning.

In this retrospective study, Random Forest Regression was used to predict cross-sectional and longitudinal Functional Ambulation Category (FAC) and Six-item Cognitive Impairment Test (6CIT) after hip fracture using ambulatory accelerometry data. Accelerometry data and BPS functional assessments were available of 49 participants of the HIPCARE study, assessing prognostic determinants of outcome after hip fracture in the elderly.

Overall, cross-sectional FAC scores three months after hip fracture could be predicted with moderately low error, and categorized regression predictions showed high precision and recall. Cross-sectional 6CIT and both longitudinal regression models underperformed, but categorized regression predictions revealed mixed but more promising precision and recall.

It is expected that the predictive performance of models can be improved by increasing participant sample size with balanced samples over population-specific, prevalent ranges of BPS outcome scales and exploring additional machine learning models. In the future, accurate accelerometry-based predictions for individual patients needing rehabilitation could support personalized treatment and improve long-term biopsychosocial functioning.

Table of Contents

List of Abbreviations

6CIT	Six-item Cognitive Impairment Test
ADL	Activities of Daily Living
BMI	Body Mass Index
BMR	Basal Metabolic Rate
BPS	Biopsychosocial
CV	Cross-Validation
DA	Discriminant Analysis
FAC	Functional Ambulation Category
ICF	International Classification of Functioning, Disability, and Health
IDE	Integrated Development Environment
IQR	Interquartile Range
MET	Metabolic Equivalent of Task
ML	Machine Learning
MVPA	Moderate-to-Vigorous Physical Activity
NAP	Napping
PA	Physical Activity
PAI	Physical Activity Intensity
PCA	Principal Component Analysis
PLS	Partial Least Squares
QoL	Quality of Life
RAR	Rest-Activity Rhythm
RF	Random Forest
RMSE	Root Mean Squared Error
SB	Sedentary Behavior
SD	Standard Deviation
TPA	Total Physical Activity
TUG	Timed Up & Go
WHO	World Health Organization

1
Introduction

As chronic diseases continue to account for a significant portion of morbidity and mortality in Western countries, healthcare systems that traditionally revolve around acute biomedical care models are facing challenges in improving patient-reported outcomes and reducing healthcare costs [1]. The increase in healthcare costs is mainly due to increased treatment and care costs and more extended diagnosis and treatment processes of increasingly complex health problems [2]. To address these healthcare needs, there is an increasing demand for the application of the biopsychosocial (BPS) model in healthcare management [1, 3]. This model offers a comprehensive approach to understanding and addressing the complex interplay of factors affecting health and well-being, making it particularly relevant in the context of chronic diseases and rehabilitation medicine [1-3]. This BPS effect of illness can also be seen in the case of elderly hip fracture patients. Not only does it (temporarily) affect mobility, but it is also related to decreased psychosocial well-being, such as social participation, depressive symptoms, fear of falling, and health-related quality of life [4-6].

While rehabilitation is essential for optimizing BPS functioning and improving the overall quality of life, creating and implementing successful rehabilitation programs is a complex undertaking that requires careful consideration of resource allocation, treatment efficacy, and the unique needs of each individual in the clinical and ambulant setting [7]. Rehabilitation does not only focus on biological recovery, it also strives to optimize BPS functioning using compensation techniques or substitution methods [8, 9]. In the field of rehabilitation medicine, wearable devices have emerged as valuable tools for implementing the principles of personalized medicine.

Human movement is a multifactorial and complex phenomenon, affected by all BPS aspects [10]. Subjective assessment of movement behavior, for example by using questionnaires, or lab-based objective assessment in a controlled environment, for example by the use of force plates, often do not reflect true day-to-day physical activity [11, 12]. Objective quantification of movement behavior in early stages of rehabilitation can be crucial for early decision-making concerning the need of (preventive) interventions to optimize BPS functioning in later phases of a health condition. Over the past years, accelerometry emerges as a noninvasive and affordable method to achieve this, being a widely accepted method for the objective quantification of physical activity and behavior [13, 14]. This technology utilizes accelerometers, often worn on the wrist or waist, to record body movement. This allows researchers to collect data on the intensity, duration, and frequency of physical movements, offering valuable insights into various aspects of human movement behavior. While the link between accelerometry and physical fitness is intuitive, its connection to psychological wellbeing may be less evident [15-17]. Growing numbers of studies, as well as multiple reviews, have assessed the association between physical activity and psychological wellbeing using accelerometry. Unfortunately, these studies are primarily cross-sectional and assess a heterogenous range of psychosocial and accelerometric measures [18-21]. Review of available literature demonstrated the potential of some accelerometric outcome measures assessing both physical activity intensity (PAI)

and rest-activity rhythms (RAR) to be an early indicator of future psychological wellbeing in adults (Appendix C).

So far, most papers either assess accelerometric features describing PAI or RAR (Appendix C). Combining these could result in more accurate findings on the interaction between physical activity and psychological wellbeing [22]. These results raise the question if accelerometry can be used to develop a model for the prediction of the expected biopsychosocial rehabilitative course of individual patients. High quality predictive models could help with the early identification of patients at a higher risk of complications or relapse during rehabilitation and to customize care plans based on individual traits. This individualized approach ensures that patients receive the most effective therapy techniques, maximizing BPS functioning. [7, 23, 24] Accelerometry can also be a fundamental component of remote rehabilitation healthcare and long-term follow-up [7, 25, 26]. Additionally, early information on the predicted course of recovery can also contribute to the efficiency of rehabilitation care concerning required equipment, staff and time when aiming to improve resource allocation and quality of treatment [7] Development of predictive models can be done using machine learning (ML), which is capable of processing large amounts of data and derive clinically relevant conclusions including diagnosis and morbidity risk assessment [7, 27, 28]. This makes it very suitable for processing raw accelerometric signals or large sets of accelerometric features for the prediction of biopsychosocial wellbeing [29-32]. There is a large range of machine learning algorithms, all using a different approach and use [33]. Random Forest (RF) models use a relatively simple technique based on the combination of a large number of decision trees and have proven to be useful for a wide range of medical applications, diagnostics and treatment considerations [34-37].

In this study it is explored how home-based accelerometry can be used for the prediction of biopsychosocial wellbeing three months and one year after hip fracture in the elderly. The aim of this study is to assess whether an accelerometry data-driven Random Forest machine learning model is able to correctly predict levels of BPS functioning three and twelve months after hip fracture. This would enable early detection of patients vulnerable for decline or stagnation of BPS functioning, paving the way for further personalized adaptation of treatment plans.

2

Background

For a deeper understanding of biopsychosocial functioning and preventive, predictive medicine, additional information on both topics is provided in this chapter.

2.1 Biopsychosocial Model

The biopsychosocial model emerged as a response to the limitations of the traditional biomedical model in explaining health and illness [1, 3]. Relying solely on diagnosis does not provide adequate prediction of necessary care, hospitalization duration and BPS functioning, nor the likelihood of (social) (re)integration [38]. Alternatively, the social model of disability only views disability as only an issue that arises from societal constructs and, although partially valid, it is like the biomedical model not adequate enough [38]. The central tenet of the biopsychosocial model is the recognition that health and illness result from intricate interactions among biological, psychological and social factors [1, 3, 38]. This model has gained widespread recognition and application in various aspects of healthcare such as in research on complex healthcare interventions, clinical practice, and the development of clinical guidelines, and serves as the foundation for the World Health Organization's International Classification of Functioning, Disability, and Health (ICF, Figure 1) [1, 38]. This model attempts to encourage a holistic approach of the patient, although it is not yet fully complete, for example by not taking into account the effect of time in the process of improvement or deterioration of disability [2].

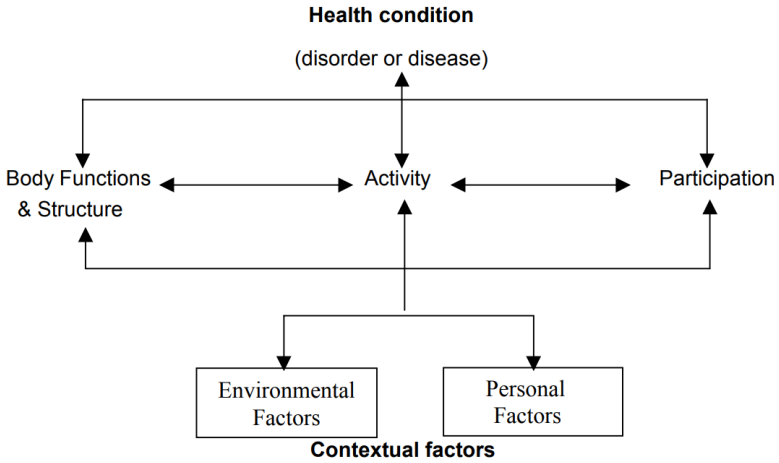


Figure 1. Overview of the International Classification of Functioning, Disability, and Health (ICF) model. Reprinted from *Towards a Common Language for Functioning, Disability and Health: ICF*, WHO, 2002 [38].

The interaction between the different aspects within the biopsychosocial model holds significant relevance in the field of rehabilitation medicine, where the focus is on optimizing patients' functional recovery and overall quality of life, which are highly related to each other as well as to both biomedical illness and contextual factors of the patient [2]. Many studies highlight these

multidirectional interactions between psychosocial factors and functional outcomes assessing a wide range of outcomes during and after rehabilitation, including function, well-being, perceived pain and disability, work status, social participation and overall quality of life (QoL) [4, 39-43].

For example, Adomaviciene and colleagues concluded that during the rehabilitation process after rotator cuff repair pain and psychosocial factors were of influence on functional shoulder recovery and that vice versa motor function, ability and pain relief had a long-term effect on subjective well-being [44]. Bordne et al. showed that geriatric rehabilitation did improve affect as part of subjective well-being, but that psychological parameters like personality should be taken into account as part of geriatric rehabilitation when determining if patients are at risk for poor overall subjective well-being [39].

Eleuteri et al. highlight the importance of the use of a biopsychosocial approach in the treatment management of hip fracture patients and their caregivers [40]. There is an association between hip fractures and a reduction of QoL which is influenced by pre-existing care dependence, reduced function and cognitive state and depressive symptoms [40, 45]. A cohort of hip fracture patients reported a decrease of mobility and increased rates of depressive symptoms and activity participation in the first year after fracture, with a larger social network reducing the detrimental effect of fracture [4]. These psychosocial factors, as well as delirium, are associated with poorer functional recovery and increased mortality rate [40]. Finally, fear of falling can also have a negative effect on recovery [5, 40]. Falls have a negative impact on physical health, but also on psychological well-being and social participation, which are believed to be mediated by respectively perceived control and frailty [5, 41, 43]. Social isolation on its turn is associated with future functional status such as gait speed and (instrumental) activities of daily living ((I)ADL) disability, demonstrating the tight relations between biopsychosocial aspects [46, 47].

3.2 P4 Medicine

Predictive, Preventive, Personalized, Participatory (P4) medicine is a transformative approach in healthcare that is revolutionizing disease management by emphasizing a proactive rather than reactive strategy. This approach focuses on individual patients and their unique characteristics, moving away from the traditional standardized model of healthcare. P4 medicine is characterized by the use of predictive methods, early disease identification, a personalized approach that considers each patient as a unique entity, and finally a responsibility of the individual for the optimization of their health. Furthermore, it is very suitable for the inclusion of all aspects of the biopsychosocial model, assessing physical, psychological and cognitive health. [7, 27, 48-50]

The development P4 medicine depends on large databases of health related factors including, but not limited to, socio-demographic, biological and genetic details [27]. These databases can aid in the characterization of patient groups at risk of certain disease, and if taken even further be used to guide decision making and initiate screening or preventive treatment of patients at risk before actual disease onset [7, 27, 49]. It is important that these databases contain valuable, quantified data on different domains. Wearable devices have the potential to provide continuous health monitoring across various domains, including environmental, behavioral, physiological, and psychological factors, which all play a role in an individual's health and well-being. A sophisticated wearable could for example track exposure to sunlight (environmental), number of steps (behavioral) and heart rate (physiological). The integration of data from wearable devices allows for a more comprehensive and personalized assessment of a patient's health status. [27, 49] It aligns with the predictive aspect of P4 medicine by enabling the early detection of symptoms or changes in health status, promoting preventive measures, and ultimately optimizing the rehabilitation process.

3
Method

3.1 Study Design

This study is a part of the HIPCARE single-center mixed-method inception cohort study assessing prognostic determinants of outcome after hip fracture in the elderly [51, 52]. Data was collected between January 2019 and November 2021 and assessed retrospectively. The study cohort consists of community-dwelling elderly aged 70 years or older with unilateral proximal femoral fractur, who were eligible for (geriatric) rehabilitation. All participants have given written informed consent. Patients received standard care as defined in the National guideline treatment protocol of the proximal femoral fracture in the older population of the Dutch trauma surgery society (Nederlandse Vereniging voor Heelkunde) and the HMC Bronovo care pathway for patients with hip fracture ('zorgpad heupfracturen', METC-nr: 18-029). Participants had outpatient check-up at three months and twelve months after surgery, at which a range of biopsychosocial questionnaires were administered (Figure 2). Additionally, home-based physical activity over seven days was assessed using accelerometry at three month check-up (Figure 2).

3.2 Data Collection

3.2.1 Accelerometry Data

Participants were requested to wear the MoveMonitor triaxial accelerometer (McRoberts B.V. The Hague, The Netherlands) at the three month check-up timepoint for seven consecutive days, removing the accelerometer only during water activities (bathing, swimming) and main sleep period at night. The MoveMonitor is worn on the lower back and is attached to an elastic band which is worn around the waist at belt height. This triaxial accelerometer has a sample frequency of 100Hz, range of 8g and resolution of 1mg. Data is stored on a 1Gb flash memory which can contain up to 14 days of collected data. The McRoberts MoveMonitor and corresponding movement behavior detection algorithm has previously been validated for multiple populations, including the elderly [53, 54]. An extensive report on types of activities, transitions between activities, steps, movement durations, movement frequencies, movement intensities and energy expenditure is available in a minute-by-minute format [55].

3.2.2 Biopsychosocial Outcomes

Assessments concerning biopsychosocial functioning administered at three and twelve months after surgery were included for analysis, with the three month timepoint as baseline measurement (Figure 2). Outcomes were the clinically validated Six-item Cognitive Impairment Test (6CIT) and Functional Ambulation Category (FAC) tools (Table 1). Additional questionnaires and tests including Short Physical Performance Battery SPPB, Harris Hip Score, Timed Up & Go test, Katz ADL scale, EuroQol 5D and Parker Mobility Score were administered but not included for initial analysis.

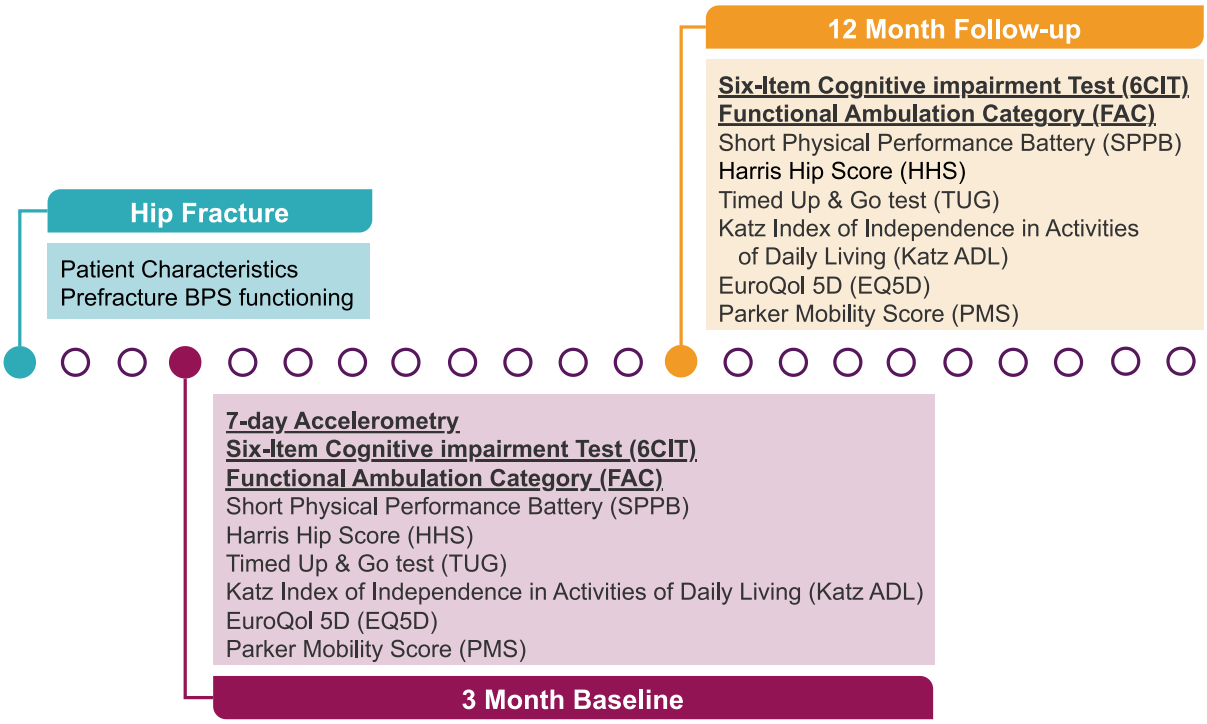


Figure 2. Participant timeline with biopsychosocial functioning assessed at three and twelve month check-up, and accelerometry assessed at 3-month check-up. *BPS, Biopsychosocial.*

3.3 Data Preprocessing

All data processing and predictive model training and evaluation were performed using the Python programming language (Python Software Foundation, https://www.python.org/) within the Spyder IDE (https://www.spyder-ide.org/) and multiple open source packages including, but not limited to, NumPy, pandas and Scikit-Learn [56-61].

3.3.1 Biopsychosocial Outcomes

Biopsychosocial outcomes were analyzed as absolute value for regression analysis. For post-hoc classification purposes, true and predicted outcomes were grouped in three categories, roughly distinguishing between the (1) unimpaired, (2) impaired and (3) either a mildly impaired or risk group (Table 1).

Table 1. Biopsychosocial assessments and interpretation of scores into functional categories.

Tool	Assesses	Functional Categories
6CIT [62]	Cognition	0-7 normal 8-9 risk 10-28 impaired
FAC [63]	Mobility	0-3 dependent or not functional 4 independent but limited 5 independent

6CIT, Six-Item Cognitive Impairment Test; FAC, Functional Ambulation Category.

3.3.2 Accelerometry Data

A subset of McRoberts' movement parameters were selected for inclusion in the data analysis based on previously performed literature review (Appendix C). Movement parameters were included if they described total physical activity, moderate-to-vigorous physical activity, sedentary behavior or napping (Table 2). Even though not described as most relevant movement parameter, sedentary behavior was included as it was expected that the included participants would engage in little moderate to vigorous physical activity. Multiple preprocessing steps were followed (Figure 3A), starting by redefining days to start and end at 03:00 to reduce spillage of late-night activity to the next day. Then, first and last day were removed since they do not accurately describe full day activity. Even though patients were asked to remove the accelerometer during the main sleep period at night, not all patients complied to this request. As a result, for some patients sleep data was removed by detecting periods of consecutive minutes lying down, with a duration of at least 300 minutes and onset between 16:00 and 06:00. Next, days with less than 12 hours of accelerometry data were removed and patients with less than 2 days of valid data were excluded from analysis.

After data cleaning, minute-to-minute movement parameters were grouped in to one-hour blocks and transformed into multiple features, including metrics and statistical measures such as the sum, mean, median and standard deviation, describing average movement behavior per hour (Figure 3B). Next, a moving average with a 10-minute window was applied to the movement parameters to decrease the power of short-lasting events (Figure 4). The maximum value per hour was included as feature as an indication of both power and duration of events. Combining 27 movement parameters, six descriptive transformations and 24 hour blocks resulted in 3888 features that were averaged over all valid days for each patient. Finally, included features were scaled based on train data using the Scikit-Learn StandardScaler which centers the data by removing the mean and scales the data to unit variance [64].

Figure 3A. Preprocessing steps of accelerometry data.

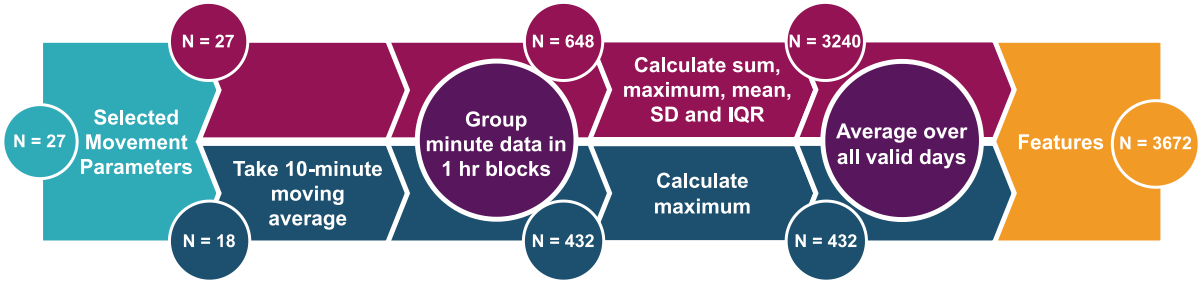


Figure 3B. Transformation steps of movement parameters into features. *SD, standard deviation; IQR, interquartile range.*

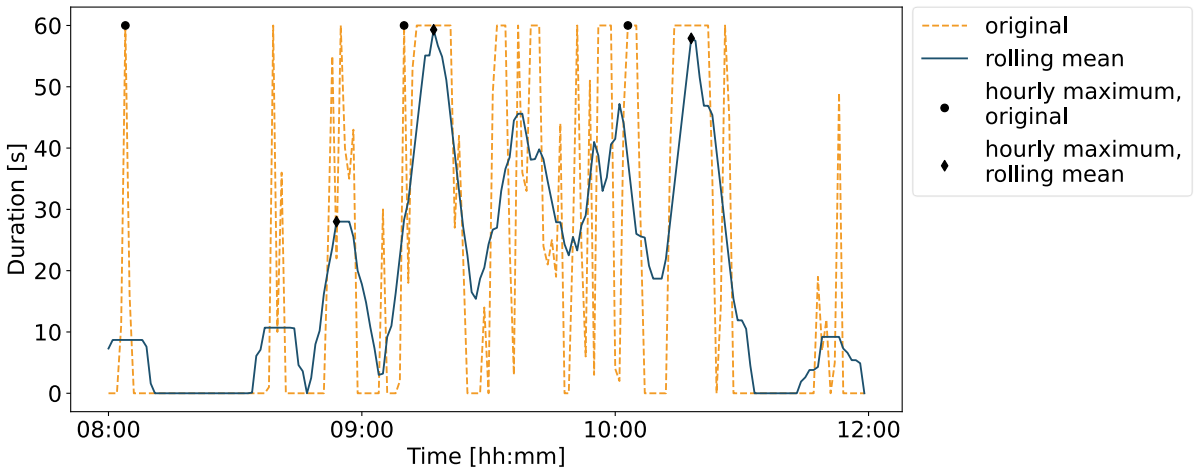


Figure 4. Example of difference between original feature and ten minute rolling mean of feature and corresponding hourly maxima.

3.4 Predictive Model Development

To assess whether home-based accelerometry data can be used to predict both cross-sectional and longitudinal BPS functioning, multiple supervised regression and classification methods were tested. These included the more basic linear and decision tree models as well as a regression tree model. Linear regression attempts to describe the relationship between variables and outcome using a linear function, whereas logistic regression predicts the outcome based on a logistic function. Decision tree models use a hierarchical tree structure, dividing samples into multiple groups based on feature based constraints (Figure 5A). Random forest trains multiple decision trees, with the most common label outputted by each tree being assigned to the sample (Figure 5B). [33]

Since we expected best results for the prediction of cross-sectional FAC score as they both describe mobility at the same time point, initial feature selection and model development were based on this outcome, although method was similar for other BPS outcomes.

A common problem in machine learning is overfitting, and cross-validation (CV) is a widely accepted method to reduce the effects of this pitfall [65]. Therefore, 4-fold CV was used, during which included samples were split into four batches, with each batch once being used as the test set and for the remaining three folds a part of the train set. This way the model is trained on three quarters of the data and its performance tested on a quarter of the data for each fold (Figure 5C). To increase accuracy each CV cycle was repeated five times, with performance estimation being based on 20 train-and-test cycles. For estimation of final model performance after feature selection, repeats were increased to 10.

Table 2. Included movement parameters as described by McRoberts.

Movement parameter	UNIT	MEASURE	DESCRIPTION
COUNTS_total	amount	TPA	the sum of the counts for worn periods
DUR_total_active	seconds	TPA	total duration of standing, shuffling, walking, stair walking and cycling periods combined
DUR_total_inactive	seconds	SB	total duration of lying and sitting periods combined
DUR_total_lying	seconds	NAP	total duration of lying periods
DUR_total_moving	seconds	TPA	total duration of walking, stair walking and cycling periods combined
DUR_total_sitting	seconds	SB	total duration of sitting periods
DUR_total_static	seconds	SB	total duration of standing and shuffling periods combined
MET_mean	kcal/minute	TPA	average MET value for worn periods
METS_moderate_time	seconds	MVPA	total duration of periods above or equal to 3 METs and below 6 METs when not taking bouts* into account
METS_sedentary_time	seconds	SB	total duration of periods below 3 METs when not taking bouts into account
METS_vigorous_time	seconds	MVPA	total duration of periods above or equal to 6 METs when not taking bouts into account
MI_active	mg	TPA	mean movement intensity of standing, shuffling, walking, stair walking and cycling periods combined
MI_inactive	mg	SB	mean movement intensity of lying and sitting periods combined
MI_lying	mg	NAP	mean movement intensity of lying periods
MI_moving	mg	TPA	mean movement intensity of walking, stair walking and cycling periods combined
MI_worn	mg	TPA	mean movement intensity of worn periods
PAR_active	ratio	TPA	ratio between total energy expenditure and BMR of standing, shuffling, walking, stair walking and cycling periods combined
PAR_inactive	ratio	SB	ratio between total energy expenditure and BMR of lying and sitting periods combined
PAR_lying	ratio	NAP	ratio between total energy expenditure and BMR of lying periods
PAR_moving	ratio	TPA	ratio between total energy expenditure and BMR of walking, stair walking and cycling periods combined
PERIODS_active	amount	TPA	number of periods of consecutive standing, shuffling, walking, stair walking and cycling
PERIODS_inactive	amount	SB	number of periods of consecutive lying and sitting
PERIODS_lying	amount	NAP	number of lying periods
PERIODS_moving	amount	TPA	number of periods of consecutive walking, stair walking and cycling
STEPS	amount	TPA	number of steps (including steps from walking stairs)
TRANSITIONS_ly	amount	-	number of transitions from lying to shuffling/standing/walking
TRANSITIONS_si	amount	-	number of transitions from sitting to shuffling/standing/walking

*TPA, total physical activity; SB, sedentary behavior; NAP, napping or daytime sleeping; MVPA, moderate-to-vigorous physical activity; MET, metabolic equivalent of task; BMR, basal metabolic rate. *A bout is a period during which the same movement behavior is performed for at least ten minutes with an allowed interruption of one minute.*

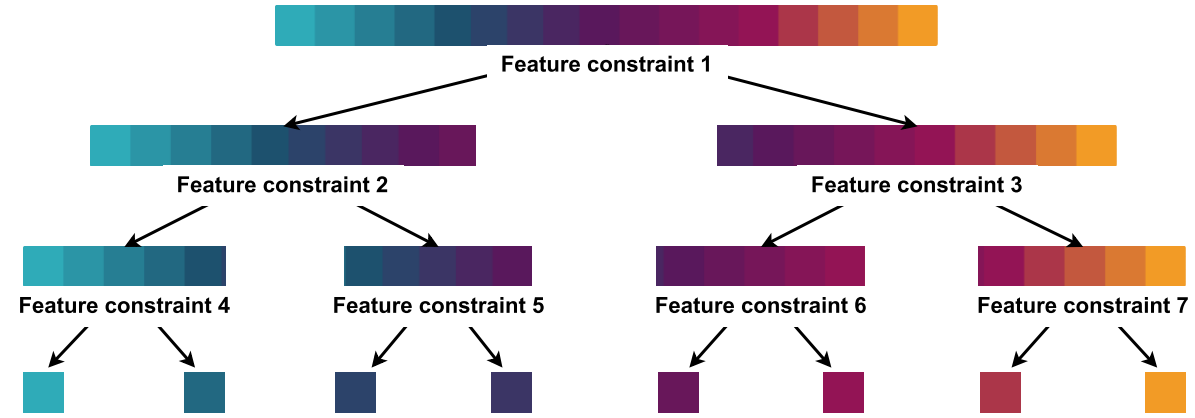


Figure 5A. Visualization of Decision Tree structure with a depth of 3. Splits are based on feature constraints.

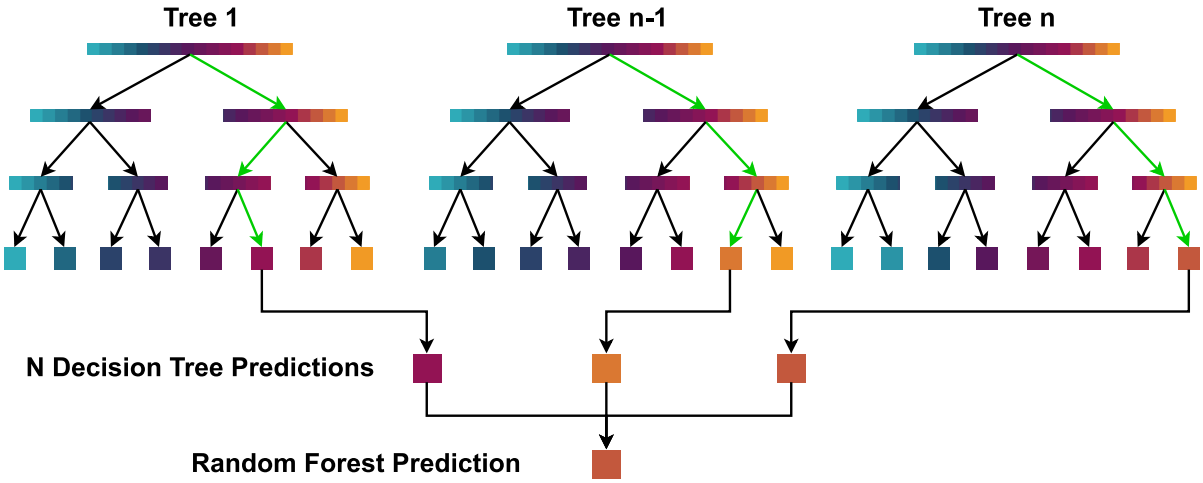


Figure 5B. Visualization of Random Forest with a depth of 3 and n-estimators.

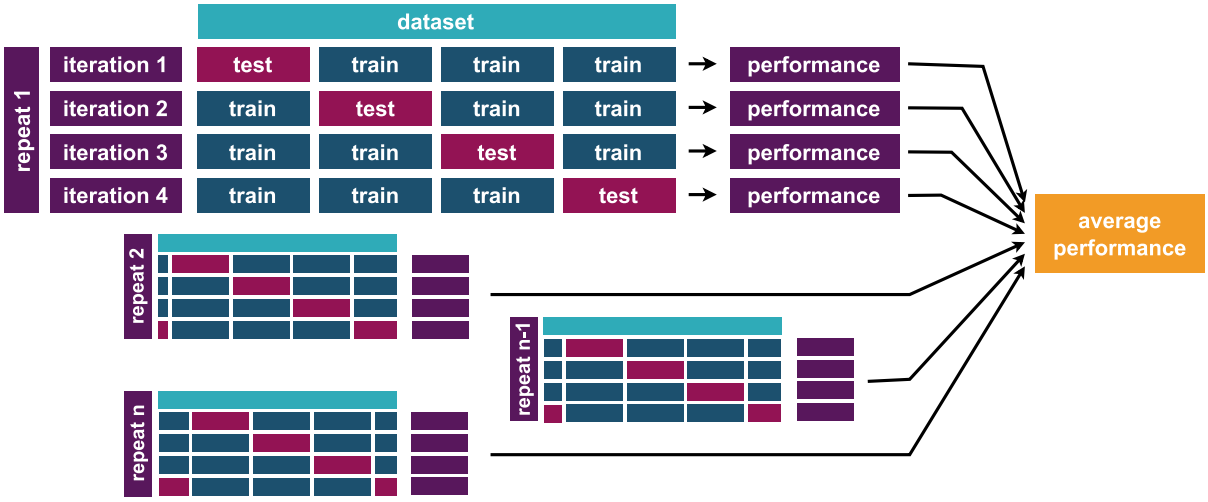


Figure 5C. Visualization of repetitive cross-validation.

3.4.1 Feature Selection

Large feature sets can compromise predictive performance of a machine learning model. Feature selection reduces the amount of correlated and redundant features whilst keeping those with high predictive value, resulting in the development of a model with lower complexity, improved predictions and smaller errors. [66] Ideally, selected features are highly correlated with the target (in this case, BPS performance) but minimally correlated with each other [67].

Many feature selection methods can be used, including basic methods such as the removal of constant features based on variance or assessing feature correlation and removing all but one of those with high correlation. [67] Initial basic feature selection was performed on the entire dataset to improve efficiency. First, correlations between features are assessed (Figure 6). It is expected that features that are highly correlated are equally informative and have similar predictive value. For example, Interquartile range and standard deviation both describe the spread of values in the dataset, where the IQR is less affected by outliers. It is expected that the IQR and std are equally informative and have similar predictive value, which can demonstrated by their correlation.

Based on pairwise correlation between feature metrics the two transformations with highest correlation to remaining metrics and lowest correlation between each other were chosen. Entire metric feature group was removed, opposed to specific movement parameter features to ensure a more homogenous feature set. Next, constant features were removed by selecting all features with zero variance.

Further feature selection was performed using the feature importance method. In this method, a model is trained using all features using cross-validation after which for each feature the average feature importance was determined. A higher feature importance indicates that the respective feature has a larger effect on the model and therefore predictions. By selecting only those features with highest importance, model dimensionality can be reduced without removing those features that are likely to have a high predictive power.

First, the regression model was trained and evaluated using all features and top-20 features to assess most important time windows. Eight time windows of three hour blocks starting at 03:00 (e.g. 03:00-06:00, 06:00-09:00, … 00:00-03:00) were separately included for model training and prediction after which features from time windows with poorest performance were removed from the feature set.

Then, a range between 2-500 features with highest importance were used to train the same model again, determining the amount of features with highest performance. To validate used method, the SelectKBest method available from the Scikit-Learn package version 1.3.2 was used, which selects a given amount of best features based on univariate linear regression test F-statistic. [59, 68]

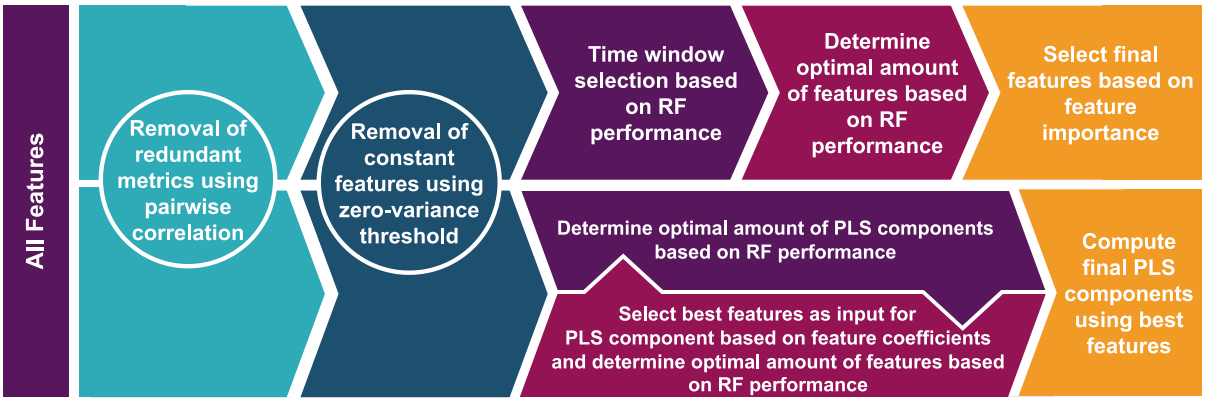


Figure 6. Process of feature selection. Upper half describes feature selection based on Random Forest feature importance, lower half describes feature selection based on feature coefficient and PLS component computation based on RF performance. *RF, random forest; PLS, partial least squares.*

3.4.2 Partial Least Squares

An alternative method for dimensionality reduction can be the use of Principal Component Analysis (PCA) or the lesser known Partial Least Squares (PLS). Both methods combine features into a predefined number of components after which these components are used as features in the predictive model. Methods like these find and combine accelerometric outcome parameters that have the highest possibility of containing relevant information. This is especially valuable in situations where only small sample sizes are available and parameters show significant covariance, which is the case for most physical activity intensity outcome measures, where an increase of a certain behavior always leads to a decrease of another behavior [69]. There are slight differences between PCA and PLS methods. PCA is a unsupervised method and aims to preserve as much variance in the original data as possible. PLS is a supervised method which aims to preserve as much covariance as possible between the accelerometric data and BPS outcomes. [70] In situations where the aim is to determine the components most predictive of known classes, PLS is preferred. [71, 72]

For optimal performance, PLS is combined with feature selection. Therefore, all feature selection based on correlation and variance as described above were applied before continuing with the optimization process using PLS. Multiple parameters were tuned during this cross-validates process, including the amount of PLS components to be used for prediction (1-5 components) and the amount of features to be used for the construction of these PLS components based on their value to the PLS component (similar to feature importance described earlier).

3.4.3 Hyperparameter Tuning

Machine learning models use certain predefined configurations called hyperparameters. These hyperparameters guide and set boundaries for how the model trains on the data. Hyperparameter tuning is the process where for a selection of configurations a range of values are tried and evaluated on performance after which optimal values are used for the model with given data. For Random Forest models, many hyperparameters can be set, with the two most commonly tuned being *n_estimators* and *max_depth*. These hyperparameters respectively determine how many trees are built and how many levels can be used within a tree (Figure 5A-B) and were tuned using a Leave-One-Out Cross-Validated grid search scoring values using negated root mean squared [73]. Initial grids were defined as *[2,3,5]* for *max_depth* and *[50,70,90,150]* for *n_estimators*. For the remainder of the hyperparameters standard values were used which can be found in documentation of *scikit-*

learn version 1.3.2. [74-76] After feature selection, models were trained once more with a more extensive hyperparameter grid including as *[2,3,5,7,9]* for *max_depth* and *[50,70,90,150,200,250,300]* for *n_estimators.*

3.5 Performance Evaluation

Both during iterative model development and for final model performance of regression models were evaluated using Root Mean Squared Error (RMSE) and coefficient of determination or R^2 over all iterations. The RMSE is a measure of the standard deviation of the difference between predicted and true values. Increased RMSE indicate worse performance. R^2 represents how well the model fits the data by looking at the proportion of the variance of the data that is explained by the model. A R^2 of 1 indicates that entire variance is explained by the model, a negative value indicates that the model is worse than when using the sample mean as predicted value. [77]

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(yi - \hat{y})^2}$$

with $\hat{y}$ being the predicted value

$$R^2 = 1 - \frac{\sum(yi - \hat{y})^2}{\sum(yi - \bar{y})^2}$$

with $\hat{y}$ being the predicted value and $\bar{y}$ the mean of the observed data

Finally, regression results were also grouped into functional categories as described in Table 1. Even though not predicted using a classification model, results could be interpreted in a similar matter, assessing performance with overall accuracy and class specific precision and recall.

$$accuracy = \frac{tp}{tp + fp + tn + fn} \qquad precision = \frac{tp}{tp + fp} \qquad recall = \frac{tp}{tp + fn}$$

With *tp* = true positives, *fp* = false positives, *tn* = true negatives, *fn* = false negatives

4

Results

4.1 Patient Characteristics

Accelerometry data of 55 participants were gathered, of which 49 resulted in both valid accelerometry and BPS outcome data, suitable for analysis. Baseline and follow-up characteristics of included patients are shown in Table 3. Overall, 65% of participants was female, mean age at time of fracture was 81.5 and 78% of participants lived independent without any additional care. Not all BPS outcome measures were available for each participant, resulting in a variable selection of participants for each analysis.

Table 3. Patient Characteristics and biopsychosocial functioning.

	At time of fracture (n=49)	Baseline (n=49)	Follow-up (n=39)*
Female, n (%)	32 (65%)		
Age (y), mean +- SD	81.5 +- 6.8		
Fracture side right, n (%)	25 (51%)		
Living situation, n			
Independent	*38*	*34*	*30*
Independent with additional care	*8*	*11*	*6*
Care facility	*3*	*1*	*2*
Unknown	*0*	*3*	*1*
Readmitted, n (%)		2 (4%)	7 (18%)
Physiotherapy in last 6 weeks, yes, n (%), frequency, mean/week ± SD		47 (96%), 3.0 ± 1.25	
FAC, mean ± SD		4.04 ± 0.61	4.39 ± 0.84
FAC, group, n			
Unknown		*1*	*1*
0-3 dependent		*8*	*2*
4 independent but limited		*30*	*16*
5 independent		*10*	*20*
6CIT, mean ± SD		3.67 ± 4.35	4.35 ± 4.76
6CIT, group, n			
Unknown		*1*	*5*
0-7 normal		*38*	*26*
8-9 risk		*5*	*3*
10-28 impaired		*5*	*5*

*6CIT, Six-Item Cognitive Impairment Test; FAC, Functional Ambulation Category; SD, standard deviation. *two patients were lost to follow-up, eight patients provided neither FAC or 6CIT.*

4.2 Feature Selection

First, correlation between feature metrics computed over original movement parameters were assessed. It was expected that the IQR and SD are equally informative and have similar predictive value, which is demonstrated by their mean correlation coefficient of *R* = 0.85 (SD 0.12) (Table 4). This also holds for SD and the maximum with *R* = 0.86 (SD 0.17), but in lesser rates for SD and mean with an *R* = 0.73 (SD 0.25). In this case, the removal of all features describing the interquartile range, maximum and the sum could be justified since most features show a good correlation with the standard deviation and/or mean (Table 4). Features describing maximum value of 10-minute rolling mean was not compared to original feature metrics and was included in feature set.

Zero-variance filtering resulted in the removal of seven features when all participants were included, all describing hourly METS_vigorous_time between 00:00-06:00, with six out of seven being a measure of standard deviation. Since not all BPS outcomes were available for all participants, it is possible that exclusion of participants based on missing data resulted in zero-variance features. For example when removing all patients without baseline FAC score resulted in the removal of one feature: the maximum of the moving window average "METS_sedentary_time" at 19.00, which was 60 seconds for all participants (one excluded patient had 59.1 seconds of sedentary time).

Table 4. Correlations between feature metrics.

Metric	**IQR**			**Max**			**Mean**			**SD**		
R (abs)	*Mean (SD)*	*≥0.70*	*≥0.90*	*Mean (SD)*	*≥0.70*	*≥0.90*	*Mean (SD)*	*≥0.70*	*≥0.90*	*Mean (SD)*	*≥0.70*	*≥0.90*
Sum	0.68 (0.27)	54%	27%	0.77 (0.18)	70%	25%	0.83 (0.21)	76%	55%	0.70 (0.26)	59%	29%
SD	0.85 (0.12)	84%	37%	0.86 (0.17)	85%	60%	0.73 (0.25)	65%	34%			
Mean	0.69 (0.26)	54%	29%	0.84 (0.15)	84%	45%						
Max	0.68 (0.22)	48%	16%									

R, correlation coefficient; abs, absolute value; SD, standard deviation; max, maximum; IQR, interquartile range.

4.2.1 3-Month Cross-sectional analysis

Since we expected best results for the prediction of cross-sectional FAC score, and both initial linear and decision tree regression showed poor performance compared to random forest regression, initial feature selection based on feature importance was based on the prediction of baseline FAC score using a cross-validated and repeated Random Forest regression model. This initial model used the following hyperparameter grid: *max depth: 2,3,5; n_estimators: 50,70,90,150* and *CV random_state: 346469* for reproducibility reasons. Time slot analysis using all features and top-20 features showed best performance for timeslots between 12:00-21:00 (Table 5). An additional analysis was run for all features within this larger timeslot, as well as the SelectKBest feature selection to verify generalized performance during that timeslot using correlation instead of feature importance. Based on these results, features describing movement behavior before 12:00 or after 21:00 were excluded from the feature set. Since feature selection based on feature importance was better compared to the results of the SelectKBest approach, this method was continued.

Next, optimal amount of features with highest importance was determined by adding different amounts of features *(2-11,13,15,20,25,30,40,50,75,150,250,500)* to the Random Forest model and assessing which amount resulted in smallest errors (RMSE) and highest R^2 score (Figure 7A). Cross-validated and repeated model was run twice to improve accuracy of results. An optimal performance

was found when only top-six features with highest importance were included. Repetitive running of the model resulted in slight shifts in which features were in the top-six of highest importance. These features described mainly duration and intensity of moving (non-sedentary) behavior (Table 6). As can be seen in Table 6, feature importance cannot be explained based on post-hoc analysis of linear correlation between feature and baseline FAC score. Of selected features, only *DUR_total_moving* parameters demonstrate a high interfeature correlation of $R = 0.87$, overall mean interfeature correlation $R = 0.43$, SD 0.16 (Appendix B, Table B1).

Similar methods were applied to the prediction of cross-sectional 6CIT score, demonstrating optimal performance with features describing movement behavior between 06:00-09:00 and 12:00-18:00 (Table 7). Optimal amount of features was established at five, which included measures of energy expenditure and transitions during both sedentary and active time (Figure 7B, Table 8). Linear correlations between features with highest importance and cross-sectional 6CIT score were low with a mean of 0.45 (SD 0.16). Of selected features mean interfeature correlation $R = 0.34$, SD 0.21 (Appendix B, Table B2), with highest interfeature correlation between *PAR_moving_mean_14:00:00* and *PAR_active_max_14:00:00* ($R = 0.74$)

Table 5. 3-month FAC, time window performance

		03:00-06:00	***06:00-09:00***	***09:00-12:00***	***12:00-15:00***	***15:00-18:00***	***18:00-21:00***	***21:00-00:00***	***00:00-03:00***	***12:00-21:00***
All features	*RMSE*	0.65	0.56	0.61	**0.48**	**0.47**	**0.5**	0.59	0.68	0.44
	R^2	-0.05	0.17	-0.02	**0.33**	**0.42**	**0.32**	0.11	-0.22	0.47
Top 20 features*	*RMSE*	0.63	0.54	0.57	**0.42**	**0.4**	**0.46**	0.54	0.68	0.36
	R^2	-0.01	0.23	0.09	**0.45**	**0.55**	**0.41**	0.21	-0.18	0.61
Select-KBest[†]	*RMSE*									0.46
	R^2									0.40

*Results of three best performing time windows shown in bold. FAC, functional ambulation category; RMSE, root mean square error; R^2, coefficient of determination. * based on mean feature importance of all features.* [†]*k=[4,6,10,20,50,75]*

Table 6. 3-month FAC, features with highest importance

Feature			**Mean feature importance**	**Correlation (Pearson)**
Movement parameter	*Metric*	*Timeslot*		*FAC*
DUR_total_moving*	Mean	14:00-15:00	0.0265	0.55
DUR_total_moving*	SD	14:00-15:00	0.026	0.63
MI_active	Mean	19:00-20:00	0.0292	0.59
MI_active	SD	13:00-14:00	0.0279	0.67
MI_active	SD	15:00-16:00	0.0919	0.66
MI_active**[†]	SD	16:00-17:00	0.0263	0.44
MI_moving	SD	15:00-16:00	0.0411	0.48

*Descriptions of movement parameters can be found in Table 2. SD, standard deviation; FAC, functional ambulation category. * featured in top-six features of highest importance in 3/4 repetitions, ranked seventh in remaining repetition. ** featured in top-six features of highest importance in 2/4 repetitions, ranked seventh and eighth in remaining repetitions.* [†]*not included in final feature set.*

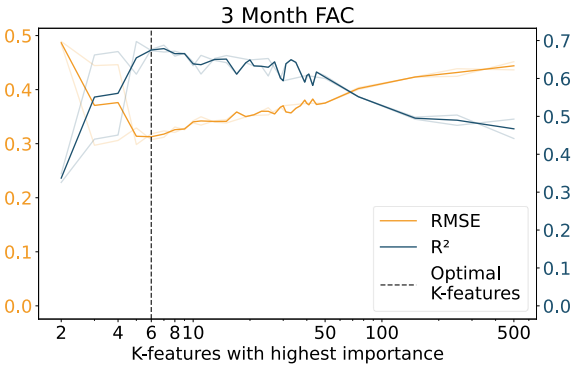


Figure 7A. Amount of features with highest performance for cross-sectional FAC. *Highest performance at 6 features.*

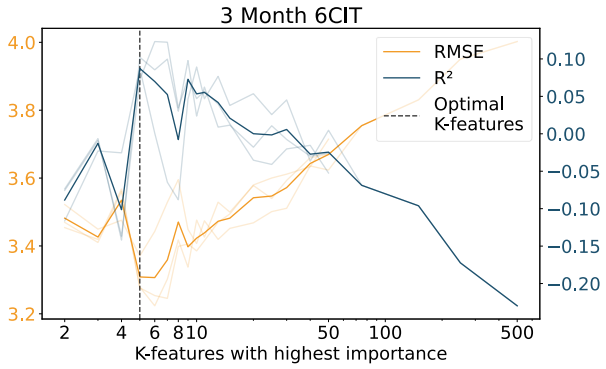


Figure 7B. Amount of features with highest performance for cross-sectional 6CIT. *Highest performance at 5 features.*

FAC, Functional ambulation category; 6CIT, Six-item Cognitive Impairment Test; RMSE, root mean square error; R^2, coefficient of determination

Table 7. 3-month 6CIT, time window performance

		03:00-06:00	***06:00-09:00***	***09:00-12:00***	***12:00-15:00***	***15:00-18:00***	***18:00-21:00***	***21:00-00:00***	***00:00-03:00***	***06:00-09:00 & 12:00-18:00***
All features	*RMSE*	4.41	**4.01**	4.25	**4.15**	**4.08**	4.57	4.22	4.35	4.06
	R^2	-0.61	**-0.25**	-0.37	**-0.31**	**-0.26**	-0.68	-0.39	-0.42	-0.24
Top 20 features*	*RMSE*	4.18	**3.65**	3.91	**3.69**	**3.72**	4.23	3.84	4.24	3.46
	R^2	-0.35	**-0.05**	-0.17	**-0.08**	**-0.07**	-0.41	-0.13	-0.34	0.06

*Results of three best performing time windows shown in bold. 6CIT, Six-item Cognitive Impairment Test; RMSE, root mean square error; R^2, coefficient of determination. * based on mean feature importance of all features.*

Table 8. 3-month 6CIT, features with highest importance

Feature			**Mean feature**	**Correlation (Pearson)**
Movement parameter	*Metric*	*Timeslot*	**importance**	*6CIT*
METS_sedentary_time*	Mean	06:00-07:00	0.0163	0.38
METS_vigorous_time*** †	SD	15:00-16:00	0.016	0.07
PAR_active	Max	14:00-15:00	0.0464	0.46
PAR_inactive*** †	Max	16:00-17:00	0.0116	0.55
PAR_inactive	SD	16:00-17:00	0.0322	0.64
PAR_moving**	Mean	14:00-15:00	0.0167	0.58
PERIODS_lying*** †	SD	16:00-17:00	0.0155	0.45
TRANSITIONS_ly	SD	17:00-18:00	0.0289	0.50

*Descriptions of movement parameters can be found in Table 2. SD, standard deviation; max, maximum; 6CIT, Six-item Cognitive Impairment Test. * featured in top-five features of highest importance in 3/4 repetitions, ranked nineth in remaining repetition. ** featured in top-five features of highest importance in 2/4 repetitions, ranked sixth and seventh in remaining repetitions. *** featured in top-five features of highest importance in 1/4 repetitions. †not included in final feature set.*

4.2.2 12-Month Longitudinal analysis

For prediction of one year follow-up FAC score, the inclusion of only features describing movement behavior between 09:00-12:00 resulted in the best performance. An additional analysis was done including both 09:00-12:00 and 15:00-18:00 but since performance did not improve, a smaller time window and therefore feature set was chosen for further analysis (Table 9). Contrary to cross-sectional analysis, the best amount of features to be included was not evidently clear (Figure 8A). By assessing mainly RMSE and R^2 score, it was decided that best 37 features would be included. Over four iterations, 52 individual features were included in top-37 features, of which 23 features were present in all iterations and 9 were present in three out of four iterations. For the final model, these features, along with five features with highest mean importance out of the 2/4 group were included (Table 10). Of selected features mean interfeature correlation R = 0.34, SD 0.21 (Appendix B, Table B4), with highest absolute interfeature correlation between *DUR_total_inactive_mean_11:00:00* and *DUR_total_active_mean_11:00:00* (R = -1)

Again, feature selection was applied for the prediction of follow-up 6CIT score. Time window with highest performance was between 03:00-06:00 and 21:00-00:00. An additional analysis was done including 03:00-06:00, 09:00-15:00 and 21:00-00:00 but since performance did not improve, a smaller time window and therefore feature set was chosen for further analysis (Table 11). Optimal amount of features was established at five, which included measures of vigorous activity, mean movement intensity, sedentary behavior and energy expenditure whilst lying (Figure 8B, Table 12). Of selected features, *METS_vigorous_time* parameters demonstrate a high interfeature correlation of R = 0.87 , overall mean absolute interfeature correlation is 0.28, SD 0.24 (Appendix B, Table B3).

Table 9. 12-month FAC, time window performance

		03:00-06:00	***06:00-09:00***	***09:00-12:00***	***12:00-15:00***	***15:00-18:00***	***18:00-21:00***	***21:00-00:00***	***00:00-03:00***
All features	*RMSE*	0.78	0.8	**0.7**	0.78	0.81	0.85	0.9	0.82
	R^2	-0.59	-0.8	**-0.2**	-0.79	-0.9	-1.12	-1.48	-0.73
Top 20 features*	*RMSE*	0.78	0.78	**0.68**	0.75	0.67	0.76	0.81	0.73
	R^2	-0.56	-0.69	**-0.03**	-0.49	-0.1	-0.57	-0.81	-0.27

*Results of best performing time window shown in bold. FAC, Functional ambulation category; RMSE, root mean square error; R^2, coefficient of determination. * based on mean feature importance of all features.*

Table 10. 12-month FAC, features with highest importance

Feature			Mean feature	Correlation (Pearson)
Movement parameter	*Metric*	*Timeslot*	**importance**	*FAC*
COUNTS_total	SD	09:00-10:00	0.008544	0.23
DUR_total_active	Max	11:00-12:00	0.010737	0.22
DUR_total_active*	Mean	11:00-12:00	0.00855	0.01
DUR_total_inactive	Mean	11:00-12:00	0.010334	-0.01
DUR_total_inactive**	SD	11:00-12:00	0.008554	0.22
DUR_total_moving**	SD	10:00-11:00	0.006946	0.42
DUR_total_sitting*	Max	09:00-10:00	0.007975	-0.14
DUR_total_sitting*	Mean	10:00-11:00	0.009669	-0.21
DUR_total_sitting	SD	11:00-12:00	0.012651	0.2
DUR_total_static	Max	11:00-12:00	0.009575	0.12
DUR_total_static	Mean	11:00-12:00	0.024169	-0.07
DUR_total_static	SD	09:00-10:00	0.015094	0.05
DUR_total_static	SD	11:00-12:00	0.019182	0.14
MET_mean*	Max	10:00-11:00	0.008244	-0.12
MET_mean	Mean	09:00-10:00	0.009678	0.11
METS_moderate_time**	Max	11:00-12:00	0.007278	0.24
METS_moderate_time*	Mean	09:00-10:00	0.007624	-0.43
METS_moderate_time	SD	10:00-12:00	0.01372	-0.18
MI_active	Mean	10:00-11:00	0.011249	-0.45
MI_active	SD	10:00-11:00	0.126699	0.43
MI_active	SD	11:00-12:00	0.031494	0.22
MI_moving	Mean	10:00-11:00	0.009348	-0.03
MI_moving	Mean	11:00-12:00	0.011896	-0.52
MI_moving*	SD	10:00-11:00	0.007738	0.48
MI_moving*	SD	11:00-12:00	0.009687	-0.66
PAR_active	Max	10:00-11:00	0.024782	0.34
PAR_active	Mean	10:00-11:00	0.008582	-0.19
PAR_active	SD	10:00-11:00	0.047444	0.42
PAR_inactive*	Max	10:00-11:00	0.009107	-0.45
PAR_lying*	Max	10:00-11:00	0.00892	-0.52
PAR_moving	Max	11:00-12:00	0.009434	-0.2
PAR_moving	SD	09:00-10:00	0.018824	0.28
PAR_moving	SD	10:00-11:00	0.021145	0.31
PERIODS_active	Mean	11:00-12:00	0.013642	0.03
PERIODS_inactive	SD	11:00-12:00	0.015588	0.15
TRANSITIONS_si**	Max	10:00-11:00	0.009	0.32
TRANSITIONS_si**	Mean	11:00-12:00	0.007611	-0.01

*Descriptions of movement parameters can be found in Table 2. SD, standard deviation; max, maximum, FAC, Functional ambulation category. * featured in top-37 features of highest importance in 3/4 repetitions. ** featured in top-37 features of highest importance in 2/4 repetitions.*

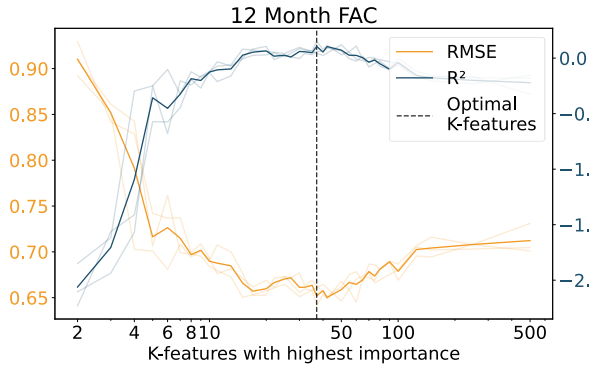


Figure 8A. Amount of features with highest performance for longitudinal FAC. *Highest performance at 37 features.*

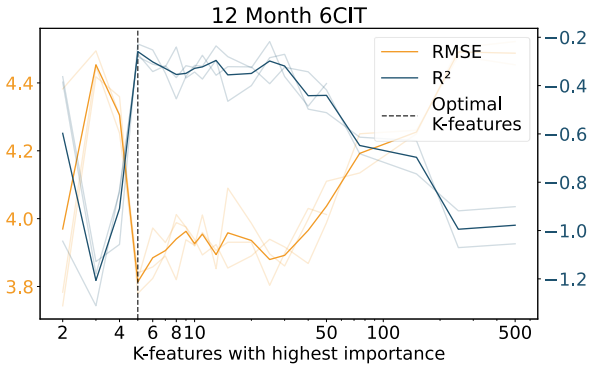


Figure 8B. Amount of features with highest performance for longitudinal 6CIT. *Highest performance at 5 features.*

FAC, Functional ambulation category; 6CIT, Six-item Cognitive Impairment Test; RMSE, root mean square error; R^2, coefficient of determination

Table 11. 12-month 6CIT, time window performance

		03:00-06:00	***06:00-09:00***	***09:00-12:00***	***12:00-15:00***	***15:00-18:00***	***18:00-21:00***	***21:00-00:00***	***00:00-03:00***	***03:00-06:00 + 21:00 - 00:00***
All features	*RMSE*	**4.51**	4.91	**4.61**	4.72	4.88	5.07	**4.55**	4.71	4.40
	R^2	**-1.26**	-1.92	**-1.22**	-1.47	-1.67	-1.84	**-1.1**	-1.5	-0.88
Top 20 features*	*RMSE*	**4.14**	4.52	4.19	4.22	4.29	4.37	**4.06**	4.49	3.90
	R^2	**-0.71**	-1.37	-0.97	-0.91	-1.01	-0.81	**-0.45**	-1.2	-0.30

*Results of best performing time windows shown in bold. 6CIT, Six-item Cognitive Impairment Test; RMSE, root mean square error; R^2, coefficient of determination. * based on mean feature importance of all features.*

Table 12. 12-month 6CIT, features with highest importance

Feature			**Mean feature importance**	**Correlation (Pearson)**
Movement parameter	*Metric*	*Timeslot*		*6CIT*
DUR_total_sitting*	Mean	22:00-23:00	0.0156	-0.31
METS_vigorous_time	Max	23:00-24:00	0.0418	0.18
METS_vigorous_time	Mean	23:00-24:00	0.1311	0.28
MI_inactive** †	Mean	22:00-23:00	0.0147	-0.19
MI_moving** †	Mean	22:00-23:00	0.0139	-0.17
MI_moving	SD	23:00-24:00	0.0347	0.07
PAR_lying	Max	22:00-23:00	0.0219	-0.05

*Descriptions of movement parameters can be found in Table 2. SD, standard deviation; max, maximum; 6CIT, Six-item Cognitive Impairment Test. * featured in top-five features of highest importance in 3/4 repetitions ** featured in top-five features of highest importance in 2/4 repetitions. †not included in final feature set.*

4.3 Model performance

To estimate model performance, models were run once more using 4-fold cross-validation repeated 10 times. Only best performing features were included and a larger hyperparameter grid was used.

Cross-sectional FAC model could explain 68% of variance (R^2 = 0.68) and predicted scores with a mean RMSE of 0.30 (Table 13). When assessing all predictions over all iterations, it is clear that there is a tendency to over-predict values as beween 3.5-4.5 (Figure 9A). Keeping in mind the categories described in Table 1, this results in mainly the misclassification of both fully independent walkers and dependent walkers as independent but limited walkers (Figure 9B). Evaluation of these grouped regression predictions results in good to excellent precision and recall for all functional categories (Table 14).

Follow-up FAC score regression showed poor performance, with only 5% of variance being explained by the model and a RMSE of 0.66 (Table 13). When assessing individual sample predictions, it is evident that the majority of participants with FAC scores 4 and 5 are predicted in the correct range, whereas participants with FAC score 1 are not (Figure 10A-B). This is clearly demonstrated when looking at regression results grouped by functional categories. This shows moderate to excellent precision and recall for FAC groups with normal/independent and risk/limited function, but poor predictive performance of those with lowest functioning (Table 14).

Both cross-sectional and longitudinal 6CIT score regression models show poor performance with respectively a RMSE of 3.07 and 3.81 and R^2 of 0.08 and -0.29 (Table 13). Regression results grouped by functional categories show good to excellent precision and recall for cross-sectional 6CIT prediction of the cognitively normal, as well as moderate precision for the cognitively impaired (Table 14). Remaining cross-sectional performance was poor. Similarly, good to excellent performance was seen for the longitudinal prediction of the cognitively normal. Moderate precision and high recall was seen for those at risk for cognitive impairment, and respectively poor performance for those cognitively impaired.

Addition of baseline FAC and 6CIT score to the feature set for the prediction of longitudinal scores did not result in improved performance of the regression model. Substituting part of feature selection with PLS also did not improve performance (Table 13).

Table 13. Final model performances of cross-sectional and longitudinal analyses.

Analysis		Based On	RMSE	R^2
Cross-sectional	FAC	Features	0.30	0.68
		PLS	0.56	0.14
	6CIT	Features	3.07	0.08
		PLS	4.41	-0.58
Longitudinal	FAC	Features	0.66	0.05
		PLS	1.08	-2.18
		Features + baseline	0.67	0.02
	6CIT	Features	3.81	-0.29
		PLS	6.23	-4.72
		Features + baseline	3.87	-0.38

Final model performances of cross-sectional and longitudinal 6CIT and FAC regression models using feature selection, PLS and for longitudinal analysis only feature selection with baseline 6CIT and FAC included. FAC, Functional ambulation category; 6CIT, Six-item Cognitive Impairment Test; RMSE, root mean square error; R^2, coefficient of determination; PLS, partial least squares.

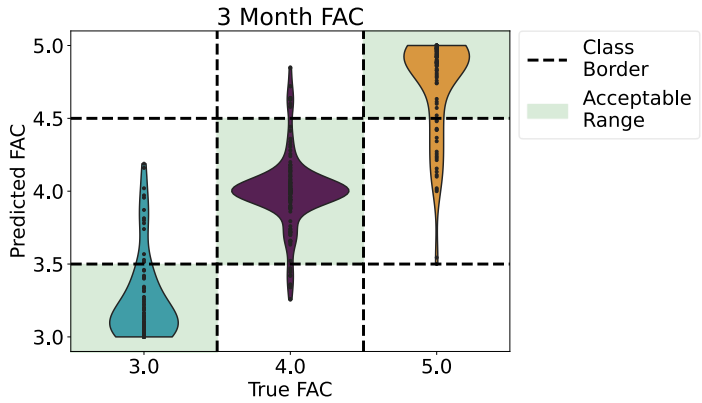


Figure 9A. Distribution of predicted FAC scores per true 3-month FAC score.

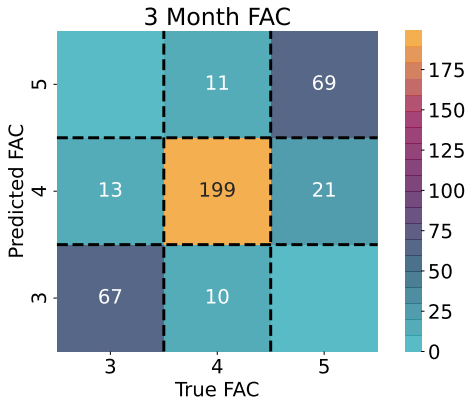


Figure 9B. Discrete predicted FAC scores per true 3-month FAC score.

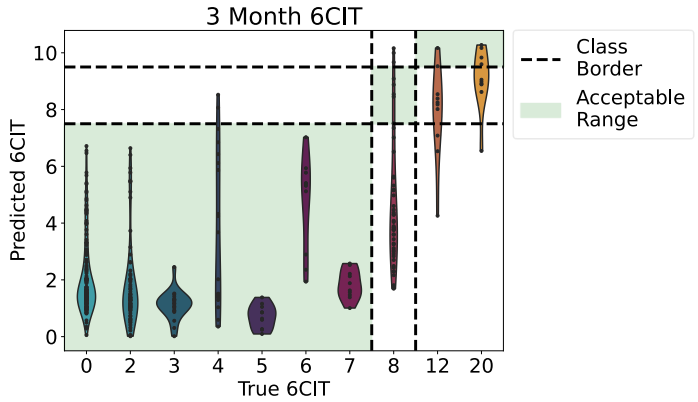


Figure 9C. Distribution of predicted 6CIT scores per true 3-month 6CIT score.

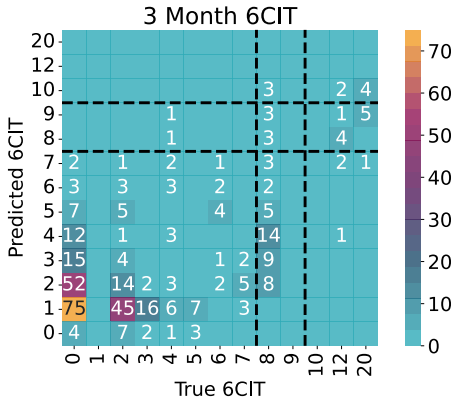


Figure 9D. Discrete predicted 6CIT scores per true 3-month 6CIT score.

Class border and acceptable range are defined based on functional categories as described in Table 1. FAC, Functional ambulation category; 6CIT, Six-item Cognitive Impairment Test.

Table 14. Model performance based on regression predictions grouped by functional categories.

	Functional Category	Cross-sectional		Longitudinal	
		FAC	6CIT	FAC	6CIT
Overall accuracy*		85.6%	84.6%	72.8%	87.2%
Precision	Normal/Independent	**0.87**	**0.88**	**0.78**	**0.88**
	Risk/Limited	**0.85**	0.28	**0.66**	**0.67**
	Impaired/Dependent	**0.84**	**0.64**	0.0	0.0
Recall	Normal/Independent	**0.77**	**0.99**	**0.74**	**1.0**
	Risk/Limited	**0.91**	0.10	**0.77**	**0.80**
	Impaired/Dependent	**0.81**	0.35	0.0	0.0

*Functional categories can be found in Table 1. Good to excellent precision and recall scores displayed in bold. FAC, Functional ambulation category; 6CIT, Six-item Cognitive Impairment Test. * note that high overall accuracy scores are biased due to data imbalance.*

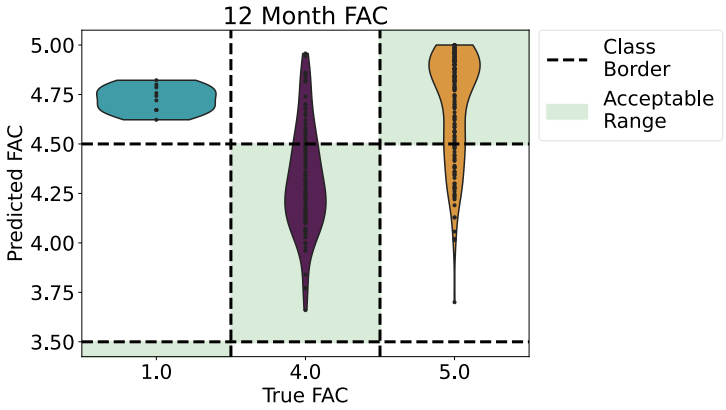


Figure 10A. Distribution of predicted FAC scores per true 12-month FAC score.

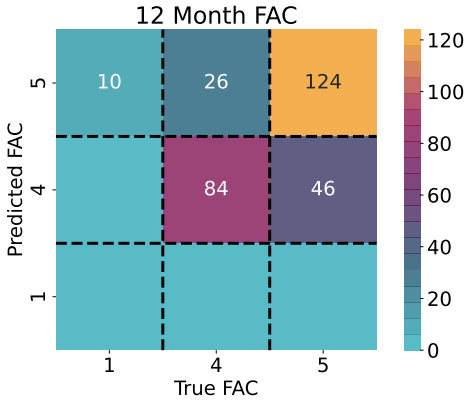


Figure 10B. Discrete predicted FAC scores per true 12-month FAC score.

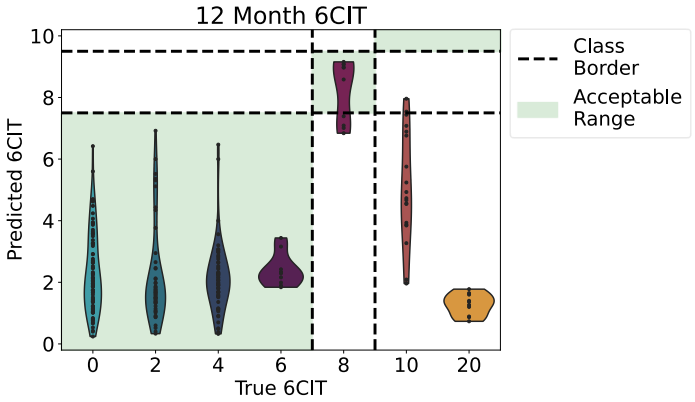


Figure 10C. Distribution of predicted 6CIT scores per true 12-month 6CIT score.

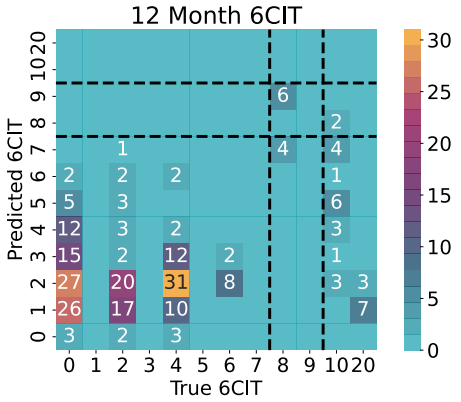


Figure 10D. Discrete predicted 6CIT scores per true 12-month 6CIT score.

Class border and acceptable range are defined based on functional categories as described in Table 1. FAC, Functional ambulation category; 6CIT, Six-item Cognitive Impairment Test.

5 Discussion

5.1 Key Findings

In this study it was found that cross-sectional Functional Ambulation Categories could be predicted using accelerometry data and a Random Forest Regression model with RMSE of 0.30 and R^2 of 0.68, indicating that the model fits the data well (Table 13). When grouping regression results into functional categories, high precision and recall values were found for all functional categories (Table 14). For cross-sectional 6CIT and both longitudinal models, regression performance was poor (Table 13). However, when assessing grouped regression predictions, performance differed between functional categories (Table 14). For cross-sectional 6CIT, the cognitively normal group could be predicted with high precision and recall (respectively 0.88 and 0.99). Prediction of those at risk of cognitive impairment was poor (precision 0.28; recall 0.10), and of the cognitively impaired had fairly good precision (0.64) but poor recall (0.35). For both longitudinal FAC and 6CIT, overall performance for normal and risk (6CIT) or limited function (FAC) groups was good to excellent, but poor for group with lowest functioning.

When taking a closer look at feature selection, it is remarkable to see that different time windows and behavior parameters have highest predictive value for each model. For cross-sectional FAC movement behavior in the afternoon and early evening (12:00-21:00) had highest predictive power (Table 5). For cross-sectional 6CIT, this respectively was the morning and afternoon (06:00-09:00, 12:00-18:00; Table 7). Different timeslots were seen for longitudinal predictions, with respectively the late morning (09:00-12:00) for FAC and early morning and evening (03:00-06:00, 21:00-00:00) for 6CIT (Table 9, Table 11). Especially the large differences between cross-sectional and longitudinal time windows were not expected. Surprisingly, available cross-sectional and longitudinal studies describing associations between physical activity throughout the day and BPS outcomes comply with found cross-sectional time windows of FAC and 6CIT, but not with found longitudinal time windows. Yi Lee and colleagues showed that participants with a delayed (after 15:00) acrophase, which describes moment of peak activity throughout the day, had worse cognitive scores after one year compared to those with averaged acrophase (between 13:24 – 15:00), which was supported by multiple other studies [78-80]. Daytime napping was also described as a risk factor for cognitive impairment, which corresponds with some features that describe sedentary behavior and lying behavior in the afternoon (Table 8) [81, 82]. Regarding feature selection for cross-sectional FAC prediction, multiple studies showed that higher total daily activity or less sedentary behavior was associated with better (long-term) physical functional status [83-87]. Schrack and colleagues also demonstrated that higher age was associated with a decrease in mainly afternoon activity, possibly explained by the preservation of routine morning activities, which could explain why the physical activity in the afternoon and evening hours may be the most informative for our predictions [83]. It is important to realize that main sleeping period at night was not included in the data. This possibly affects the results of time window analysis, reducing valuable nighttime information.

5.2 Model Performance

There are many factors of influence on how well a model is able to predict. The most important limitations and alternative methods are described in the following paragraphs, as well as the effect of covariates on predictive performance.

5.A.1 Participant Sample Size

It is important to keep in mind that model performance is biased due to small sample size and data imbalance. Only a small dataset was available for the development of the predictive models. This implies even smaller sample sizes for train and test sets when using (nested) cross-validation. Small datasets contain little data for model training, which makes them more prone to overfitting, resulting in a poorly generalized model with low performance. Due to this small sample size, an additional set of unseen data to validate a final model was also not available. Compared to other studies using accelerometry for the prediction of BPS outcomes, sample size was also low [88-90].

5.2.2 Partial Least Squares

Including PLS dimensionality reduction did not improve regression performance (Table 13). When training PLS using nested cross-validation for the determination of optimal features and PLS components, performance of tests within this extra layer of CV was superior to feature selection performance, but test set performance was equal or inferior. This suggests that possibly overfitting might (partially) explain poor performance compared to only feature selection, which is something this method is prone to [70]. Additionally, this extra layer of CV requires the already small train set to be split, reducing sample size and generalizability. Alternatively, PLS might not be a suited method for this population. Kikkert and colleagues found that gait-characteristic based PLS-discriminant analysis (PLS-DA) did correctly differentiate healthy old adults from geriatric adults, but did not differentiate between geriatric adults with and without cognitive impairment [91]. This was possibly due to non-linearity which is not captured by discriminant analysis or the effect of comorbidities and polypharmacy [90]. Since a random forest regression model was used, the problem of non-linearity should be resolved. However, Zhou and colleagues found that a random forest classification model did not result in adequate classification of geriatric patients with and without cognitive impairment, where the use of an Artificial Neural Network did [90]. This suggests that PLS may still be beneficial if the right model is chosen.

5.2.3 Baseline BPS Outcomes as Features

Surprisingly, including baseline 6CIT and FAC scores for longitudinal analysis did not improve regression performance compared to accelerometry features alone (Table 13). In fact, when adding baseline 6CIT and FAC to both longitudinal analysis as features, their importance was low compared to remainder of included features. Even if model performance would have improved, the use of baseline questionnaires as input feature in the longitudinal predictive model is an interesting methodological consideration. If the addition of this information improves model performance, which was expected, it may be useful in clinical practice. On the other hand, it increases the dependency on questionnaires with suboptimal performance themselves as described above. If the aim is to reduce our dependency on subjective questionnaires, adding their results as input of the predictive model along the accelerometric data will inhibit this development. Furthermore, it will require that these questionnaires (or more extensive alternatives) keep being conducted at baseline, possibly increasing patient burden.

5.2.4 Covariates

Different aspects of BPS functioning are closely related, suggesting that clinical variables should be considered as covariates when predicting future BPS functioning. Zhou and colleagues found that

adding clinical variables such as maximal hand grip strength, frailty index, polypharmacy and BMI did not improve RF classification, but did improve Artificial Neural Network classification [90]. Casanova and colleagues used Random Forest Classification to determine cognitive decline based on demographic and clinical characteristics, which showed that top ranked predictors included education, age, gender, stroke, NSES, diabetes, BMI and APOE ε_4 carrier status [92]. Shi and colleagues showed that the additional use of accelerometry features next to demographical and clinical variables improved the prediction of one-year and five-year cognitive decline [93]. Additionally, a previous study on HIPCARE data demonstrated that recovery success by three months (cross-sectional timepoint) was significantly related to comorbidity (American Society of Anesthesiologists classification), prefracture mobility, prefracture fear of falling and prefracture independence [51]. The included population consists of elderly hip fracture patients, who commonly have high comorbidity rates [94]. These comorbidities can be confounders when predicting BPS outcomes based on accelerometry [95].

A limitation of this study is the effect of the COVID-19 pandemic. Most participants had the 3-month baseline measurements just before the arrival of the COVID-19 in the Netherlands early 2020. The pandemic may have influenced the results of the biopsychosocial functioning at follow-up, both affecting overall biopsychosocial wellbeing of the general population due to for example restricted social interactions, reduction of exercise possibilities and the effects of possible COVID-19 infection [96]. Furthermore, it is assumed that cross-sectional analysis of accelerometry and baseline biopsychosocial functioning is not influenced by the pandemic, but since accelerometry and biopsychosocial assessment at 3-month baseline was conducted before COVID-19 restrictions were imposed for all but one participant, longitudinal analysis with follow-up biopsychosocial functioning is expected to be affected by lockdown measures. On the bright side, since only one patient was discharged from clinical geriatric rehabilitation institution after March 2020, largest part of initial (intensive) rehabilitation care had been conducted before the arrival of COVID-19.

5.2.5 Recommendations

Confounders and bias are known to be a pitfall for machine learning algorithms, since they are solely trained to minimize their error and do not differentiate between bias and true features of interest [97, 98]. To address the effect of covariates, future research should assess the effect of age, sex, BMI, smoke status, polypharmacy, previous hip fracture, comorbidity classification (ASA), living situation at time of admittance, measures of prefracture ADL independence (KATZ), mobility (PMS), fear of falling and nutritional state (SNAQ, MNA-SF), discharge destination and mobility at discharge, which are all available in the HIPCARE study. Furthermore, a more intricate machine learning regression model such as Artificial Neural Network could improve predictive performance of future BPS wellbeing based on accelerometry.

5.3 Biopsychosocial Outcomes

Model performance relies on high quality data, in this case both accelerometric data as well as biopsychosocial outcomes. In this chapter, it is explained how BPS assessments affect model performance and how this could be improved.

5.3.1 Assessment of Biopsychosocial Outcomes

A strength of this study is that true labels were determined based on clinically validated assessments, making interpretation of predicted scores easy and clinically valuable. On the other hand, these assessments still largely rely on subjective interpretation of either participant or researcher and deal with a rather rough scale with little room for nuance. Furthermore, both assessments deal with floor and ceiling effects, with large proportions of participants scoring highest

or lowest possible scores, lowering sensitivity at respective end of the scale [99]. Ideally, true labels would be determined based on more objective and extensive observation of biopsychosocial wellbeing with a large functional coverage. Or alternatively on no questionnaires at all, relying on for example data-driven, unsupervised models using clustering methods to identify patient groups and assessing their biopsychosocial characteristics after clustering.

A possible explanation for the spread between predicted cross-sectional FAC scores 3-5 is the fact that patients functional mobility could be in transition between these scores. The FAC scale is quite rough and even though clinically validated, a range of patients are grouped within each category. A possible method to improve differentiation between patient mobility is the additional assessment of the Short Physical Performance Battery (SPPB) and/or Timed Up-and-Go test results which were also conducted at both baseline and follow-up. Unfortunately, there were multiple shortcomings concerning these results in this study. First, SPPB and TUG were only performed for participants with a FAC score equal to or higher than 3 out of 5, resulting in the loss of data of a select group of participants in the data analysis, introducing bias, reducing representativeness of samples, and further decreasing overall sample size [100]. Since this may be the most relevant patient group (those with poor functional performance may benefit the most from early intervention), these BPS outcomes were not included for initial analysis. An alternative method could have been used to fill missing data by arbitrarily assigning SPPB score of 0 and TUG time of 60 seconds to this group. This does introduce a new bias, especially for regression models, assigning equal score to all poor performing patients. Second, presumably due to COVID-19 restrictions, both SPPB and TUG assessments were not conducted for the large majority of patients at time of follow-up. Since all patients at baseline had a FAC score of 3 or higher, a post-hoc analysis was done for the prediction of cross-sectional TUG (Appendix X). This demonstrates good performance of both regression predictions itself, as well as grouped predictions. This supports the hypothesis that an additional assessment with more precise scaling could be used to differentiate between patients within a single FAC score.

5.3.2 Data Imbalance

Not all BPS outcome score levels were evenly represented. For example, for cross-sectional FAC score only scores 3-5 out of the 0-5 scale were represented in the dataset, which becomes a problem if it is expected that this does not represent true patient population.

Additionally, distribution between represented scores was not even. For baseline FAC score, scores 3-5 were represented by respectively 8, 30 and 10 participants. Imbalanced data can result in a prediction preference by the model towards the values with highest occurrence in the dataset. This bias can be seen in both classification and regression models, where the model is rewarded for high performance within the training set. Favoring the majority class or value region will result in a decreased RMSE for the largest amount of samples and therefore the overall performance, whilst it actually reduces predictive performance for minority classes. [101] For example, if you have a sample of one participant with FAC score 3, four with score 4 and one with score 5, simply predicting mean score 4 for all samples results in a RMSE of 0.58, whereas if the distribution between the scores was even, RMSE would have been 0.82.

There are multiple methods to decrease the bias of imbalanced data in classification, including data-based methods (oversampling minority classes or undersampling majority classes) or model-based methods, by for example class weights or adjusting loss functions. Whilst methods like these could be used for discrete variables with limited range such as the 0-5 FAC score, they are not suited for continuous data or data where interpolation or extrapolation is needed for non-represented

outcome values. Furthermore, most resampling methods are not very suited for very small datasets since undersampling will even further reduce sample size and too little data is available for representative oversampling. Alternative methods for data imbalance in regression problems with continuous values are available but are less common. [102-104]

5.3.3 Recommendations

To improve BPS outcome prediction, a larger dataset with uniform distribution over the entire range expected in the hip fracture population, should be used. Additionally, if still slightly skewed, more advanced methods for data imbalance could be used. Finally, addition of TUG prediction next to FAC prediction could improve insights into patient mobility.

5.4 Movement Parameters and Feature Selection

Predictive power of accelerometry data largely depends on how raw data is processed into features and how features are selected. In this chapter considerations, strengths and recommendations concerning the selection of accelerometric data is discussed.

5.4.1 Movement Parameters

A strength of our study is the use of validated movement parameters as extracted by the McRoberts accelerometry algorithm [53]. On the other hand, it complicates the comparison between studies or the analysis of additional datasets that do not use McRoberts' MoveMonitor since we have little insight in their raw data analysis. Furthermore, movement parameters are developed to address a wide range of people, not specifically for the elderly. For example, using universal cut-off points for vigorous activity may not be valid for the elderly population [105].

The selection of which movement parameters to include in the analysis was based on previously performed literature review (Appendix C). This shows that total PA, moderate-to-vigorous PA, daytime sleeping, rest-activity-rhythm amplitude and relative amplitude are most feasible to have a high predictive value for future biopsychosocial wellbeing. Since our population contains community-dwelling elderly, we decided to add measures sedentary behavior to our analysis, expecting very little moderate or vigorous physical activity in our population and more predictive value within low energy activities. This selection of movement parameters was then processed and resulted in a large number of features.

5.4.2 Interfeature Correlations

There are multiple possibilities to reduce the amount of features that strongly correlate within the prediction model. First, all individual features that show high correlation with another feature could be removed, keeping only one. For example, if the feature that describes "maximum amount of steps per minute, between 14:00-15:00 during weekdays" is highly correlated to the feature that describes not the maximum but the standard deviation, one of them could be removed. This would possibly result in a heterogenous feature set, with different metrics for different movement parameters and time windows, and possible an overfitted model. The alternative is the removal of larger selections of features, possibly removing some individual features that do not correlate as much with others, but keeping a more homogenous feature set. After selecting features with highest predictive power for each model, some features showed high interfeature correlations. This shows that the removal of large sections of features based on metric correlations is not sufficient.

It is interesting to see that some correlations are lower than expected, for example the correlation between mean and sum. Further inspection of individual correlations showed that some features showed near-zero correlation between mean and sum. These correlations could be explained by non-wear time. Partially missing data within an hour can result in low sum scores and high mean

scores. For example, one patient wore the accelerometer for 15 minutes during one hour and remained sedentary during that time. This will result in a mean of 60 sedentary seconds per minute and a sum of 900 sedentary seconds for that hour. If this person had worn the accelerometer for the entire hour, a mean of 60 sedentary seconds per minute would result in a sum of 3600 sedentary seconds for that hour. If this occurs for multiple patients, correlation between sum and max reduces.

In our data analysis, we performed feature selection before tuning our hyperparameters. Theoretically, feature selection affects optimal hyperparameters, but choice of hyperparameters may also affect feature selection. Ideally, feature selection and hyperparameter tuning should be performed "at the same time", assessing all possible combinations, resulting in a very computationally expensive process which can be simplified by splitting these optimization steps. Although this is a valid approach, it may result in slightly less optimized model. [106, 107]

5.4.3 Recommendations

For future research, it would be interesting to investigate the added value of other available movement parameters, for example those describing specific activities such as walking, stair walking or cycling, as alternative measures for light, moderate and vigorous PA. Moreover, addition of heart rate monitoring could improve the quantification of movement intensity and energy expenditure, especially at lower intensities or during non-walking based activities [105, 108, 109]. Either an additional wearable or a wearable combining accelerometry and heart rate detection such as commercially available smart watches could provide this additional information.

Correlation based feature selection was a very important step in the development of the predictive models. It is recommended that correlations between metrics are further investigated, especially when correlations are lower than expected. This can be useful for the decision which metric to remove from feature set. After rough selection, additional feature removal based on interfeature correlation should be performed to prevent final feature sets containing features with high correlation.

5.5 Study Protocol

This study used a retrospective approach, which has the consequence that study protocol could not be influenced. Even though the protocol had its strengths such as the choice of accelerometer, it also resulted in some shortcomings that could be improved for future research.

5.5.1 Accelerometry

A strength of this study is the use and placement of a tri-axial accelerometer on the lower back. This captures movement with more precision compared to wrist-worn or uniaxial accelerometers [110, 111]. Keep in mind that although the placement of an accelerometer on the lower back is more desirable for data acquisition compared to placement on the (non-dominant) wrist, it may hinder patients from sleeping comfortably. Next, the use of the advanced McRoberts algorithm eliminates the need of the complementary activity diary for interpretation of accelerometry data. This reduces patient burden and decreases reliability on accurate log behavior of the patient, especially important for those with cognitive impairment.

A downside of the study protocol is that patients included in the database were asked to not wear the accelerometer at night and during water activities such as bathing. First of all, this reduces the ability to determine Rest-Activity Rhythms and corresponding features such as amplitude or relative amplitude. It also removes potentially relevant information on nighttime PA. Gianfredi and colleagues showed that patients developing depressive symptoms had a significantly higher sedentary time at night (00:00-03:00) [22]. It was also demonstrated that the relative amplitude,

which describes the ratio between daytime and nighttime activity, could be an indicator of cognitive impairment and/or dementia, which requires both valid day and night time accelerometry data [112, 113]. Second, some patients did not adhere to these instructions, requiring manual removal of main periods lying down. Both shortcomings resulted in a less uniform dataset, with some patients removing accelerometer shortly before going to sleep and replacing shortly after waking up in the morning, some patient removing the accelerometer for longer and some not removing them at all with manual removement of sleeping period. Third, by allowing participants to remove the accelerometer, some patients did not wear the accelerometer for large periods of time during the day, reducing accuracy of data analysis. Even more complicating the matter, the behavior of wearing may also reflect someone's cognitive functioning [114].

5.5.2 Duration and Timing of Accelerometric Measurement

A strength of our study is that we took the average movement behavior over multiple days. Taking the average over multiple days creates a more general image of someone's behavior and therefore is an indicator for behavior over a larger period of time [115, 116]. In this study, patients were instructed to wear the accelerometer for 7 constructive days. However, since not all participants adhered to these instructions, those with at least two valid days were included for analysis. There is no clear consensus whether there is an added value to separately analyze week- and weekend days [22, 117]. Since patients with at least two valid days were included, this sometimes resulted in either missing data if there was no week or weekend data available, or not being able to take the average over multiple days if only a single day was available. This reduces the generalizability of the samples, possible reducing quality of the predictive model. Combined with the knowledge that most of included patients were retired, week and weekend days were not analyzed separately.

Participants in our study were asked to wear the accelerometer three months after hip fracture. Declined BPS functioning including mobility, social isolation and depressive symptoms after hip fracture can take up to 2-3 years until returned to peer group levels [4]. It is possible that the timing of accelerometry, 3-months after hip fracture, may not be ideal to predict long-term BPS outcomes for this patient population.

5.5.4 Recommendations

There are some considerations that should be kept in mind when further developing the use of accelerometry. Firstly, homogeneity of accelerometry data collection should be increased, for example by using stick-on accelerometers with waterproof coating to enable 24h/day measuring without the need of patients to adhere to specific instructions. It would also be useful to assess how many days are needed at a minimum for the analysis of movement behavior to keep patient burden as low as possible whilst keeping the highest data quality for analysis [115, 116, 118-120]. Additionally, other locations should be considered if patient burden and wear compliance is an important factor. Furthermore, the type of accelerometer should be considered. Alternatives for devices like the MoveMonitor could be the use of widely available smartwatches or fitness trackers such as a Fitbit devices or smartphones. Tradeoffs between patient comfort, accuracy and costs should be considered. Finally, alternative timelines for accelerometric measurement should be considered, e.g. after 1 month (for even earlier intervention), 6 months (more time for rehabilitation before assessment) or personalized to each patient, for example after clinical rehabilitation facility discharge.

5.6 Clinical Relevance

The applicability of new techniques in the medical scene is arguably the most important part of the development process. Here it is illustrated how results could affect medical care, but especially how clinical issues could influence model development.

5.6.1 Interpretation of Physical Activity

Although some studies show that physical activity behavior changes can alter for example mental wellbeing or decrease risk of depression [121-123], it can also be hypothesized that physical activity behavior as described in this study does not cause functional decline but is a marker of underlying (subclinical) morbidities causing long-term decline or lack of improvement [124, 125]. For example, Maasakkers and colleagues found some associations between physical activity and follow-up measures of cognition, but also some associations between physical activity and preceding cognitive changes [126]. Furthermore, it has been suggested that subtle changes in physical activity may be a marker for impending functional decline and disability [105]. This indicates that differences in physical activity throughout the day are not necessarily the cause, but merely a symptom or marker of BPS functioning.

5.6.2 Intervention Oriented Modeling

Intervention availability determines which patient groups should be detected with highest sensitivity and specificity. For example, the predictive model could be trained to have high sensitivity for those patients that are expected to be at risk of impairment or show mild impairment at follow-up if they are most likely to benefit from additional preventive treatment. Additionally, even though BPS outcomes are clinically validated, ranges could be slightly altered for the interpretation of model predictions. For example, if a patient would score a 6CIT score at the edge between normal and risk group, a medical specialist may be tempted to monitor the patient more closely compared to someone scoring at the far end of the normal range. This effect of caution may even increase when interpreting long-term predictions. Knowing which patient groups are main focus of intervention can determine how to interpret predictive error magnitude.

It is possible that not endpoint functioning but rate of decline or improvement is most valuable for the initiation of treatment. Both for BPS outcomes with a larger scale range such as the 6CIT, where a decline within the same functional category may indicate a subclinical deterioration which could be stopped with adequate intervention, as well as small scale range such as the FAC, where a long-term prediction of say, 3.9, could indicate a steady state for a patient with a baseline score of 4, but a prediction of 3.6 could indicate a slow decline over time which could become clinically relevant a few months later. Assuming that this is possible, adding additional follow-up measurements or a more detailed mobility assessment would be essential to demonstrate a trend of deterioration or improvement since FAC score itself is limited to discrete values and therefore can only describe larger changes in functioning

5.6.3 Recommendations

First of all, it is important to determine what early interventions can be performed to alter long-term BPS outcome after hip fracture. A separate study should demonstrate that early intervention results in improved BPS functioning on the long-term. Similar to the general principles of screening, there should be an acceptable treatment or for predicted risks and impairments [127]. If no intervention is available, acquired knowledge should for example enable carefully considered choices improving quality of life. It could also be considered if only a subset of patients should be included for accelerometry monitoring, keeping both costs and patient burden as low as possible. For example, it

could be assessed if patients that are admitted for clinical rehabilitation benefit more or less of this prediction model than those discharged from the hospital with only outpatient rehabilitation care.

This being established, it is important to determine which patient groups should be identified with highest accuracy and what level of predictive accuracy is needed to add value to patient treatment in daily practice. If satisfactory predictive performance is achieved, methods and models should be validated on a larger patient population to determine generalizability.

5.7 Take Home Message

To my knowledge, this study is the first attempt to use objective home-based accelerometric data-driven machine learning models to predict cross-sectional and longitudinal biopsychosocial functioning of individual patients after hip fracture in the elderly. Overall, cross-sectional predictive model for mobility (FAC) had a good performance. Cross-sectional cognition (6CIT) and both longitudinal regression models underperformed, but categorized regression predictions revealed more promising performance. It is expected that predictive performance of models can be improved by increasing participant sample size with balanced samples over population specific prevalent range of BPS outcome scales and the exploration of additional machine learning models. In the future, accurate accelerometry-based predictions for individual patients in need of rehabilitation could support personalized treatment and improve long-term biopsychosocial functioning.

References

1. Wade DT, Halligan PW. The biopsychosocial model of illness: a model whose time has come. Clin Rehabil. 2017;31(8):995-1004.
2. Wade DT. Complexity, case-mix and rehabilitation: the importance of a holistic model of illness. Clin Rehabil. 2011;25(5):387-95.
3. Engel GL. The need for a new medical model: a challenge for biomedicine. Science. 1977;196(4286):129-36.
4. Swayambunathan J, Dasgupta A, Bhattacharyya T. The Pronounced Impact of Hip Fractures on Psychosocial Well-being. J Am Acad Orthop Surg. 2021;29(1):e22-e30.
5. Visschedijk J, Achterberg W, Van Balen R, Hertogh C. Fear of falling after hip fracture: a systematic review of measurement instruments, prevalence, interventions, and related factors. J Am Geriatr Soc. 2010;58(9):1739-48.
6. Alexiou KI, Roushias A, Varitimidis SE, Malizos KN. Quality of life and psychological consequences in elderly patients after a hip fracture: a review. Clin Interv Aging. 2018;13:143-50.
7. Sharma M, Khan SA, Chhabra C. Predictive Analytics In Rehabilitation: A Guide Using Artificial Intelligence Towards Better Care. 2023. In: AI Horizons: Exploring Multidisciplinary Frontiers Volume - II [Internet]. Book Rivers; [102-22]. Available from: https://www.researchgate.net/publication/373927294_CHAPTER_-7_PREDICTIVE_ANALYTICS_IN_REHABILITATION_A_GUIDE_USING_ARTIFICIAL_INTELLIGENCE_TOWARDS_BETTER_CARE.
8. Cirstea MC, Levin MF. Compensatory strategies for reaching in stroke. Brain. 2000;123 (Pt 5):940-53.
9. Tomaszewski Farias S, Schmitter-Edgecombe M, Weakley A, Harvey D, Denny KG, Barba C, Gravano JT, Giovannetti T, Willis S. Compensation Strategies in Older Adults: Association With Cognition and Everyday Function. Am J Alzheimers Dis Other Demen. 2018;33(3):184-91.
10. Godfrey A, Conway R, Meagher D, G OL. Direct measurement of human movement by accelerometry. Med Eng Phys. 2008;30(10):1364-86.
11. Rodrigues B, Encantado J, Carraca E, Sousa-Sa E, Lopes L, Cliff D, Mendes R, Silva MN, Godinho C, Santos R. Questionnaires measuring movement behaviours in adults and older adults: Content description and measurement properties. A systematic review. PLoS One. 2022;17(3):e0265100.
12. Chen PH, Wang RL, Liou DJ, Shaw JS. Gait Disorders in Parkinson's Disease: Assessment and Management. Int J Gerontol. 2013;7(4):189-93.
13. Gebruers N, Truijen S, Engelborghs S, De Deyn PP. Predictive value of upper-limb accelerometry in acute stroke with hemiparesis. J Rehabil Res Dev. 2013;50(8):1099-106.
14. Bento T, Cortinhas A, Leitao JC, Mota MP. Use of accelerometry to measure physical activity in adults and the elderly. Rev Saude Publica. 2012;46(3):561-70.
15. Peluso MA, Guerra de Andrade LH. Physical activity and mental health: the association between exercise and mood. Clinics (Sao Paulo). 2005;60(1):61-70.
16. Rice J, Corp DT, Swarowsky A, Cahalin LP, Cabral DF, Nunez C, Koch S, Rundek T, Gomes-Osman J. Greater Cognitive-Motor Interference in Individuals Post-Stroke During More Complex Motor Tasks. J Neurol Phys Ther. 2022;46(1):26-33.
17. Al-Yahya E, Dawes H, Smith L, Dennis A, Howells K, Cockburn J. Cognitive motor interference while walking: a systematic review and meta-analysis. Neurosci Biobehav Rev. 2011;35(3):715-28.
18. Vancampfort D, Firth J, Schuch FB, Rosenbaum S, Mugisha J, Hallgren M, Probst M, Ward PB, Gaughran F, De Hert M, Carvalho AF, Stubbs B. Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta-analysis. World Psychiatry. 2017;16(3):308-15.
19. Dutil C, Podinic I, Sadler CM, da Costa BG, Janssen I, Ross-White A, Saunders TJ, Tomasone JR, Chaput JP. Sleep timing and health indicators in children and adolescents: a systematic review. Health Promot Chronic Dis Prev Can. 2022;42(4):150-69.
20. Schuch F, Vancampfort D, Firth J, Rosenbaum S, Ward P, Reichert T, Bagatini NC, Bgeginski R, Stubbs B. Physical activity and sedentary behavior in people with major depressive disorder: A systematic review and meta-analysis. J Affect Disord. 2017;210:139-50.

21. Domingos C, Pego JM, Santos NC. Effects of physical activity on brain function and structure in older adults: A systematic review. Behav Brain Res. 2021;402:113061.
22. Gianfredi V, Schaper NC, Odone A, Signorelli C, Amerio A, Eussen S, Kohler S, Savelberg H, Stehouwer CDA, Dagnelie PC, Henry RMA, van der Kallen CJH, van Greevenbroek MMJ, Schram MT, Koster A. Daily patterns of physical activity, sedentary behavior, and prevalent and incident depression-The Maastricht Study. Scand J Med Sci Sports. 2022;32(12):1768-80.
23. Choo YJ, Chang MC. Use of Machine Learning in Stroke Rehabilitation: A Narrative Review. Brain & Neurorehabilitation. 2022;15(3).
24. Tschuggnall M, Grote V, Pirchl M, Holzner B, Rumpold G, Fischer MJ. Machine learning approaches to predict rehabilitation success based on clinical and patient-reported outcome measures. Informatics in Medicine Unlocked. 2021;24.
25. Mennella C, Maniscalco U, Pietro G, Esposito M. A deep learning system to monitor and assess rehabilitation exercises in home-based remote and unsupervised conditions. Comput Biol Med. 2023;166:107485.
26. Mennella C, Maniscalco U, De Pietro G, Esposito M. The Role of Artificial Intelligence in Future Rehabilitation Services: A Systematic Literature Review. IEEE Access. 2023;11:11024-43.
27. Boffetta P, Collatuzzo G. Application of P4 (Predictive, Preventive, Personalized, Participatory) Approach to Occupational Medicine. Med Lav. 2022;113(1):e2022009.
28. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford). 2020;2020.
29. Ross MK, Tulabandhula T, Bennett CC, Baek E, Kim D, Hussain F, Demos AP, Ning E, Langenecker SA, Ajilore O, Leow AD. A Novel Approach to Clustering Accelerometer Data for Application in Passive Predictions of Changes in Depression Severity. Sensors (Basel). 2023;23(3).
30. Nait Aicha A, Englebienne G, van Schooten KS, Pijnappels M, Krose B. Deep Learning to Predict Falls in Older Adults Based on Daily-Life Trunk Accelerometry. Sensors (Basel). 2018;18(5).
31. Dasgupta P, VanSwearingen J, Sejdic E. "You can tell by the way I use my walk." Predicting the presence of cognitive load with gait measurements. Biomed Eng Online. 2018;17(1):122.
32. Barth J, Lohse KR, Bland MD, Lang CE. Predicting later categories of upper limb activity from earlier clinical assessments following stroke: an exploratory analysis. J Neuroeng Rehabil. 2023;20(1):24.
33. Brownlee J. "A Tour of Machine Learning Algorithms." Available from: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/ (accessed
34. Koehrsen W. "Random Forest Simple Explanation." Available from: https://williamkoehrsen.medium.com/random-forest-simple-explanation-377895a60d2d (accessed 06-01-2024).
35. Fernandez-Lozano C, Hervella P, Mato-Abad V, Rodriguez-Yanez M, Suarez-Garaboa S, Lopez-Dequidt I, Estany-Gestal A, Sobrino T, Campos F, Castillo J, Rodriguez-Yanez S, Iglesias-Rey R. Random forest-based prediction of stroke outcome. Sci Rep. 2021;11(1):10071.
36. Shi G, Liu G, Gao Q, Zhang S, Wang Q, Wu L, He P, Yu Q. A random forest algorithm-based prediction model for moderate to severe acute postoperative pain after orthopedic surgery under general anesthesia. BMC Anesthesiol. 2023;23(1):361.
37. Naveed SS. Prediction of Breast Cancer Through Random Forest. Curr Med Imaging. 2023;19(10):1144-55.
38. Towards a Common Language for Functioning, Disability and Health: ICF. Geneva: WHO; 2002.
39. Bordne S, Rietz C, Schulz RJ, Zank S. Subjective well-being of geriatric patients during and after inpatient geriatric rehabilitation: a biopsychosocial prediction model. Eur Geriatr Med. 2019;10(6):965-75.
40. Eleuteri S, de Lima MEB, Falaschi P. The Psychological Health of Patients and their Caregivers. In: Falaschi P, Marsh D, editors. Orthogeriatrics: The Management of Older Patients with Fragility Fractures. 2nd ed. Cham (CH)2021. p. 223-37.
41. Jonsdottir HL, Ruthig JC. A longitudinal study of the negative impact of falls on health, well-being, and survival in later life: the protective role of perceived control. Aging Ment Health. 2021;25(4):742-8.
42. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJ, Ostelo RW, Guzman J, van Tulder MW. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev. 2014(9):CD000963.
43. Pin S, Spini D. Impact of falling on social participation and social support trajectories in a middle-aged and elderly European sample. SSM Popul Health. 2016;2:382-9.

44. Adomaviciene A, Daunoraviciene K, Sidlauskaite R, Griskevicius J, Kubilius R, Varzaityte L, Raistenskis J. Analysis of Functional Recovery and Subjective Well-Being after Arthroscopic Rotator Cuff Repair. Medicina (Kaunas). 2021;57(7).
45. Buecking B, Struewer J, Waldermann A, Horstmann K, Schubert N, Balzer-Geldsetzer M, Dodel R, Bohl K, Ruchholtz S, Bliemel C. What determines health-related quality of life in hip fracture patients at the end of acute care?--a prospective observational study. Osteoporos Int. 2014;25(2):475-84.
46. Guo L, An L, Luo F, Yu B. Social isolation, loneliness and functional disability in Chinese older women and men: a longitudinal study. Age Ageing. 2021;50(4):1222-8.
47. Shankar A, McMunn A, Demakakos P, Hamer M, Steptoe A. Social isolation and loneliness: Prospective associations with functional status in older adults. Health Psychol. 2017;36(2):179-87.
48. Alonso SG, de la Torre Diez I, Zapirain BG. Predictive, Personalized, Preventive and Participatory (4P) Medicine Applied to Telemedicine and eHealth in the Literature. J Med Syst. 2019;43(5):140.
49. Haghi M, Danyali S, Ayasseh S, Wang J, Aazami R, Deserno TM. Wearable Devices in Health Monitoring from the Environmental towards Multiple Domains: A Survey. Sensors (Basel). 2021;21(6).
50. Longo UG, Carnevale A, Massaroni C, Lo Presti D, Berton A, Candela V, Schena E, Denaro V. Personalized, Predictive, Participatory, Precision, and Preventive (P5) Medicine in Rotator Cuff Tears. J Pers Med. 2021;11(4).
51. van der Sijp MPL, van Eijk M, Niggebrugge AHP, Putter H, Blauw GJ, Achterberg WP. Prognostic Factors for Short-term Recovery of Independence in a Multistate Model for Patients With a Hip Fracture. J Am Med Dir Assoc. 2021;22(6):1307-12.
52. Mattiazzo GF, Drewes YM, van Eijk M, Achterberg WP. Geriatric rehabilitation care after hip fracture. Eur Geriatr Med. 2023;14(2):295-305.
53. Dijkstra B, Kamsma Y, Zijlstra W. Detection of gait and postures using a miniaturised triaxial accelerometer-based system: accuracy in community-dwelling older adults. Age Ageing. 2010;39(2):259-62.
54. Langer D, Gosselink R, Sena R, Burtin C, Decramer M, Troosters T. Validation of two activity monitors in patients with COPD. Thorax. 2009;64(7):641-2.
55. "McRoberts Modules." Available from: https://www.mcroberts.nl/module/ (accessed 04-01-2024).
56. Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng. 2007;9(3):90-5.
57. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE. Array programming with NumPy. Nature. 2020;585(7825):357-62.
58. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. J Mach Learn Res. 2017;18.
59. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825-30.
60. Waskom ML. seaborn: statistical data visualization. The Journal of Open Source Software. 2021;6:3021.
61. McKinnew W. Data structures for statistical computing in python. 9th Python in Science Conference2010. p. 56-61.
62. "Six Item Cognitive Impairment Test (6CIT)." Available from: https://gpnotebook.com/pages/geriatric-medicine/six-item-cognitive-impairment-test-(6cit) (accessed 06-11-2023).
63. J. UNODBFM. UNCO-MOB 2.1 Dé fysiotherapeutische meetinstrumentenset voor de geriatrische cliënt. 2018. [33-5]. Available from: https://unoamsterdam.nl/wp-content/uploads/2017/06/UNCO-MOB-2.1.pdf.
64. "sklearn.preprocessing.StandardScaler." Available from: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html (accessed 19-12-2023).
65. Poldrack RA, Huckins G, Varoquaux G. Establishment of Best Practices for Evidence for Prediction: A Review. JAMA Psychiatry. 2020;77(5):534-40.
66. Chen RC, Dewi C, Huang SW, Caraka RE. Selecting critical features for data classification based on machine learning methods. J Big Data-Ger. 2020;7(1).
67. Banerjee P. "Comprehensive Guide on Feature Selection." Available from: https://www.kaggle.com/code/prashant111/comprehensive-guide-on-feature-selection (accessed 15-12-2023).

68. "sklearn.feature_selection.SelectKBest." Scikit-Learn. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html (accessed 16-12-2023).
69. Cabanas-Sanchez V, Esteban-Cornejo I, Garcia-Esquinas E, Ortola R, Ara I, Rodriguez-Gomez I, Chastin SFM, Rodriguez-Artalejo F, Martinez-Gomez D. Cross-sectional and prospective associations of sleep, sedentary and active behaviors with mental health in older people: a compositional data analysis from the Seniors-ENRICA-2 study. Int J Behav Nutr Phys Act. 2021;18(1):124.
70. Ruiz-Perez D, Guan HB, Madhivanan P, Mathee K, Narasimhan G. So you think you can PLS-DA? Bmc Bioinformatics. 2020;21.
71. Liu C, Zhang X, Nguyen TT, Liu J, Wu T, Lee E, Tu XM. Partial least squares regression and principal component analysis: similarity and differences between two popular variable reduction approaches. Gen Psychiatr. 2022;35(1):e100662.
72. "Principal Component Regression vs Partial Least Squares Regression." Available from: https://scikit-learn.org/stable/auto_examples/cross_decomposition/plot_pcr_vs_pls.html (accessed 19-12-2023).
73. "Metrics and scoring: quantifying the quality of predictions." Available from: https://scikit-learn.org/stable/modules/model_evaluation.html (accessed 19-12-2023).
74. "sklearn.ensemble.RandomForestRegressor." Skicit-Learn. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html (accessed 19-12-2023).
75. Ellis C. "Max depth in random forests." Available from: https://crunchingthedata.com/max-depth-in-random-forests/ (accessed 19-12-2023).
76. "sklearn.model_selection.GridSearchCV." Available from: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (accessed 19-12-2023).
77. Chugh A. "MAE, MSE, RMSE, Coefficient of Determination, Adjusted R Squared — Which Metric is Better?" Available from: https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e (accessed 19-12-2023).
78. Yi Lee PM, Ling Kwok BH, Ting Ma JY, Tse LA. A population-based prospective study on rest-activity rhythm and mild cognitive impairment among Hong Kong healthy community-dwelling older adults. Neurobiol Sleep Circadian Rhythms. 2021;10:100065.
79. Tranah GJ, Blackwell T, Stone KL, Ancoli-Israel S, Paudel ML, Ensrud KE, Cauley JA, Redline S, Hillier TA, Cummings SR, Yaffe K, Group SOFR. Circadian activity rhythms and risk of incident dementia and mild cognitive impairment in older women. Ann Neurol. 2011;70(5):722-32.
80. Rogers-Soeder TS, Blackwell T, Yaffe K, Ancoli-Israel S, Redline S, Cauley JA, Ensrud KE, Paudel M, Barrett-Connor E, LeBlanc E, Stone K, Lane NE, Tranah G, Osteoporotic Fractures in Men Study Research G. Rest-Activity Rhythms and Cognitive Decline in Older Men: The Osteoporotic Fractures in Men Sleep Study. J Am Geriatr Soc. 2018;66(11):2136-43.
81. Leng Y, Redline S, Stone KL, Ancoli-Israel S, Yaffe K. Objective napping, cognitive decline, and risk of cognitive impairment in older men. Alzheimers Dement. 2019;15(8):1039-47.
82. Li P, Gao L, Yu L, Zheng X, Ulsa MC, Yang HW, Gaba A, Yaffe K, Bennett DA, Buchman AS, Hu K, Leng Y. Daytime napping and Alzheimer's dementia: A potential bidirectional relationship. Alzheimers Dement. 2023;19(1):158-68.
83. Schrack JA, Zipunnikov V, Goldsmith J, Bai J, Simonsick EM, Crainiceanu C, Ferrucci L. Assessing the "physical cliff": detailed quantification of age-related differences in daily patterns of physical activity. J Gerontol A Biol Sci Med Sci. 2014;69(8):973-9.
84. Thralls KJ, Godbole S, Manini TM, Johnson E, Natarajan L, Kerr J. A comparison of accelerometry analysis methods for physical activity in older adult women and associations with health outcomes over time. J Sports Sci. 2019;37(20):2309-17.
85. Reuter C, Bellettiere J, Liles S, Di CZ, Sears DD, LaMonte MJ, Stefanick ML, LaCroix AZ, Natarajan L. Diurnal patterns of sedentary behavior and changes in physical function over time among older women: a prospective cohort study. Int J Behav Nutr Phy. 2020;17(1).
86. Floegel TA, Dickinson JM, DerAnanian C, McCarthy M, Hooker SP, Buman MP. Association of Posture and Ambulation With Function 30 Days After Hospital Discharge in Older Adults with Heart Failure. J Card Fail. 2018;24(2):126-30.
87. Sun R, Tomkins-Lane C, Muaremi A, Kuwabara A, Smuck M. Physical activity thresholds for predicting longitudinal gait decline in adults with knee osteoarthritis. Osteoarthritis Cartilage. 2021;29(7):965-72.

88. Liu Z, Zhang L, Wu J, Zheng Z, Gao J, Lin Y, Liu Y, Xu H, Zhou Y. Machine learning-based classification of circadian rhythm characteristics for mild cognitive impairment in the elderly. Front Public Health. 2022;10:1036886.
89. Hasegawa S, Mizokami F, Kameya Y, Hayakawa Y, Watanabe T, Matsui Y. Machine learning versus binomial logistic regression analysis for fall risk based on SPPB scores in older adult outpatients. Digit Health. 2023;9:20552076231219438.
90. Zhou YvC, J.; Hortobágyi, T.; Lamoth, C. Artificial neural network to classify cognitive impairment using gait and clinical variables. Intelligence-Based Medicine. 2022;6.
91. Kikkert LHJ, Vuillerme N, van Campen JP, Appels BA, Hortobagyi T, Lamoth CJC. Gait characteristics and their discriminative power in geriatric patients with and without cognitive impairment. J Neuroeng Rehabil. 2017;14(1):84.
92. Casanova R, Saldana S, Lutz MW, Plassman BL, Kuchibhatla M, Hayden KM. Investigating Predictors of Cognitive Decline Using Machine Learning. J Gerontol B Psychol Sci Soc Sci. 2020;75(4):733-42.
93. Shi C, Babiker N, Urbanek JK, Grossman RL, Huisingh-Scheetz M, Rzhetsky A. Free-living wrist and hip accelerometry forecast cognitive decline among older adults without dementia over 1- or 5-years in two distinct observational cohorts. NPJ Aging. 2022;8(1):7.
94. Lloyd R, Baker G, MacDonald J, Thompson NW. Co-morbidities in Patients with a Hip Fracture. Ulster Med J. 2019;88(3):162-6.
95. Duffy G, Clarke SL, Christensen M, He B, Yuan N, Cheng S, Ouyang D. Confounders mediate AI prediction of demographics in medical imaging. NPJ Digit Med. 2022;5(1):188.
96. Oliveira MR, Sudati IP, Konzen VM, de Campos AC, Wibelinger LM, Correa C, Miguel FM, Silva RN, Borghi-Silva A. Covid-19 and the impact on the physical activity level of elderly people: A systematic review. Exp Gerontol. 2022;159:111675.
97. Arjovsky MB, L.; Gulrajani, I.; Lopez-Paz D. Invariant Risk Minimization2020. Available from: https://arxiv.org/pdf/1907.02893.pdf.
98. . causaLens: Causal AI. Available from: https://causalens.com/resources/blogs/confounders-machine-learnings-blindspot/.
99. Gulledge CM, Smith DG, Ziedas A, Muh SJ, Moutzouros V, Makhni EC. Floor and Ceiling Effects, Time to Completion, and Question Burden of PROMIS CAT Domains Among Shoulder and Knee Patients Undergoing Nonoperative and Operative Treatment. JB JS Open Access. 2019;4(4).
100. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013;64(5):402-6.
101. Lin WJ, Chen JJ. Class-imbalanced classifiers for high-dimensional data. Brief Bioinform. 2013;14(1):13-26.
102. Tang Y. "Strategies and Tactics for Regression on Imbalanced Data." Available from: https://towardsdatascience.com/strategies-and-tactics-for-regression-on-imbalanced-data-61eeb0921fca (accessed 28-12-2023).
103. "10 Techniques to Solve Imbalanced Classes in Machine Learning." Available from: https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/#:~:text=Non%2Dfraudulent).-,Resampling%20Techniques%20to%20Solve%20Class%20Imbalance,class%20(over%2Dsampling (accessed 28-12-2023).
104. Batista GPCM, M. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter. 2004;6(1):20-9.
105. Schrack JA, Leroux A, Fleg JL, Zipunnikov V, Simonsick EM, Studenski SA, Crainiceanu C, Ferrucci L. Using Heart Rate and Accelerometry to Define Quantity and Intensity of Physical Activity in Older Adults. J Gerontol A Biol Sci Med Sci. 2018;73(5):668-75.
106. Kuhn MJ, K. Feature Engineering and Selection: A Practical Approach for Predictive Models. 2019. Taylor & Francis Group. Available from: https://bookdown.org/max/FES/summary-8.html.
107. Soemers D. How should Feature Selection and Hyperparameter optimization be ordered in the machine learning pipeline? Stack Exchange2018.
108. Carbonell-Hernandez L, Pastor D, Jimenez-Loaisa A, Ballester-Ferrer JA, Montero-Carretero C, Cervello E. Lack of Correlation between Accelerometers and Heart-Rate Monitorization during Exercise Session in Older Adults. Int J Environ Res Public Health. 2020;17(15).
109. Akerberg A, Koshmak G, Johansson A, Linden M. Heart rate measurement as a tool to quantify sedentary behavior. Stud Health Technol Inform. 2015;211:105-10.

110. Duncan MJ, Rowlands A, Lawson C, Leddington Wright S, Hill M, Morris M, Eyre E, Tallis J. Using accelerometry to classify physical activity intensity in older adults: What is the optimal wear-site? Eur J Sport Sci. 2020;20(8):1131-9.
111. Sagelv EH, Ekelund U, Pedersen S, Brage S, Hansen BH, Johansson J, Grimsgaard S, Nordstrom A, Horsch A, Hopstock LA, Morseth B. Physical activity levels in adults and elderly from triaxial and uniaxial accelerometry. The Tromso Study. PLoS One. 2019;14(12):e0225670.
112. Posner AB, Tranah GJ, Blackwell T, Yaffe K, Ancoli-Israel S, Redline S, Leng Y, Zeitzer JM, Chen DM, Webber KR, Stone KL. Predicting incident dementia and mild cognitive impairment in older women with nonparametric analysis of circadian activity rhythms in the Study of Osteoporotic Fractures. Sleep. 2021;44(10).
113. Xiao Q, Sampson JN, LaCroix AZ, Shadyab AH, Zeitzer JM, Ancoli-Israel S, Yaffe K, Stone K, Osteoporotic Fractures in Men Study G. Nonparametric Parameters of 24-Hour Rest-Activity Rhythms and Long-Term Cognitive Decline and Incident Cognitive Impairment in Older Men. J Gerontol A Biol Sci Med Sci. 2022;77(2):250-8.
114. Chen JH, Lauderdale DS. Cognitive Function, Consent for Participation, and Compliance With Wearable Device Protocols in Older Adults. J Gerontol A Biol Sci Med Sci. 2019;74(2):269-73.
115. Sasaki JE, Junior JH, Meneguci J, Tribess S, Marocolo Junior M, Stabelini Neto A, Virtuoso Junior JS. Number of days required for reliably estimating physical activity and sedentary behaviour from accelerometer data in older adults. J Sports Sci. 2018;36(14):1572-7.
116. van Schooten KS, Rispens SM, Elders PJ, Lips P, van Dieen JH, Pijnappels M. Assessing physical activity in older adults: required days of trunk accelerometer measurements for reliable estimation. J Aging Phys Act. 2015;23(1):9-17.
117. Marshall S, Kerr J, Carlson J, Cadmus-Bertram L, Patterson R, Wasilenko K, Crist K, Rosenberg D, Natarajan L. Patterns of Weekday and Weekend Sedentary Behavior Among Older Adults. J Aging Phys Act. 2015;23(4):534-41.
118. Kocherginsky M, Huisingh-Scheetz M, Dale W, Lauderdale DS, Waite L. Measuring Physical Activity with Hip Accelerometry among U.S. Older Adults: How Many Days Are Enough? PLoS One. 2017;12(1):e0170082.
119. Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int J Behav Nutr Phys Act. 2011;8:62.
120. Aguilar-Farias N, Martino-Fuentealba P, Salom-Diaz N, Brown WJ. How many days are enough for measuring weekly activity behaviours with the ActivPAL in adults? J Sci Med Sport. 2019;22(6):684-8.
121. Nuzum H, Stickel A, Corona M, Zeller M, Melrose RJ, Wilkins SS. Potential Benefits of Physical Activity in MCI and Dementia. Behav Neurol. 2020;2020:7807856.
122. Thomas E, Battaglia G, Patti A, Brusa J, Leonardi V, Palma A, Bellafiore M. Physical activity programs for balance and fall prevention in elderly: A systematic review. Medicine (Baltimore). 2019;98(27):e16218.
123. Heissel A, Heinen D, Brokmeier LL, Skarabis N, Kangas M, Vancampfort D, Stubbs B, Firth J, Ward PB, Rosenbaum S, Hallgren M, Schuch F. Exercise as medicine for depressive symptoms? A systematic review and meta-analysis with meta-regression. Br J Sports Med. 2023;57(16):1049-57.
124. Ortlieb S, Gorzelniak L, Nowak D, Strobl R, Grill E, Thorand B, Peters A, Kuhn KA, Karrasch S, Horsch A, Schulz H. Associations between multiple accelerometry-assessed physical activity parameters and selected health outcomes in elderly people--results from the KORA-age study. PLoS One. 2014;9(11):e111206.
125. Dooley EE, Chen L, Ghazi L, Hornikel B, Martinez-Amezcua P, Palta P, Bowling CB, Muntner P, Lewis CE, Pettee Gabriel K. Multimorbidity is associated with lower total 24-hour movement activity among US adults. Prev Med Rep. 2023;36:102483.
126. Maasakkers CM, Claassen J, Scarlett S, Thijssen DHJ, Kenny RA, Feeney J, Melis RJF. Is there a bidirectional association between sedentary behaviour and cognitive decline in older adults? Findings from the Irish Longitudinal Study on Ageing. Prev Med Rep. 2021;23:101423.
127. Nielsen C, Lang RS. Principles of screening. Med Clin North Am. 1999;83(6):1323-37, v.

Appendix

Appendix A. Cross-sectional Timed Up-and-Go

Table A1. 3-month TUG, time window performance

		03:00-06:00	***06:00-09:00***	***09:00-12:00***	***12:00-15:00***	***15:00-18:00***	***18:00-21:00***	***21:00-00:00***	***00:00-03:00***	***15:00-21:00***
All features	*RMSE*	11.54	10.09	9.41	8.82	**7.83**	**7.36**	8.68	12.49	7.63
	R^2	0.07	0.26	0.38	0.4	**0.57**	**0.6**	0.45	-0.15	0.59
Top 20 features*	*RMSE*	10.95	9.7	8.83	7.74	**6.51**	**6.4**	8.22	12.34	6.11
	R^2	0.16	0.32	0.46	0.54	**0.7**	**0.69**	0.49	-0.12	0.74

*Results of best performing time windows shown in bold. TUG, timed up & go; RMSE, root mean square error; R^2, coefficient of determination. * based on mean feature importance of all features.*

Table A2. 3-month TUG, features with highest importance

Feature			**Mean feature importance**
Movement parameter	*Metric*	*Timeslot*	
COUNTS_total	Max	15:00:00	0.030191
COUNTS_total	SD	15:00:00	0.025221
COUNTS_total**	SD	19:00:00	0.017535
METS_moderate_time	Mean	17:00:00	0.01754
METS_moderate_time** †	Mean	18:00:00	0.013177
METS_sedentary_time	Mean	15:00:00	0.018861
METS_sedentary_time** †	Mean	17:00:00	0.012775
METS_sedentary_time	Mean	19:00:00	0.018499
MI_active** †	Mean	18:00:00	0.01552
MI_active	SD	15:00:00	0.022269
MI_moving	Mean	15:00:00	0.015293
MI_moving	Mean	17:00:00	0.053175
MI_moving	Mean	19:00:00	0.024674
MI_moving	Mean	20:00:00	0.02704
MI_worn*	SD	15:00:00	0.015602
PAR_moving	SD	15:00:00	0.024209

*Descriptions of movement parameters can be found in Table 2. SD, standard deviation; max, maximum; TUG, timed up & go. * featured in top-13 features of highest importance in 3/4 repetitions. ** featured in top-13 features of highest importance in 2/4 repetitions. †not included in final feature set.*

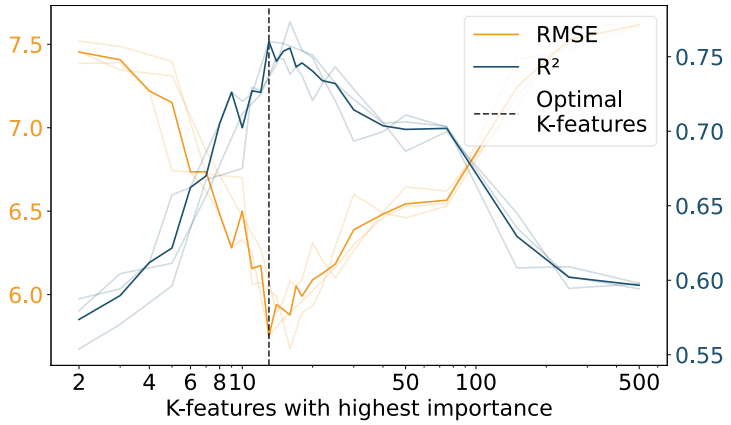


Figure A1. Amount of features with highest performance for cross-sectional TUG. *Highest performance at 13 features. TUG, timed up & go; RMSE, root mean square error; R^2, coefficient of determination*

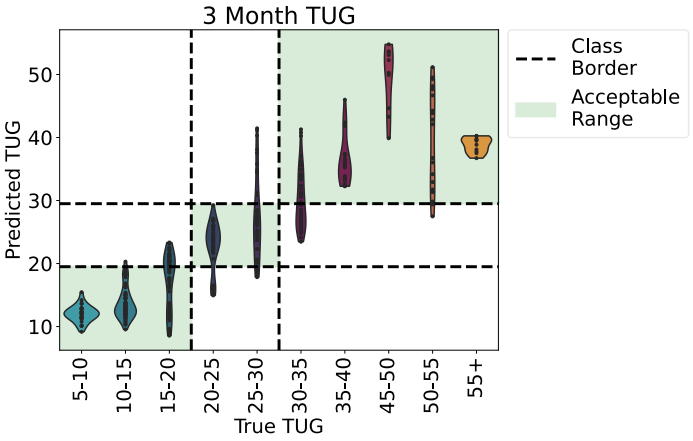


Figure A2a. Distribution of predicted TUG scores per true 5 second range 3-month TUG score.

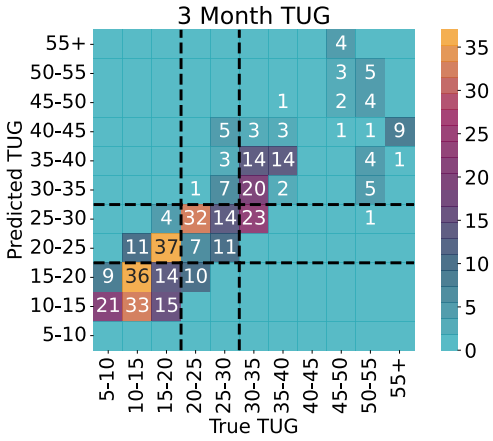


Figure A2b. 5 Second range predicted TUG scores per true 5 second range 3-month TUG score.

Class border and acceptable range are defined based on functional categories as described in Table A3. Five second ranges were chosen to improve interpretability of figures. TUG, timed up & go.

Table A3. Regression and grouped regression predictive performance of 3-month TUG

Functional Category	**Regression**		**Grouped Regression**		
	RMSE	*R^2*	*Accuracy*	*Precision*	*Recall*
Overall	5.95	0.74	73,8%		
0-19 low fall risk [63]				0.93	0.71
20-29 high fall risk [63]				0.46	0.71
30+ dependent [63]				0.86	0.80

TUG, timed up & go; RMSE, root mean square error; R^2, coefficient of determination; PLS, partial least squares.

Appendix B. Interfeature Correlations

Table B1. Interfeature Correlations, 3-month FAC.

	DUR_total_moving_mean_14:00:00	DUR_total_moving_std_14:00:00	MI_active_mean_19:00:00	MI_active_std_13:00:00	MI_active_std_15:00:00	MI_moving_std_15:00:00
DUR_total_moving_mean_14:00:00	1					
DUR_total_moving_std_14:00:00	0.87	1				
MI_active_mean_19:00:00	0.2	0.33	1			
MI_active_std_13:00:00	0.43	0.43	0.54	1		
MI_active_std_15:00:00	0.35	0.33	0.29	0.58	1	
MI_moving_std_15:00:00	0.32	0.28	0.38	0.59	0.51	1

Table B2. Interfeature Correlations, 3-month CIT.

	METS_sedentary_time_mean_06:00:00	PAR_active_max_[14:00:00	PAR_inactive_std_16:00:00	PAR_moving_mean_14:00:00	TRANSITIONS_ly_std_17:00:00
METS_sedentary_time_mean_06:00:00	1				
PAR_active_max_14:00:00	0.16	1			
PAR_inactive_std_16:00:00	0.39	0.36	1		
PAR_moving_mean_14:00:00	0.28	0.74	0.44	1	
TRANSITIONS_ly_std_17:00:00	0.26	0.04	0.62	0.13	1

Table B3. Interfeature Correlations, 12-month CIT.

	DUR_total_sitting_mean_22:00:00	METS_vigorous_time_max_23:00:00	METS_vigorous_time_mean_23:00:00	MI_moving_std_23:00:00	PAR_lying_max_ 22:00:00
DUR_total_sitting_mean_22:00:00	1.0				
METS_vigorous_time_max_23:00:00	0.09	1.0			
METS_vigorous_time_mean_23:00:00	0.13	0.87	1.0		
MI_moving_std_23:00:00	0.22	0.28	0.35	1.0	
PAR_lying_max_ 22:00:00	0.08	0.47	0.32	-0.03	1.0

Table B4. Interfeature Correlations, 12-month FAC.

ft #	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					
9																					
10																					
11																					
12																					
13																					
14																					
15																					
16																					
17	1																				
18	0.17	1																			
19	0.72	0.2	1																		
20	0.6	0.24	0.54	1																	
21	0.68	0.33	0.52	0.76	1																
22	0.62	0.21	0.82	0.76	0.66	1															
23	0.63	0.1	0.88	0.52	0.62	0.79	1														
24	0.53	0.12	0.23	0.7	0.7	0.52	0.28	1													
25	0.55	0.01	0.82	0.31	0.45	0.6	0.91	0.04	1												
26	0.53	0.38	0.53	0.72	0.68	0.64	0.48	0.6	0.27	1											
27	0.61	0.28	0.81	0.57	0.57	0.77	0.7	0.41	0.54	0.84	1										
28	0.39	0.33	0.25	0.7	0.62	0.44	0.31	0.63	0.1	0.88	0.61	1									
29	0.53	0.36	0.35	0.06	0.17	0.18	0.41	0.07	0.44	0.05	0.18	0.12	1								
30	0.24	0.15	0.27	0.05	0.03	0.07	0.32	0.09	0.39	0.11	0.1	0.15	0.61	1							
31	0.37	0.1	0.53	0.5	0.56	0.62	0.69	0.46	0.5	0.73	0.73	0.66	0.18	0.12	1						
32	0.25	0.39	0.1	0.45	0.44	0.19	0.13	0.3	0.01	0.46	0.24	0.61	0.24	-0.2	0.25	1					
33	0.32	0.15	0.09	0.46	0.44	0.21	0.14	0.46	0	0.48	0.27	0.66	0.06	0.21	0.33	0.91	1				
34	0.17	0.32	0.09	0.16	0.34	0.18	0.02	0.24	0.13	0.24	0.21	0.11	0.02	0	0.17	0.07	0.01	1			
35	0.16	0.33	0.09	0.28	0.46	0.27	0.09	0.37	-0.1	0.37	0.28	0.27	0.04	0.06	0.32	0.17	0.1	0.92	1		
36	0.34	0.04	0.14	0.49	0.58	0.42	0.22	0.64	0.02	0.44	0.28	0.41	0.28	0.14	0.4	0.08	0.19	0.47	0.58	1	
37	0.06	0.28	0.12	0.1	0.26	0.22	0.11	0.18	0.06	0.26	0.24	0.13	0.08	0.06	0.22	0.03	0.04	0.76	0.79	0.54	1

Feature	ft #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
COUNTS_total_std_09:00:00	1	1															
DUR_total_active_max_[11:00:00	2	0.32	1														
DUR_total_active_mean_11:00:00	3	0.2	0.86	1													
DUR_total_inactive_mean_11:00:00	4	-0.2	0.86	-1	1												
DUR_total_inactive_std_11:00:00	5	0.27	0.96	0.77	0.77	1											
DUR_total_moving_std_10:00:00	6	0.6	0.5	0.31	0.31	0.53	1										
DUR_total_sitting_max_09:00:00	7	0.16	0.03	0.15	0.15	0	0	1									
DUR_total_sitting_mean_10:00:00	8	0.19	0.41	0.39	0.39	0.41	0.11	0.27	1								
DUR_total_sitting_std_11:00:00	9	0.19	0.75	0.5	-0.5	0.85	0.44	0.22	0.34	1							
DUR_total_static_max_11:00:00	10	0.22	0.94	0.86	0.86	0.87	0.28	0.05	0.52	0.65	1						
DUR_total_static_mean_11:00:00	11	0.08	0.76	0.95	0.95	0.64	0.08	-0.2	0.44	0.35	0.85	1					
DUR_total_static_std_09:00:00	12	0.29	0.56	0.43	0.43	0.6	0.26	0.37	0.32	0.47	0.56	0.4	1				
DUR_total_static_std_11:00:00	13	0.2	0.95	0.79	0.79	0.95	0.35	0.02	-0.5	0.79	0.95	0.74	0.65	1			
METS_moderate_time_max_11:00:00	14	0.37	0.42	0.24	0.24	0.51	0.59	0.01	0.11	0.43	0.24	0.03	0.16	0.36	1		
METS_moderate_time_mean_09:00:00	15	0.53	0.06	0.03	0.03	0.06	0.34	0.19	0.05	0.06	0.04	0.16	0.03	0.02	0.27	1	
METS_moderate_time_std_10:00:00	16	0.34	0.19	0.05	0.05	0.23	0.3	0.05	0.04	0.19	0.14	0.02	0.05	0.18	0.44	0.6	1
MET_mean_max_10:00:00	17	0.53	0.19	0.2	-0.2	0.16	0.65	0.02	0.04	0.08	0.06	0.06	0.12	0.04	0.41	0.49	0.44
MET_mean_mean_09:00:00	18	0.69	0.37	0.26	0.26	0.34	0.42	0.56	0.18	0.32	0.29	0.14	0.54	0.32	0.2	0.49	0.17
MI_active_mean_10:00:00	19	0.4	0.09	0.04	0.04	0.12	0.41	0.02	0.05	0.04	0.01	0.06	0.02	0.02	0.34	0.75	0.75
MI_active_std_10:00:00	20	0.63	0.33	0.13	0.13	0.37	0.7	0.11	0.25	0.33	0.21	0.01	0.22	0.27	0.52	0.34	0.52
MI_active_std_11:00:00	21	0.7	0.45	0.29	0.29	0.47	0.68	0.07	0.26	0.35	0.32	0.12	0.19	0.35	0.68	0.48	0.51
MI_moving_mean_10:00:00	22	0.47	0.28	0.09	0.09	0.34	0.59	0.13	0.17	0.26	0.16	0.05	0.11	0.25	0.5	0.61	0.73
MI_moving_mean_11:00:00	23	0.36	0.08	0.05	0.05	0.12	0.33	0.01	0.03	0.07	0	0.14	0.03	0.06	0.43	0.73	0.7
MI_moving_std_10:00:00	24	0.45	0.43	0.23	0.23	0.48	0.75	0.18	0.23	0.36	0.28	0.06	0.16	0.34	0.59	0.07	0.11
MI_moving_std_11:00:00	25	0.26	0.13	0.16	0.16	0.11	0.08	0.07	0.13	0.09	0.16	0.21	-0.1	0.14	0.24	0.73	0.62
PAR_active_max_10:00:00	26	0.58	0.38	0.16	0.15	0.44	0.73	0	0.17	0.36	0.22	0.02	0.37	0.32	0.62	0.39	0.42
PAR_active_mean_10:00:00	27	0.4	0.3	0.15	0.15	0.36	0.56	0.02	0.03	0.22	0.16	0.01	0.25	0.25	0.52	0.56	0.61
PAR_active_std_10:00:00	28	0.55	0.31	0.06	0.06	0.39	0.68	0.05	0.21	0.34	0.18	0.08	0.45	0.3	0.56	0.15	0.24
PAR_inactive_max_10:00:00	29	0.09	0.09	0.08	0.08	0.12	0.12	0.22	0.02	0.23	0.01	0.17	0.29	0.07	0.09	0.05	0.12
PAR_lying_max_10:00:00	30	0.02	0.13	0.02	0.02	-0.1	0.23	0.01	0.14	0.14	-0.1	0.01	0.11	0.08	0.11	0.2	0.03
PAR_moving_max_11:00:00	31	0.29	0.34	0.09	0.09	0.44	0.43	0.01	0.16	0.37	0.25	0.02	0.35	0.39	0.66	0.33	0.5
PAR_moving_std_09:00:00	32	0.52	0.2	0.05	0.05	0.19	0.41	0.27	0.05	0.15	0.15	0.02	0.2	0.17	0.38	0.16	0.33
PAR_moving_std_10:00:00	33	0.37	0.12	0.01	0.01	0.14	0.45	0.07	0.05	0.06	0.05	0.08	0.08	0.09	0.42	0	0.29
PERIODS_active_mean_11:00:00	34	0.21	0.86	0.92	0.92	0.81	0.38	0.21	0.46	0.6	0.83	0.84	0.38	0.81	0.3	0.12	0.16
PERIODS_inactive_std_11:00:00	35	0.25	0.91	0.8	-0.8	0.9	0.45	0.14	0.49	0.76	0.87	0.7	0.4	0.89	0.41	0.16	0.26
TRANSITIONS_si_max_10:00:00	36	0.24	0.42	0.3	-0.3	0.46	0.48	0.35	-0.3	0.37	0.37	0.21	0.06	0.41	0.37	0.12	0.2
TRANSITIONS_si_mean_11:00:00	37	0.1	0.56	0.47	0.47	0.6	0.35	-0.2	0.39	0.58	0.47	0.35	0.15	0.54	0.26	0.31	0.21

Appendix C. Literature Review (TM30003)

Using Home-Based Accelerometry for Prediction of Future Psychological Wellbeing: A Scoping Review

Nadine L.A. de Jong

Abstract

To keep up with increasing pressure on health care systems, innovative and cost-effective solutions are essential. Personalized rehabilitation care seeks for better indicators of individual wellbeing to efficiently apply early therapeutic interventions. Accelerometry could be a noninvasive and affordable tool to contribute to the prediction of future biopsychosocial wellbeing. This review explores the available literature on studies relating baseline accelerometry-quantified physical activity and subsequent psychological wellbeing.

A PUBMED database search resulted in the inclusion of 35 papers. Accelerometric and psychological outcome measures were extracted and findings were summarized.

This review showed the potential of some accelerometric outcome measures assessing both physical activity intensity and rest-activity rhythms to be an early indicator of future psychological wellbeing in adults. Nevertheless, this is not the case of all frequently described accelerometric outcome measures. Large variations in patient cohorts, accelerometric methods, accelerometric and psychological outcome measures and follow-up duration complicate the comparison of study findings.

Combining multiple accelerometric parameters, including measures of total physical activity, moderate-to-vigorous physical activity, napping, amplitude and relative amplitude, will likely improve prediction of future psychological wellbeing.

List of Abbreviations

AD	Alzheimer's Disease
AD8	Ascertain Dementia Questionnaire
AP	Anteroposterior
C&D	Ccognition and Dementia
CVLT-II	California Verbal Learning Test
D-KEFS	Delis-Kaplan Executive Function System
EE	Energy Expenditure
EEG	Electroencephalography
GDS	Geriatric Depression Scale
ICF	International Classification of Functioning, Disability and Health
IQR	Interquartile Range
IS	Interdaily Stability
IV	Intradaily Variability
L5	Activity during 5 hours of lowest activity
LPA	Light Physical Activity
M10	Activity during 10 hours of highest activity
MCI	Mild Cognitive Impairment
METmin	Metabolic Equivalent of Task Minutes
mg	milligravities

MH&D	Mental Health and Depression
(M)MMS	(Modified) Mini Mental State
MPA	Moderate Physical Activity
MrOS	Osteoporotic Fractures in Men Study
MVPA	Moderate-to-Vigorous Physical Activity
PA	Physical Activity
PAI	Physical Activity Intensity
PCA	Principle Component Analysis
PHQ-9	Patient Health Questionnaire
PLS-DA	Partial Least Squares Discriminant Analysis
RA	Relative Amplitude
RAR	Rest-Activity Rhythms
SB	Sedentary Behavior
SDQ	Strengths and Difficulties Questionnaire
SOF	Study of Osteoporotic Fractures
TPA	Total Physical Activity
UK	United Kingdom
USA	United Stated of America
VPA	Vigorous Physical Activity
WAIS	Wechsler Adult Intelligence Scale

1 Introduction

The pressure on the health care system is rising. With increased longevity, the population of elderly with multimorbidity and chronic conditions is growing. [1, 2] Furthermore, healthcare advancements have led to significant improvements in survival rates for traumatic injuries and acquired brain injuries such as stroke [3-5]. The incidence of stoke in the Netherlands was 2.0 per 1000 inhabitants in 2020, and it substantially increases with age. In 2019, mortality rate within 30 days was 11.4% and 35.5% for respectively ischemic and hemorrhagic stroke. [6] Surviving patients often need rehabilitation till some extend, dealing with both immediate and delayed (in)visible effects [7]. For instance, this may involve primary care physiotherapy, multidisciplinary rehabilitation at an outpatient clinic or temporary admittance to a nursing home.

To minimalize the consequences of especially the delayed and invisible effects of health conditions, personalized, secondary preventive medicine can play an important role [8, 9]. The International Classification of Functioning, Disability, and Health (ICF) model emphasizes biopsychosocial wellbeing as a comprehensive approach to healthcare [10, 11]. Further development of rehabilitation methods using this ICF model seeks for better prediction of individual future wellbeing to efficiently apply early therapeutic interventions. Personalized rehabilitation programs could offer all-round care based on the ICF model whilst preventing unnecessary costs and patient burden. Objective quantification is crucial for early decision-making, and accelerometry emerges as a noninvasive and affordable method to achieve this [12, 13]. While the link between accelerometry and physical fitness is intuitive, its connection to psychological wellbeing may be less evident [14-16].

Growing numbers of studies, as well as multiple reviews, have assessed the association between physical activity and psychological wellbeing using accelerometry. Unfortunately, these studies are primarily cross-sectional and assess a heterogenous range of psychosocial and accelerometric measures. [17-20] This review aims to fill this gap by exploring the available literature on studies relating baseline accelerometry-quantified physical activity and subsequent psychological wellbeing, including cognition and mental health, providing valuable insights for enhancing patient care and

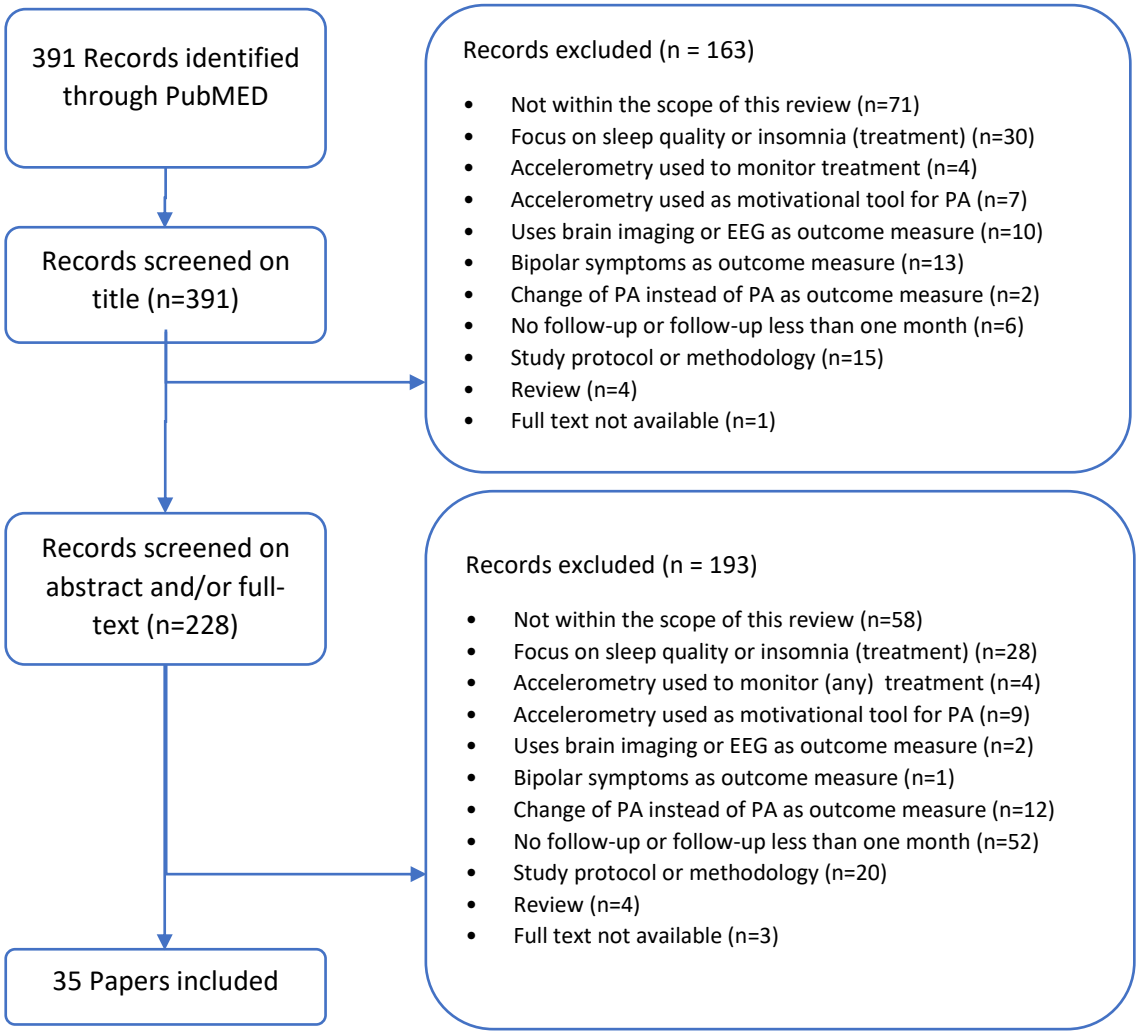
early intervention strategies. Specifically, an overview will be provided of (1) the use of accelerometry as an indicator for future psychological wellbeing, (2) common methods utilized for the analysis of accelerometry data in this context, and (3) the most promising methods for application as predictor for biopsychosocial wellbeing.

2 Methods

2.1 Search Strategy

A literature search was performed to identify prospective studies reporting on the relationship between quantification of home-based physical activity (PA) using accelerometry data and psychological wellbeing. The search was performed in the PubMed electronic database in May 2023 using a combination of keywords aiming at objectively measured physical activity (*Acceleromet* OR actigraph* OR gyroscop* OR movemonitor OR "physical activity monitoring" OR "fitness trackers"*), psychological wellbeing and cognition (*"mental health" OR "clinical course" OR "psychological health" OR cognit* OR biopsychosocial*), and study method (*prospective OR (longitudinal AND follow up)*).

Figure 1. Flowchart of article search and selection.



Abbreviations: EEG, electroencephalography; PA, physical activity

2.2 Exclusion Criteria

Papers were excluded if they were a review, study protocol or if full text was not available. Furthermore, they were excluded if they had no follow-up or had a follow-up time of less than one month. Papers were not eligible for inclusion if they did not use objective physical activity assessment, assessment was laboratory-based and not home-based, used sleep quality, insomnia as outcome measure or used accelerometry as motivational tool for PA. Papers assessing only the change in physical activity were also excluded, as this requires at least two accelerometric measurement periods instead of only one at baseline. Studies were also excluded if they did not include psychological and/or social outcome measures, or used neurophysiological or neuroimaging as outcome measurement. Papers reporting bipolar symptoms were excluded since they are known to be present in episodes. The research flow diagram describing paper selection is shown in figure 1.

2.3 Data Extraction

For each paper, data associated with study design, accelerometric materials and methods and psychological assessment methods was extracted and summarized. This included global patient characteristics, accelerometer type, wearing time and location, both PA and psychological assessment methodologies and time between baseline (PA assessment) and follow-up (psychological state).

3 Results

PubMED search resulted in 391 articles published between 2011 and 2023, which were first screened on title, resulting in exclusion of 163 records (figure 1). Remaining articles were screened on abstract and/or full-text, after which an additional 193 papers were excluded.

3.1 Study Design Characteristics

3.1.1 Population and Sample Size

There were several papers reporting on the same studies. Four papers reported the association with physical activity intensity and psychological measures as a part of the two-year follow-up study targeting community-dwelling people aged ≥65 years in the Hunei District, Kaohsiung, Taiwan [21-24]. Four articles were part of the Osteoporotic Fractures in Men Study (MrOS), which included men aged ≥65 years, or its ancillary study assessing sleep. This United States of America (USA)-based multicenter cohort study recruited community-dwelling men aged ≥65 years. [25-28] Three papers reported on the Rush Memory and Aging Project (USA), with participants mainly being recruited from retirement communities [29-31]. Three included papers described results found in the multicenter Study of Osteoporotic Fractures (SOF), targeting women aged ≥65 years in the USA [32-34]. Two papers were a part of the Women's Health Initiative (USA), including postmenopausal women [28, 35]. Finally, three papers included their samples from the United Kingdom (UK) Biobank, a large prospective study with participants between ages 40-69 years at baseline [36-38]. In total, the 35 included papers described 22 unique studies.

Not all papers describing the same study included the same number of participants due to their specific study criteria. Due to this discrepancy, the following population characteristics are calculated using all papers (including duplicate studies). Included papers had a large range of included sample size, ranging from 58 [39] to 60235 [38], median sample size was 1203. Most papers included only older participants, only 5 studies reported a mean or median age below 25 years old [39-43]. Median age of remaining studies was 74.52 (IQR 65.75-81.47). All papers reported gender of participants. Four papers only included males (MrOS), five papers only included females (SOF, Women's Health Initiative), median percentage female was 54.4% (IQR 51-76).

Table 1. Included papers and main information on method.

1st author (publ. year) [ref nr.]	Study	Psych. Cat.	Acc. Cat.	N	Age (years)	Sex (%women)	Follow-up (years)	Acc Type	Acc Location	Acc Wear days
Ekegren(2021) [44]	-	MH&D	PAI	58	range 19-69	44,8	0.5†	ActiGraph GT3X+ & activPAL3	anterior mid-thigh & right hip	4 - 10†
Chong (2021) [43]	-	MH&D	PAI	88	11.8 (SD 0.4)	59,1	1 (SD 0.17)	GENEActiv	ND wrist	3 - 6†
Yi Lee (2021) [45]	-	C&D	RAR	174	75.6 (SD 7.1)	79,3	1†	GENEActiv	ND wrist	5 - 7†
Slykerman (2020) [41]	Auckland Birthweight Collaborative study (ABC study)	MH&D	PAI	547	7**	51	4†	ActiGraph AM71256	not reported	1†
Sewell (2023) [46]	Australian Imaging, Biomarkers and Lifestyle study (AIBL)	C&D	PAI	199	68.7 (SD 5.9)	55,8	8.4 (SD 2.5)	ActiGraph GT1M	waist	7†
Tian (2021) [47]	Baltimore Longitudinal Study of Aging	C&D	RAR	520	73 (SD 8)	51	7.3 (SD 2.7)	ActiHeart	chest	3 - 7†
Whitaker (2021) [48]	Coronary Artery Risk Development in Young Adults (CARDIA)	C&D	PAI	1970	45.27 (SD 3.56)	58,27	5 & 10†	ActiGraph 7164	right hip	4 - 7†
Hsueh (2021) [24]	Hunei District, Taiwan	MH&D	PAI	274	74.52 (SD 6.12)	54,4	1.84 (SD 0.13)	ActiGraph GT3X+	waist	5 - 7†
Chen (2020) [23]	Hunei District, Taiwan	C&D	PAI	274	74.52 (SD 6.12)	54,4	1.84 (SD 0.12)	ActiGraph GT3X+	waist	5 - 7†
Ku (2017) [22]	Hunei District, Taiwan	C&D	PAI	274	74.5 (SD 6.1)	54,4	1.84 (SD 0.12)	ActiGraph GT3X+	waist	5 - 7†
Stubbs (2017) [21]	Hunei District, Taiwan	C&D	PAI	274	74.52 (SD 6.12)	54,4	1.84 (SD 0.12)	ActiGraph GT3X+	waist	5 - 7†
Maasakkers (2021) [49]	Irish Longitudinal Study on Ageing	C&D	PAI	1276	67.3 (SD 9.0)	53	4†	GENEActiv	wrist	4 - 7†
Jeon (2023) [50]	Korean Brain Aging Study for the Early Diagnosis and Prediction of Alzheimer Disease (KBASE)	C&D	RAR	129	69.3 (SD 7.7)	54,3	2.25 (SD 0.16)	ActiWatch 2	ND wrist	7.52 (SD 1.03)
Xiao (2022) [51]	Osteoporotic Fractures in Men Study (MrOS) /	C&D	RAR	2496	76	0	6.8 (SD 3.7)	SleepWatch-O	ND wrist	4.8 (SD 0.8)

1st author (publ. year) [ref nr.]	Study	Psych. Cat.	Acc. Cat.	N	Age (years)	Sex (%women)	Follow-up (years)	Acc Type	Acc Location	Acc Wear days
	MrOS Sleep									
Leng (2019) [27]	Osteoporotic Fractures in Men Study (MrOS) / MrOS Sleep	C&D	RAR	2751	76.0 (SD 5.3)	0	12†	SleepWatch-O	ND wrist	5.2 (SD 0.9)
Rogers-Soeder (2018) [26]	Osteoporotic Fractures in Men Study (MrOS) / MrOS Sleep	C&D	RAR	2754	76.0 (SD 5.3)	0	3.4 (SD 0.5)	SleepWatch-O	ND wrist	≥3†
Smagula (2015) [25]	Osteoporotic Fractures in Men Study (MrOS) / MrOS Sleep	MH&D	RAR	2124	76.24 (SD 5.48)	0	1.2 (SD 0.32)	SleepWatch-O	ND wrist	≥3†
McNeill (2020) [39]	Preschool Activity, Technology, Health, Adiposity, Behaviour and Cognition study (PATH-ABC)	Both	PAI	185	4.19 (SD 0.64)	39,5	1	ActiGraph GT3X+	right hip	6.8 (SD 1.6)
Zhu (2017) [52]	Reasons for Geographic and Racial Differences in Stroke (REGARDS)	C&D	PAI	6452	69.7 (SD 8.5)	55,3	2.9 (SD 1.1)	Actical	right hip	6.6 (SD 0.8)
de Feijter(2023) [53]	Rotterdam Study	MH&D	RAR	947	61.1 (SD 7.6)	52	6 (IQR = 5.6-6.3)	ActiWatch AW4	ND wrist	7†
Li (2023) [31]	Rush Memory and Aging Project	C&D	RAR	1203	81.42 (SD 7.47)	76,6	≥6†	Actical	ND wrist	10 (SD 1)
Li (2020) [30]	Rush Memory and Aging Project	C&D	RAR	1401	81.8* (IQR 76.3-85.7)	77	15†	Actical	ND wrist	7
Buchman (2012) [29]	Rush Memory and Aging Project	C&D	PAI	716	81.6 (SD 7.12)	76	3.5 (SD 1.54)	Actical	ND wrist	9.3 (SD 1.1)
Cabanas-Sánchez (2021) [54]	Seniors-ENRICA-2	MH&D	PAI	1679	71.40 (SD 4.15)	51,7	2.31 (SD 0.31)	ActiGraph GT9X	ND wrist	4 - 7†
Chan (2022) [55]	StandingTall	MH&D	PAI	322	75.5* (range 72.3-80.1)	62,1	2	McRoberts MoveMonitor	lower back	6* (IQR 1)
Wickel (2019) [40]	Study of Early Child Care and Youth Development	C&D	PAI	559	9**	54	6†	ActiGraph 7164	not reported	4 – 7†

1st author (publ. year) [ref nr.]	Study	Psych. Cat.	Acc. Cat.	N	Age (years)	Sex (%women)	Follow-up (years)	Acc Type	Acc Location	Acc Wear days
	(SECCYD)									
Posner (2021) [32]	Study of Osteoporotic Fractures (SOF)	C&D	RAR	1232	82.6 (SD 3.3)	100	4.9 (SD 0.6)	SleepWatch-O	ND wrist	3.6
Walsh (2014) [34]	Study of Osteoporotic Fractures (SOF)	C&D	RAR	1287	82.81 (SD 3.11)	100	5†	SleepWatch-O	ND wrist	≥3†
Tranah (2011) [33]	Study of Osteoporotic Fractures (SOF)	C&D	RAR	1282	82.69 (SD 3.34)	100	4.9	SleepWatch-O	ND wrist	≥3†
Gianfredi (2022) [56]	The Maastricht Study	MH&D	PAI	5113	60.1 (SD 8.5)	50	5.1*	ActivPal3	anterior mid-thigh	1 - 7†
Opdal (2020) [42]	The Tromsø Study	MH&D	PAI	686	16.25 (SD 0.94)	54,5	2†	ActiGraph GT3X+	dominant hip	4 - 7†
Campbell (2023) [36]	UK Biobank	C&D	PAI	34058	55.46 (SD 7.50)	49,23	8.64 (SD 1.76)	Axivity AX3	wrist	7†
Ho (2022) [37]	UK Biobank	MH&D	PAI	37327	56.41 (SD 7.76)	54,6	6.8 (IQR 6.3-7.4)	Axivity AX3	dominant wrist	6.91*
Kandola (2021) [38]	UK Biobank	MH&D	PAI	60235	55.9 (SD 7.7)	56	2	Axivity AX3	wrist	7†
Nguyen (2023) [35]	Women's Health Initiative	C&D	PAI	1346	81.8 (SD 6.2)	100	4.2 (IQR 2.1 - 6.3)	ActiGraph GT3X+	right hip	1 - 7†
Xiao (2022) [28]	Women's Health Initiative	C&D	RAR	763	83.5*	100	4.5 (SD 2.2)	ActiGraph GT3X+	right hip	7†

Included papers, what study they belong to (if applicable) and main information on method including outcome measure categories, participant characteristics, follow-up time and accelerometric methods.

*Abbreviations: Pysch. Cat. = Psychological Outcome Category; Acc. Cat. = Accerelometric Outcome Category; MH&D = Mental Health & Depression; C&D = Cognition & Dementia; PAI = Physical Activity Intensity; RAR = Rest-Activity Rhythms; N = number of participants; Acc = accelerometer; ND = non-dominant. * median, **age as inclusion criteria, no mean or median given, †as described in method*

3.1.2 Accelerometric Methods

The most commonly used accelerometer brand was ActiGraph (Actigraph Corporation, Pensacola, USA), which was used in 10 studies (Table 1) [21-24, 28, 35, 39-42, 44, 46, 48, 54]. Most studies used accelerometers worn on the (non-dominant) wrist (n = 9) [25-27, 29-34, 36-38, 43, 45, 49, 51, 53, 54] or hip (n = 6) [28, 35, 39, 42, 44, 48, 52]. Other locations included the waist (n = 2) [21-24, 46], anterior mid-thigh (n = 2) [44, 56], lower back (n = 1) [55] and chest (n = 1) [47]. For two studies, placement location of accelerometer was not reported [40, 41]. Accelerometer wear time differed between studies and ranged between a mean of 1 [41] and 10 [31] days (Table 1). Most studies aimed at approximately one week of accelerometric data. Generally, wear time during the day was recorded either exclusively during waking hours (only excluding water activities) or extended to include sleep periods. Follow-up time ranged between 6 months [44] and 15 years [30], with a median follow-up time of 4 years (IQR 1.84-6).

3.1.3 Accelerometric Outcome Measures

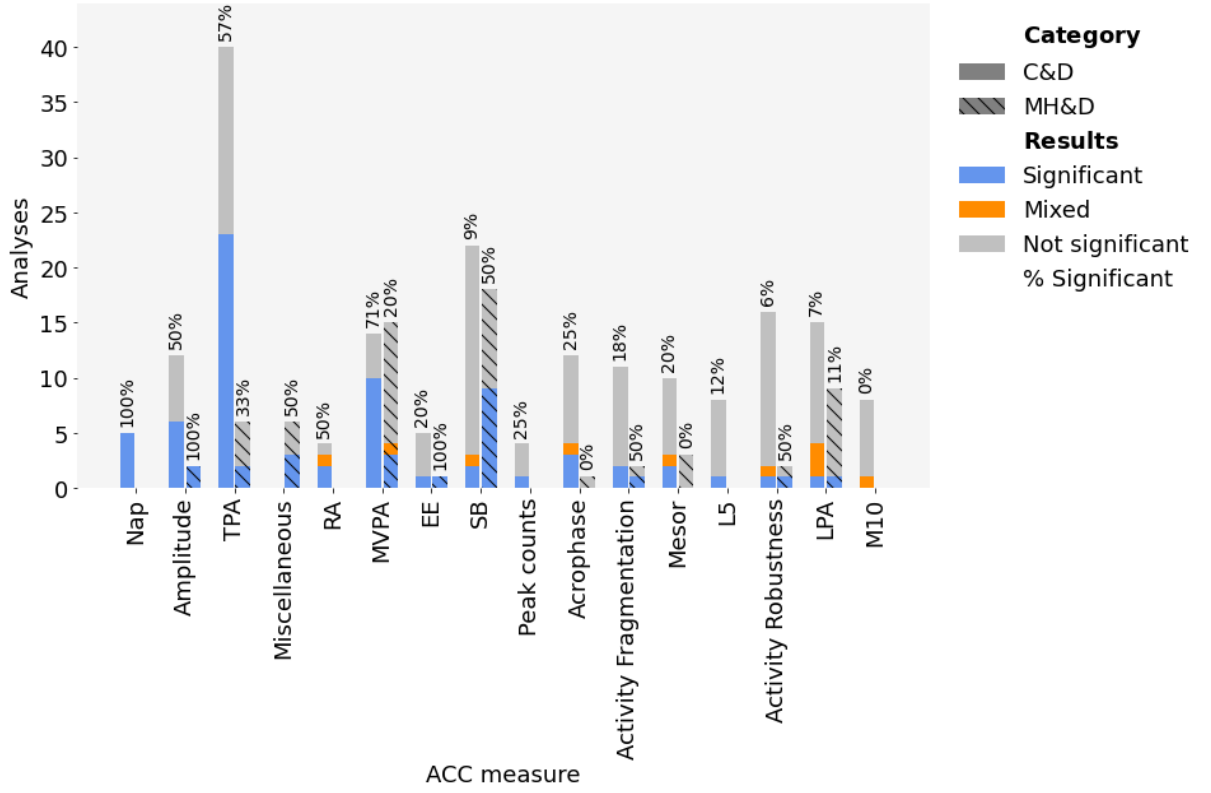
The endpoint parameters used for the analysis of accelerometric data varied significantly among studies (Table S1). Roughly, accelerometric outcome measures could divided into two groups, those mainly assessing physical activity intensity (PAI) or those mainly assessing Circadian- or Rest-Activity Rhythms (RAR). Among the included papers, 22 focused on PAI [21-24, 29, 35-44, 46, 48, 49, 52, 54-56] and 13 focused on RAR (Table 1) [25-28, 30-34, 45, 47, 51, 53]. Commonly used PAI parameters include Sedentary Behavior (SB), Light Physical Activity (LPA), Moderate-to-Vigorous Physical Activity (MVPA), Total Physical Activity (TPA, including step count) and Energy Expenditure (EE) (Figure 2, Table S1). Commonly used RAR parameters include Acrophase, Amplitude, Mesor, Pseudo-F Statistic, L5, M10, Interdaily Stability (IS) and Intradaily Variability (IV) (Figure 2, Table S1).

To further complicate the comparison of results, several studies employed similar outcome measures, but varied in their methodological approaches. Some utilized different measurement techniques, such as counts or milligravities (mg), while others adopted diverse cut-off points for determining the outcomes. For example, for the quantification of daily LPA duration, Chong et al. (2021) determined the amount of minutes the accelerometer measured gravity-corrected vector magnitude units between 52-191 mg, where Ho et al. (2022) used 30-125 mg as cut-off values [37, 43]. For the same outcome measure, multiple studies used counts/minute (with different cut-off values) [21, 35, 39, 40, 48] and finally Gianfredi et al. (2022) determined LPA using standing minutes combined with time spend moving with <100 steps/minute [56].

3.1.4. Psychological Outcome Measures

Similarly to the accelerometric outcome measures, there was also a great variation in the methodology for psychological assessment (Table S1). Studies could also be grouped into two main categories based on their psychological outcome measures, focusing on cognition and/or dementia ('C&D', n = 22) [21-23, 26-36, 40, 45-49, 51, 52], mental health and/or depression ('MH&D', n = 12) [24, 25, 37, 38, 41-44, 53-56] or both (n = 1) [39]. Commonly used assessment tools for cognition and dementia symptoms included the California Verbal Learning Test (CVLT-II), Ascertain Dementia Questionnaire (AD8), D-KEFS (mainly Trail Making subtest), Wechsler Adult Intelligence Scale (WAIS) subtests and the (Modified) Mini Mental State ((M)MMS). Common questionnaires for the assessment of depression and mental health were the Strengths and Difficulties Questionnaire (SDQ), Patient Health Questionnaire (PHQ-9) and the Geriatric Depression Scale (GDS).

Figure 2. Performed accelerometric analyses, categorized.



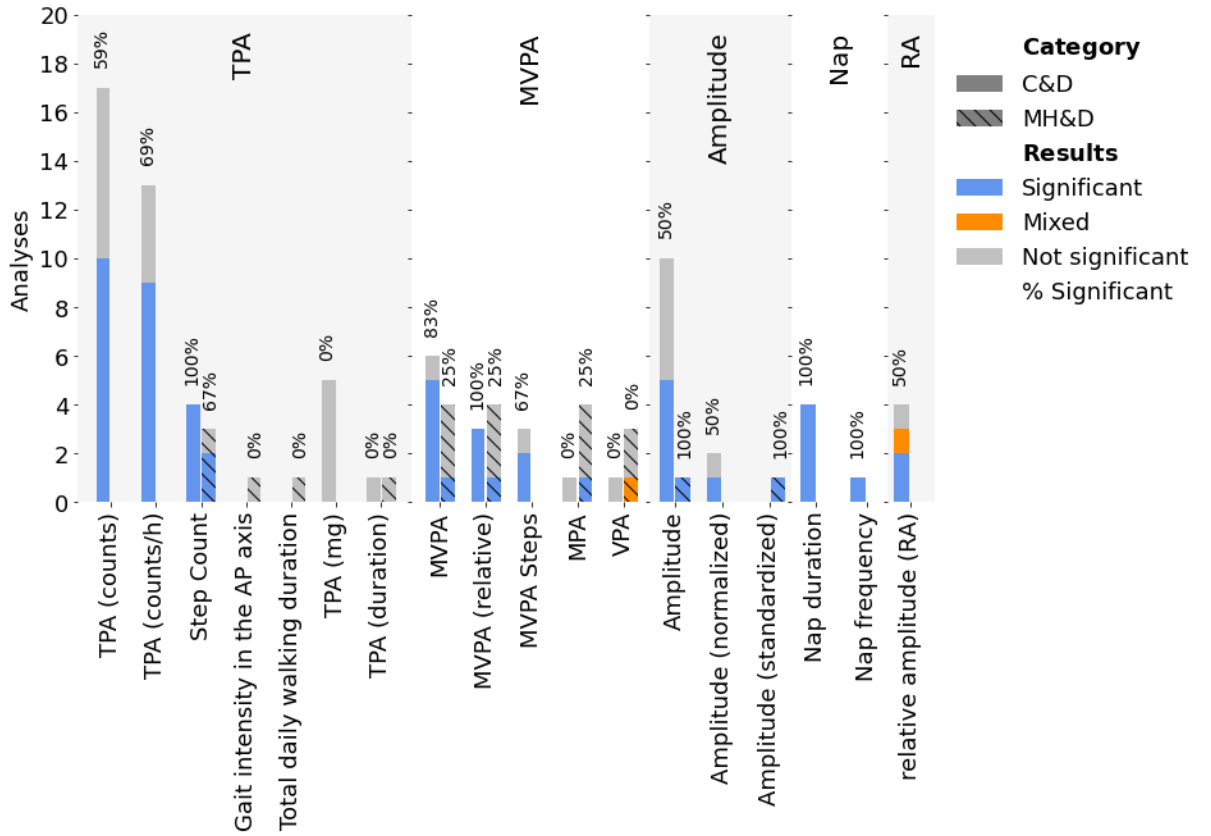
Number of analyses desribed in included papers and corresponding significance of associations with respectively C&D or MH&D psychological outcome measures. Value above each bar indicates percentage of significant findings for the respective analyses. *Abbreviations: C&D, cognition and dementia; MH&D, mental health and depression; EE, energy expenditure; TPA, total physical activity; RA, relative amplitude; MVPA, moderate-to-vigorous physical activity; SB, sedentary behavior; L5, activity during 5 hours of lowest activity; LPA, light physical activity; M10, activity during 10 hours of highest activity.*

3.2 Key Findings

Studies often included multiple analyses assessing the association between several accelerometric measures and cognitive and/or mental health outcomes. The results across these individual analyses are presented in Figure 2. Note that one study can be described in multiple papers, of which each describes the relationships between multiple accelerometric and multiple psychological outcome measures. In some cases, results were categorized as 'mixed', in that case there was no clear conclusion to be drawn from the results, e.g. there was an association for only a specific patient subgroup.

Based on the results described in Figure 2, assessment methods that had a relatively high rate of significant findings were napping, total physical activity, moderate-to-vigorous physical activity, amplitude of rest-activity rhythm and relative amplitude (RA).

Figure 3. Performed accelerometric analyses, subcategories.



Number of analyses desribed in included papers and corresponding significance of associations with respectively C&D or MH&D psychological outcome measures shown in separate columns. Only subcategories of TPA, MVPA, Amplitude, Nap and RA are shown. Value above each bar indicates percentage of significant findings for the respective analyses. *Abbreviations: C&D, cognition and dementia; MH&D, mental health and depression TPA, total physical activity; MVPA, moderate-to-vigorous physical activity; RA, relative amplitude; h, hour; AP, anteroposterior; MPA, moderate physical activity; VPA, vigorous physical activity.*

3.2.1 Total physical activity (TPA)

TPA can be determined using multiple methods, including total daily step count, total daily accelerometric counts and daily duration of physical activity (Figure 3, Table S1).

Various studies have found significant associations between accelerometer quantified overall activity and cognitive health [23, 29, 35, 46]. They found that higher levels of total daily physical activity (TPA) were associated with a reduced risk of developing Alzheimer's disease (AD), better cognition and slower cognitive decline. [29, 46] Buchman et al. (2012) found that participants with lowest daily activity had a 2.3 times higher risk to develop AD compared to those with the highest daily activity. [29] Positive associations have also been observed between TPA and aspects of cognition such as memory and processing speed [29, 46]. Similarly, higher daily step counts have been linked to lower rates of subjective cognitive decline and lower risks of incident mild cognitive impairment (MCI) and/or dementia [23, 35]. Nguyen et al. (2023) reported that participants in the highest step count quartile had a 63% lower risk of MCI/probable dementia compared to those with lowest daily step count [35]. Regarding mental health, engaging in higher levels of overall physical activity, quantified as either TPA or daily steps, appears to be correlated with a lower risk of depressive disorders and reduced depressive symptoms [24, 37, 55].

Nevertheless, not all studies have consistently demonstrated significant correlations between overall activity and cognitive or mental health outcomes [29, 36, 39, 44, 46]. Campbell et al. (2023) found correlations with neither global cognition nor any specific cognitive function [36]. Multiple studies that did report a relationship between global cognition and TPA, also found that certain specific cognitive functions (e.g. executive function or memory) did not (consistently) show significant associations [29, 39, 46]. Both studies reporting nonsignificant findings with mental health outcomes studied patient populations that diverged from the average study population, specifically young children or adults after upper/lower limb fracture [39, 44].

3.2.2 MVPA

Several studies have investigated the relationship between moderate and/or vigorous physical activity (MVPA) and cognitive and mental health outcomes. Included studies reported multiple methods for the quantification of MVPA, including time spend in MVPA, MPA and VPA based on accelerometric counts, time spend in MVPA relative to other behaviors and amount of steps taken during periods of MVPA (Figure 3, Table S1).

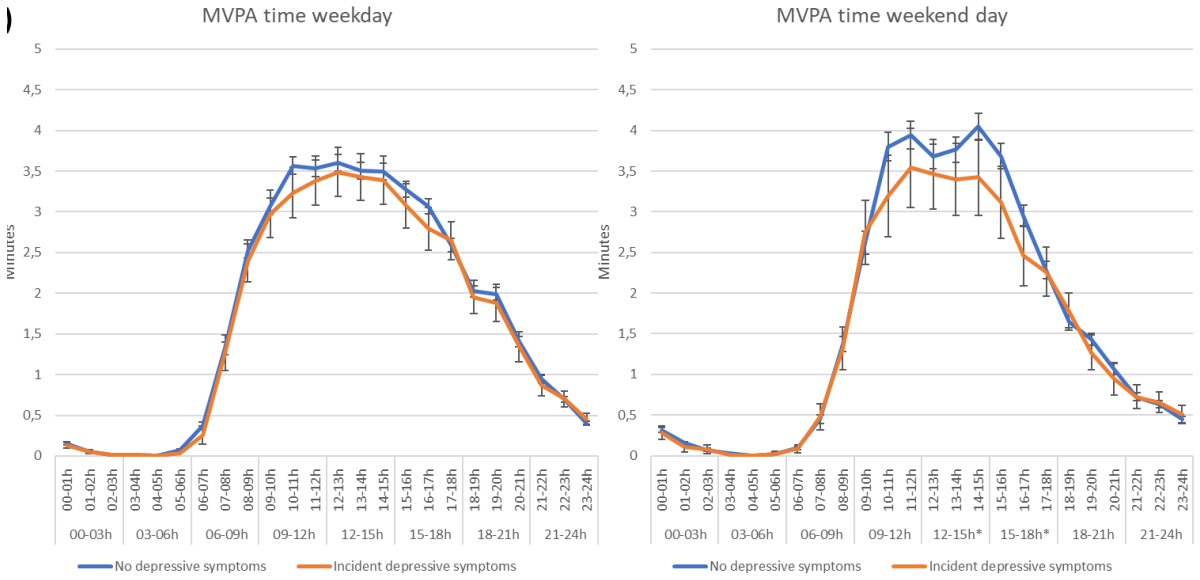
Some studies reported positive associations between MVPA and cognitive health, including a reduced rate of cognitive decline, lower risk of incident MCI and memory and executive function [21, 35, 52]. Nguyen et al. (2023) found that participants with highest MVPA had a 31% lower risk of MCI/probable dementia compared to those with lowest MVPA [35]. Zhu et al. also found that minor differences in MVPA between those least and most active resulted in similar sized effects, with a >35% difference [52]. Controversially, Wickel et al. (2018) found that in adolescents, those with higher MVPA duration had lower cognitive scores compared to those with lower MVPA [40]. Similarly, there is evidence suggesting that engaging in MVPA and moderate physical activity (MPA), is associated with lower risk and improved outcomes for global mental health, depressive symptoms and affective disorders including depression and anxiety [37, 54, 56]. Gianfredi et al. (2022) assessed the patterns of MVPA over time, concluding that the difference in MVPA between those with and without depressive symptoms was predominantly seen on weekend day afternoons (12-15h and 15-18h; Figure 4) [56].

However, some studies did not observe significant associations between MVPA and mental health measures [37, 39, 41, 43, 44, 54]. Similarly to TPA outcomes, studies including children reported no significant findings concerning the association between MVPA and emotional and behavioral problems, psychological distress, psychosocial development and executive function [39, 41, 43]. Furthermore, there were three studies reporting no or limited association between (M)VPA and depression, of which one included adults after upper/lower limb fracture [37, 44, 54].

3.2.3 Napping

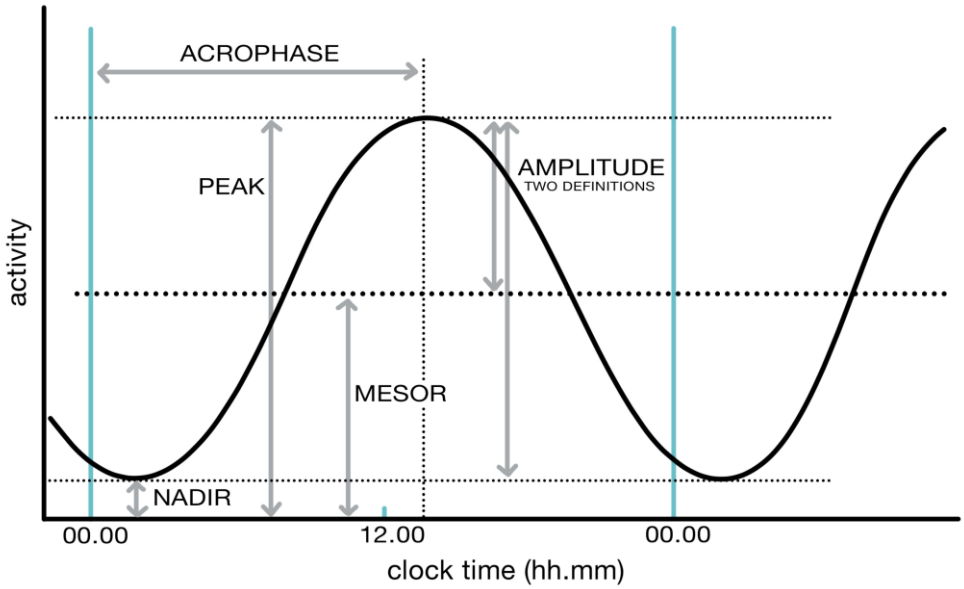
Generalized sedentary behavior was only significantly associated with psychological outcome measures in a small percentage of performed analyses. An alternative, less popular, outcome measure assessing non-active behavior is napping. Napping, also known as daytime sleeping outside of the main sleeping period, was used as an outcome measure in two research papers which both demonstrated significant associations. Both studies assessed the duration of naps, while one of them also examined the frequency of napping. Their findings revealed that longer napping durations were associated with an increased risk of developing cognitive impairment or incident Alzheimer's disease (AD). [27, 31] Additionally, Li et al. (2023) showed that a higher frequency of napping was also linked to a heightened risk of developing incident AD [31].

Figure 4. Hourly distribution of moderate-to-vigorous physical activity.



Hourly distribution of moderate-to-vigorous physical activity (MVPA) stratified by week and weekend days in individuals with (orange) and without (blue) incident depressive symptoms. Figure adapted from Gianfredi et al. (2022). Statistically significant differences in time slots are reported with *. [56]

Figure 5. Rest-activity rhythm (RAR) parameters.



3.2.4 Amplitude

As part of circadian rhythm analysis, the amplitude represents the difference between highest and either lowest or mean activity. It is calculated by either subtracting lowest activity level (nadir) or mean activity level (mesor) from the highest activity level (peak) of that day (Figure 5). It's value depends on which definition is used, as well as on the model used to describe the rhythm (e.g. an extended cosine model). A larger amplitude indicates a more pronounced difference between the two, suggesting a well-defined pattern over a 24-hour period. The correlation between (normalized) amplitude and cognition was examined in several studies. [26, 28, 30, 33, 34, 45] Two studies found

a significant association between lower amplitudes and detrimental effects on risk of dementia, MCI or cognitive decline [26, 28], against two studies that did not [34, 45]. Furthermore, Tranah et al. (2011) did find significant results for those with MCI, but not for incident dementia alone, with it being the other way around for Li et al. (2020) [30, 33]. Concerning mental health, Smagula (2015) found that after adjustments for some covariates, low (standardized) amplitude was associated with higher odds ratios for the development of depressive symptoms, although further adjustments attenuated these results into non-significance [25].

3.2.5 Relative Amplitude (RA)

Relative Amplitude is a non-parametric measure similar to (normalized) amplitude that does not use assumptions about the model of the rhythm. It can be calculated by taking the difference between the activity during the 10 consecutive hours of highest activity (M10) and the 5 consecutive hours of lowest activity (L5) and dividing that by the sum of M10 and L5. A higher value indicates relatively higher activity during waking hours and lower activity when resting/sleeping. Three studies demonstrated a significant relationship between lower relative amplitude and incident dementia [28, 32] and/or cognitive impairment [28, 51], although one study lost significance after further adjustment for sleep and daily steps [28]. Even though Posner et al. (2021) did find an association between relative amplitude and incident dementia, they did not when assessing incident MCI [32].

3.3 Additional relevant findings

3.3.1 Bidirectionality of associations

Multiple studies attempted to address possible reverse causality bias by conducting sensitivity analyses that excluded participants with confounding pre-existing conditions at baseline, such as cognitive impairment, history of depression, or difficulty with activities of daily living [21, 23, 24, 28, 29, 35, 37, 38, 51]. Some studies went further by excluding participants who developed these conditions within the first two years after baseline accelerometric assessment to further minimize reverse causation [27-29, 35, 37, 51].

A few studies directly explored the possibility of a bidirectional relationship [29-31, 40, 49, 53]. Li et al. (2023) demonstrated that nap duration or frequency was correlated to cognitive performance one year later, but also that cognitive performance was correlated to subsequent napping behavior [31]. In a previously published paper studying the same cohort Li et al. also described a bidirectional relationship between AD and circadian dysregulation [30]. This is not only the case for cognition related outcome measures, De Feijter et al. (2023) demonstrated a bidirectional association between 24h activity fragmentation and depressive symptoms [53]. However, not all studies found a consistent bidirectional relationship [29, 40, 49]. For instance, Buchman et al. (2012) found that rate of cognitive decline before accelerometric assessment was not associated with TPA and that baseline global cognition was not associated with rate of decline of TPA [29].

3.3.2 Compensatory strategies

Tian et al. (2021) investigated the association between activity fragmentation and cognitive change, which depended on the level of gait speed. Slow walkers with less fragmentation showed cognitive decline over time, whereas slow walkers with more fragmentation remained stable. The presence of more fragmentation in slow walkers indicated the ability to apply compensation strategies (more frequent rests), which served as an indicator of cognitive function. [47]

4 Discussion

This review showed the potential of some accelerometric outcome measures assessing both physical activity intensity and rest-activity rhythms to be an early indicator of future psychological wellbeing

in adults. Nevertheless, this is not the case of all frequently described accelerometric outcome measures. Large variations in patient cohorts, accelerometric methods, accelerometric and psychological outcome measures and follow-up duration complicate the comparison of study findings. Although some accelerometric outcome measurements were seldomly or never found to be significantly associated with subsequent psychological wellbeing, for most parameters findings differed between studies.

So far, few papers assessed a combination of both PAI and RAR parameters. Based on our findings, combining these different approaches to the quantification of physical activity could be a valuable approach. Chan et al. (2022) extracted many parameters from accelerometric data and performed a principal components analysis (PCA) to reduce that number. Then, they determined which initial variables contributed the most to the first five principal components to determine what aspect of gait was mainly assessed. [55] PCA aims to preserve as much variance in the original data as possible [57]. Methods like these may enable us to identify and combine accelerometric outcome parameters that have the highest possibility of containing relevant patterns. This is especially valuable in situations where parameters exhibit significant covariance, which is the case for PAI outcome measures, where an increase of a certain behavior always leads to a decrease of another behavior [54]. It can also be especially relevant if combinations of multiple parameters indicate a certain level of wellbeing, for example as described in the compensation theory of Tian et al. (2021), see compensatory strategies above [47]. An alternative approach could be Partial Least Squares Discriminant Analysis (PLS-DA), which aims to preserve as much covariance as possible between the accelerometric data and psychological state at follow-up [57]. This may help to determine which (combination of) accelerometric outcome measures at baseline predict a specific psychological outcome at follow-up. Based on our findings, when performing PCA or PLS-DA on accelerometric data, it is essential to include both PAI and RAR features. For both approaches the potential to predict future psychological wellbeing has been demonstrated, but they have not yet been combined. Assessing both PAI and RAR features will likely elevate the predictive value. This assumption is strengthened by the findings regarding the PA behavior changes during the day as described by Gianfredi et al. (2022) [56]. At a minimum, the selected parameters should include measures of TPA, MVPA, napping, amplitude, and relative amplitude.

There were some limitations to this review. First of all, assessment of reference lists of included papers revealed that the used search strategy was not exhaustive. Brief exploration of the first three additionally found papers did not reveal any findings that were not in line with included papers and therefore the search strategy was not expanded further [58-60]. It is possible that results presented in Figure 2 and 3 are not fully representable for all published papers on the topic. Second, this review did not assess cross-sectional studies, studies assessing sleep parameters, studies reporting circadian rhythm interventions or physical activity interventions or those with other interventions that were monitored using PA accelerometry, studies with laboratory based set-ups or studies that only assessed change of PA over time as accelerometric outcome measure. Some studies were also excluded because they did not specifically describe cognitive or mental health related outcome measures. Although they do not shed any insight into the possible predictive value of accelerometry or are outside the scope of this study, investigation of these studies could broaden the insights into possible accelerometric parameters that have not yet been studied in this context and provide additional methods to predict future psychologic wellbeing using accelerometry. Furthermore it is important to realize the effect of publication bias on our results. Nonsignificant findings are less likely to be published than significant associations [61]. Fortunately, many of the included papers assess basic combinations of either PAI or RAR (e.g. the RAR combination of amplitude, acrophase, mesor and pseudo-f statistic) and report both significant and nonsignificant findings. Additionally,

we did not assess the possible effect of underpowered samples in included studies, possibly leading to a faulty assumption that non-significant findings demonstrate an absent effect (false-negatives) [62]. Moreover, multiple papers have reported on the same study cohorts, which can also lead to bias in our results. We did not further investigate the possible effect of confounders between studies such as participant age, sex and comorbidities, as well as accelerometric wear duration and location, and follow-up time. Finally, we did not assess the limitations and drawbacks of the use of (solely) accelerometry in the home environment, such as faulty use, discomfort and the lack of insight into the participants social support system.

Further development of personalized, secondary preventive rehabilitation medicine requires improved prediction of individual future wellbeing to tailor early therapeutic interventions such as occupational therapy to individual needs. Unfortunately, only one of included papers studied patient recovery from a health condition, in this case lower or upper extremity fracture [44]. Based on the findings of this review, some accelerometric outcome measures show promise in predicting future psychological wellbeing in healthy adults. Combining multiple outcome measures, including both RAR and PAI parameters may be most informative on future psychological wellbeing. However, it is essential to note that this predictive capacity may not necessarily hold when being assessed after illness. Future research should study the association between accelerometric outcome measures and subsequent psychological wellbeing as part of rehabilitation after illness.

References

1. Health and care workforce in Europe: time to act. World Health Organization. Regional Office for Europe.; 2022. Available from: https://apps.who.int/iris/handle/10665/362379.
2. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. Nature. 2018;561(7721):45-56.
3. Goonewardene SS, Baloch K, Sargeant I. Road Traffic Collisions-Case Fatality Rate, Crash Injury Rate, and Number of Motor Vehicles: Time Trends Between a Developed and Developing Country. Am Surgeon. 2010;76(9):977-81.
4. Lakshminarayan K, Berger AK, Fuller CC, Jacobs DR, Jr., Anderson DC, Steffen LM, Sillah A, Luepker RV. Trends in 10-year survival of patients with stroke hospitalized between 1980 and 2000: the Minnesota stroke survey. Stroke. 2014;45(9):2575-81.
5. Wafa HA, Wolfe CDA, Emmett E, Roth GA, Johnson CO, Wang Y. Burden of Stroke in Europe: Thirty-Year Projections of Incidence, Prevalence, Deaths, and Disability-Adjusted Life Years. Stroke. 2020;51(8):2418-27.
6. Dolmans LS HL, Tjon-A-Tsien MRS, Van den Donk M, Van Noortwijk-Bonga HGC, Verburg-Oorthuizen AFE, Wiersma Tj. Beroerte. NHG-Richtlijnen: Nederlands Huisartsen Genootschap; 2022. Available from: https://richtlijnen.nhg.org/standaarden/beroerte.
7. Corr S, Wilmer S. Returning to Work after a Stroke: An Important but Neglected Area. British Journal of Occupational Therapy. 2016;66(5):186-92.
8. Liu-Ambrose T, Davis JC, Best JR, Dian L, Madden K, Cook W, Hsu CL, Khan KM. Effect of a Home-Based Exercise Program on Subsequent Falls Among Community-Dwelling High-Risk Older Adults After a Fall: A Randomized Clinical Trial. JAMA. 2019;321(21):2092-100.
9. Chohan SA, Venkatesh PK, How CH. Long-term complications of stroke and secondary prevention: an overview for primary care physicians. Singapore Med J. 2019;60(12):616-20.
10. Wade DT, Halligan PW. The biopsychosocial model of illness: a model whose time has come. Clin Rehabil. 2017;31(8):995-1004.
11. Stokes EK. International classification of functioning, disability and health (ICF). Rehabilitation Outcome Measures2011. p. 13-6.
12. Gebruers N, Truijen S, Engelborghs S, De Deyn PP. Predictive value of upper-limb accelerometry in acute stroke with hemiparesis. J Rehabil Res Dev. 2013;50(8):1099-106.
13. Bento T, Cortinhas A, Leitao JC, Mota MP. Use of accelerometry to measure physical activity in adults and the elderly. Rev Saude Publica. 2012;46(3):561-70.

14. Peluso MA, Guerra de Andrade LH. Physical activity and mental health: the association between exercise and mood. Clinics (Sao Paulo). 2005;60(1):61-70.
15. Rice J, Corp DT, Swarowsky A, Cahalin LP, Cabral DF, Nunez C, Koch S, Rundek T, Gomes-Osman J. Greater Cognitive-Motor Interference in Individuals Post-Stroke During More Complex Motor Tasks. J Neurol Phys Ther. 2022;46(1):26-33.
16. Al-Yahya E, Dawes H, Smith L, Dennis A, Howells K, Cockburn J. Cognitive motor interference while walking: a systematic review and meta-analysis. Neurosci Biobehav Rev. 2011;35(3):715-28.
17. Vancampfort D, Firth J, Schuch FB, Rosenbaum S, Mugisha J, Hallgren M, Probst M, Ward PB, Gaughran F, De Hert M, Carvalho AF, Stubbs B. Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta-analysis. World Psychiatry. 2017;16(3):308-15.
18. Dutil C, Podinic I, Sadler CM, da Costa BG, Janssen I, Ross-White A, Saunders TJ, Tomasone JR, Chaput JP. Sleep timing and health indicators in children and adolescents: a systematic review. Health Promot Chronic Dis Prev Can. 2022;42(4):150-69.
19. Schuch F, Vancampfort D, Firth J, Rosenbaum S, Ward P, Reichert T, Bagatini NC, Bgeginski R, Stubbs B. Physical activity and sedentary behavior in people with major depressive disorder: A systematic review and meta-analysis. J Affect Disord. 2017;210:139-50.
20. Domingos C, Pego JM, Santos NC. Effects of physical activity on brain function and structure in older adults: A systematic review. Behav Brain Res. 2021;402:113061.
21. Stubbs B, Chen LJ, Chang CY, Sun WJ, Ku PW. Accelerometer-assessed light physical activity is protective of future cognitive ability: A longitudinal study among community dwelling older adults. Exp Gerontol. 2017;91:104-9.
22. Ku PW, Liu YT, Lo MK, Chen LJ, Stubbs B. Higher levels of objectively measured sedentary behavior is associated with worse cognitive ability: Two-year follow-up study in community-dwelling older adults. Exp Gerontol. 2017;99:110-4.
23. Chen ST, Stevinson C, Tian T, Chen LJ, Ku PW. Accelerometer-measured daily steps and subjective cognitive ability in older adults: A two-year follow-up study. Exp Gerontol. 2020;133:110874.
24. Hsueh MC, Stubbs B, Lai YJ, Sun CK, Chen LJ, Ku PW. A dose response relationship between accelerometer assessed daily steps and depressive symptoms in older adults: a two-year cohort study. Age Ageing. 2021;50(2):519-26.
25. Smagula SF, Ancoli-Israel S, Blackwell T, Boudreau R, Stefanick ML, Paudel ML, Stone KL, Cauley JA, Osteoporotic Fractures in Men Research G. Circadian rest-activity rhythms predict future increases in depressive symptoms among community-dwelling older men. Am J Geriatr Psychiatry. 2015;23(5):495-505.
26. Rogers-Soeder TS, Blackwell T, Yaffe K, Ancoli-Israel S, Redline S, Cauley JA, Ensrud KE, Paudel M, Barrett-Connor E, LeBlanc E, Stone K, Lane NE, Tranah G, Osteoporotic Fractures in Men Study Research G. Rest-Activity Rhythms and Cognitive Decline in Older Men: The Osteoporotic Fractures in Men Sleep Study. J Am Geriatr Soc. 2018;66(11):2136-43.
27. Leng Y, Redline S, Stone KL, Ancoli-Israel S, Yaffe K. Objective napping, cognitive decline, and risk of cognitive impairment in older men. Alzheimers Dement. 2019;15(8):1039-47.
28. Xiao Q, Shadyab AH, Rapp SR, Stone KL, Yaffe K, Sampson JN, Chen JC, Hayden KM, Henderson VW, LaCroix AZ. Rest-activity rhythms and cognitive impairment and dementia in older women: Results from the Women's Health Initiative. J Am Geriatr Soc. 2022;70(10):2925-37.
29. Buchman AS, Boyle PA, Yu L, Shah RC, Wilson RS, Bennett DA. Total daily physical activity and the risk of AD and cognitive decline in older adults. Neurology. 2012;78(17):1323-9.
30. Li P, Gao L, Gaba A, Yu L, Cui L, Fan W, Lim ASP, Bennett DA, Buchman AS, Hu K. Circadian disturbances in Alzheimer's disease progression: a prospective observational cohort study of community-based older adults. Lancet Healthy Longev. 2020;1(3):e96-e105.
31. Li P, Gao L, Yu L, Zheng X, Ulsa MC, Yang HW, Gaba A, Yaffe K, Bennett DA, Buchman AS, Hu K, Leng Y. Daytime napping and Alzheimer's dementia: A potential bidirectional relationship. Alzheimers Dement. 2023;19(1):158-68.
32. Posner AB, Tranah GJ, Blackwell T, Yaffe K, Ancoli-Israel S, Redline S, Leng Y, Zeitzer JM, Chen DM, Webber KR, Stone KL. Predicting incident dementia and mild cognitive impairment in older women with nonparametric analysis of circadian activity rhythms in the Study of Osteoporotic Fractures. Sleep. 2021;44(10).

33. Tranah GJ, Blackwell T, Stone KL, Ancoli-Israel S, Paudel ML, Ensrud KE, Cauley JA, Redline S, Hillier TA, Cummings SR, Yaffe K, Group SOFR. Circadian activity rhythms and risk of incident dementia and mild cognitive impairment in older women. Ann Neurol. 2011;70(5):722-32.
34. Walsh CM, Blackwell T, Tranah GJ, Stone KL, Ancoli-Israel S, Redline S, Paudel M, Kramer JH, Yaffe K. Weaker circadian activity rhythms are associated with poorer executive function in older women. Sleep. 2014;37(12):2009-16.
35. Nguyen S, LaCroix AZ, Hayden KM, Di C, Palta P, Stefanick ML, Manson JE, Rapp SR, LaMonte MJ, Bellettiere J. Accelerometer-measured physical activity and sitting with incident mild cognitive impairment or probable dementia among older women. Alzheimers Dement. 2023.
36. Campbell T, Cullen B. Estimating the effect of physical activity on cognitive function within the UK Biobank cohort. Int J Epidemiol. 2023.
37. Ho FK, Petermann-Rocha F, Parra-Soto S, Boonpor J, Gill JMR, Gray SR, Pell JP, Celis-Morales C. Device-measured physical activity and incident affective disorders. BMC Med. 2022;20(1):290.
38. Kandola AA, Del Pozo Cruz B, Osborn DPJ, Stubbs B, Choi KW, Hayes JF. Impact of replacing sedentary behaviour with other movement behaviours on depression and anxiety symptoms: a prospective cohort study in the UK Biobank. BMC Med. 2021;19(1):133.
39. McNeill J, Howard SJ, Vella SA, Cliff DP. Longitudinal associations of physical activity and modified organized sport participation with executive function and psychosocial health in preschoolers. J Sports Sci. 2020;38(24):2858-65.
40. Wickel EE, Howie EK. Prospective bi-directional associations between sedentary time and physical activity with cognitive performance: a cohort study. J Sports Sci. 2019;37(6):630-7.
41. Slykerman RF, Budd C, Thompson JMD, Bekker M, Buckley J, Wall C, Murphy R, Mitchell EA, Waldie KE. Physical Activity, Sleep, Body Mass Index, and Associated Risk of Behavioral and Emotional Problems in Childhood. J Dev Behav Pediatr. 2020;41(3):187-94.
42. Opdal IM, Morseth B, Handegard BH, Lillevoll KR, Nilsen W, Nielsen C, Furberg AS, Rosenbaum S, Rognmo K. Is change in mental distress among adolescents predicted by sedentary behaviour or screen time? Results from the longitudinal population study The Tromso Study: Fit Futures. BMJ Open. 2020;10(2):e035549.
43. Chong KH, Parrish AM, Cliff DP, Dumuid D, Okely AD. Cross-Sectional and Longitudinal Associations between 24-Hour Movement Behaviours, Recreational Screen Use and Psychosocial Health Outcomes in Children: A Compositional Data Analysis Approach. Int J Environ Res Public Health. 2021;18(11).
44. Ekegren CL, Ashe MC, Gabbe BJ. Association between Patient-Reported Health Status and Physical Activity Six Months after Upper and Lower Limb Fractures in Working-Aged Adults. PM R. 2021;13(4):353-63.
45. Yi Lee PM, Ling Kwok BH, Ting Ma JY, Tse LA. A population-based prospective study on rest-activity rhythm and mild cognitive impairment among Hong Kong healthy community-dwelling older adults. Neurobiol Sleep Circadian Rhythms. 2021;10:100065.
46. Sewell KR, Rainey-Smith SR, Peiffer J, Sohrabi HR, Taddei K, Ames D, Maruff P, Masters CL, Rowe CC, Martins RN, Erickson KI, Brown BM. The relationship between objective physical activity and change in cognitive function. Alzheimers Dement. 2023.
47. Tian Q, Studenski SA, An Y, Kuo PL, Schrack JA, Wanigatunga AA, Simonsick EM, Resnick SM, Ferrucci L. Association of Combined Slow Gait and Low Activity Fragmentation With Later Onset of Cognitive Impairment. JAMA Netw Open. 2021;4(11):e2135168.
48. Whitaker KM, Zhang D, Pettee Gabriel K, Ahrens M, Sternfeld B, Sidney S, Jacobs DR, Jr., Palta P, Yaffe K. Longitudinal Associations of Midlife Accelerometer Determined Sedentary Behavior and Physical Activity With Cognitive Function: The CARDIA Study. J Am Heart Assoc. 2021;10(3):e018350.
49. Maasakkers CM, Claassen J, Scarlett S, Thijssen DHJ, Kenny RA, Feeney J, Melis RJF. Is there a bidirectional association between sedentary behaviour and cognitive decline in older adults? Findings from the Irish Longitudinal Study on Ageing. Prev Med Rep. 2021;23:101423.
50. Jeon SY, Byun MS, Yi D, Jung G, Lee JY, Kim YK, Sohn CH, Kang KM, Lee YJ, Lee DY, Group KR. Circadian rest-activity rhythm and longitudinal brain changes underlying late-life cognitive decline. Psychiatry Clin Neurosci. 2023;77(4):205-12.
51. Xiao Q, Sampson JN, LaCroix AZ, Shadyab AH, Zeitzer JM, Ancoli-Israel S, Yaffe K, Stone K, Osteoporotic Fractures in Men Study G. Nonparametric Parameters of 24-Hour Rest-Activity Rhythms and Long-Term Cognitive Decline and Incident Cognitive Impairment in Older Men. J Gerontol A Biol Sci Med Sci. 2022;77(2):250-8.

52. Zhu W, Wadley VG, Howard VJ, Hutto B, Blair SN, Hooker SP. Objectively Measured Physical Activity and Cognitive Function in Older Adults. Med Sci Sports Exerc. 2017;49(1):47-53.
53. de Feijter M, Kocevska D, Ikram MA, Luik AI. The bidirectional association of 24-h activity rhythms and sleep with depressive symptoms in middle-aged and elderly persons. Psychol Med. 2023;53(4):1418-25.
54. Cabanas-Sanchez V, Esteban-Cornejo I, Garcia-Esquinas E, Ortola R, Ara I, Rodriguez-Gomez I, Chastin SFM, Rodriguez-Artalejo F, Martinez-Gomez D. Cross-sectional and prospective associations of sleep, sedentary and active behaviors with mental health in older people: a compositional data analysis from the Seniors-ENRICA-2 study. Int J Behav Nutr Phys Act. 2021;18(1):124.
55. Chan LLY, van Schooten KS, Lord SR, Delbaere K, Brodie M. Short Daily-Life Walking Bouts and Poor Self-Reported Health Predict the Onset of Depression in Community-Dwelling Older People: A 2-Year Longitudinal Cohort Study. J Am Med Dir Assoc. 2022;23(7):1242-7 e3.
56. Gianfredi V, Schaper NC, Odone A, Signorelli C, Amerio A, Eussen S, Kohler S, Savelberg H, Stehouwer CDA, Dagnelie PC, Henry RMA, van der Kallen CJH, van Greevenbroek MMJ, Schram MT, Koster A. Daily patterns of physical activity, sedentary behavior, and prevalent and incident depression-The Maastricht Study. Scand J Med Sci Sports. 2022;32(12):1768-80.
57. Ruiz-Perez D, Guan HB, Madhivanan P, Mathee K, Narasimhan G. So you think you can PLS-DA? Bmc Bioinformatics. 2020;21.
58. Petermann-Rocha F, Lyall DM, Gray SR, Gill JMR, Sattar N, Welsh P, Quinn TJ, Stewart W, Pell JP, Ho FK, Celis-Morales C. Dose-response association between device-measured physical activity and incident dementia: a prospective study from UK Biobank. BMC Med. 2021;19(1):305.
59. Raudsepp L, Riso EM. Longitudinal Association Between Objectively Measured Walking and Depressive Symptoms Among Estonian Older Adults. J Aging Phys Act. 2017;25(4):639-45.
60. Kandola A, Lewis G, Osborn DPJ, Stubbs B, Hayes JF. Device-measured sedentary behaviour and anxiety symptoms during adolescence: a 6-year prospective cohort study. Psychol Med. 2022;52(14):2962-71.
61. Hedin RJ, Umberham BA, Detweiler BN, Kollmorgen L, Vassar M. Publication Bias and Nonreporting Found in Majority of Systematic Reviews and Meta-analyses in Anesthesiology Journals. Anesth Analg. 2016;123(4):1018-25.
62. Vadillo MA, Konstantinidis E, Shanks DR. Underpowered samples, false negatives, and unconscious learning. Psychon Bull Rev. 2016;23(1):87-102.