Doctoral theses at NTNU, 2020:351

Ronny Bergquist

Preventing functional decline in young seniors

Development and evaluation of interventions and smartphone-based tests

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

NTNU
Norwegian University of Science and Technology

Ronny Bergquist

Preventing functional decline in young seniors

Development and evaluation of interventions and smartphone-based tests

Thesis for the Degree of Philosophiae Doctor

Trondheim, November 2020

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Neuromedicine and Movement Science

ISBN 978-82-326-5048-4 (printed ver.)
ISBN 978-82-326-5049-1 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:351

Printed by NTNU Grafisk senter

Preventing functional decline in young seniors

- Development and evaluation of interventions and smartphone-based tests

Kandidatens navn: Ronny Bergquist

Forskningsgruppe: Forskningsgruppe for geriatri, bevegelse og slag (GeMS)

Institutt for nevromedisin og bevegelsesvitenskap, Fakultet for medisin og helsevitenskap

Norges teknisk-naturvitenskapelige universitet

Hovedveileder: Kristin Taraldsen

Biveiledere: Beatrix Vereijken, Jorunn L. Helbostad, Sabato Mellone

Finansieringskilde: RSO-stilling fra NTNU etter tilslag fra EU-programmet Horizon 2020

NORSK SAMMENDRAG

Forebygging av funksjonstap hos unge eldre

- Utvikling og evaluering av intervensjoner og smarttelefon-baserte tester

Funksjonstap hos eldre kan føre til fall, redusert aktivitetsnivå, behov for hjelp i daglige gjøremål og redusert livskvalitet. Funksjonstap er derfor en stor belastning både for den som rammes og deres pårørende, men også for samfunnet sett fra et økonomisk perspektiv. Som følge av økt forventet levealder og lavere fødselstall, ser vi en endring i befolkningssammensetningen med en større andel eldre, som er forventet å øke i årene som kommer. En konsekvens av denne endringen er et økende press på helsesektoren. Det vil derfor bli viktig å flytte fokuset fra behandling til forebygging, slik at mennesker kan leve med en god helse lengre og dermed utsette behovet for helsetjenester.

Hovedmålet med prosjektet var å generere ny kunnskap om hvordan funksjonstester instrumentert ved hjelp av smartteknologi kan brukes til å måle fysisk funksjon, og hvordan treningsprogrammer kan utvikles for å forebygge funksjonstap hos yngre eldre. Avhandlingen inkluderer en litteraturgjennomgang, en gjennomførbarhetsstudie, en brukbarhetsstudie og en metodestudie, hvor målsetningen består av fire deler:

1) å gå systematisk gjennom litteraturen som omfatter funksjonstester for styrke og balanse som har blitt brukt for å teste yngre eldre, og å evaluere måleegenskapene til de identifiserte testene.

2) å evaluere gjennomførbarheten av et livsstilsintegrert treningsprogram (aLiFE) er tilpasset yngre eldre ved å legge til flere og mer utfordrende øvelser.

3) å beskrive utviklingen og brukbarhetstestingen av tre smarttelefon-apper med selv-tester av fysisk funksjon.

4) å undersøke om en instrumentert versjon av den kliniske testen 'Timed Up and Go' (iTUG) kan predikere score på en mer avansert test av balanse og mobilitet, 'Community Balance and Mobility Scale' (CBMS), og om iTUG kunne oppnå samme eller bedre resultat sammenlignet med et batteri av tradisjonelle kliniske funksjonstester.

Litteratursøket identifiserte 3454 artikler med totalt 120 inkluderte tester for å måle fysisk funksjon hos yngre eldre. Av disse testene, ble 30 tester hyppigere brukt enn de andre. Andre del av litteratursøket identifiserte metodestudier som evaluerte måleegenskapene til 6 av de 30 hyppigst brukte testene. CBMS var den testen som ble ansett mest egnet, da den ikke var disponert for tak-effekter, og ble funnet valid og reliabel for å teste fysisk funksjon hos yngre eldre.

Gjennomførbarhetsstudien av aLiFE viste at programmet var gjennomførbart i de 4 ukene studien pågikk, hvor flertallet av deltakerne vurderte programmet med høy score på kategorier som etterlevelse, utførelsesfrekvens, nyttighet, de ulike øvelsenes mulighet til å implementer og tilpasse, vanskelighetsgrad, og trygghet. Studien var ikke designet for å måle effektivitet, men resultatene indikerte en liten forbedring i fysisk funksjon målt med CBMS, i tillegg til en ikke-signifikant økning i daglig fysisk aktivitet. En større klinisk kontrollert studie med lengre oppfølgingstid vil være nødvendig både for å evaluere effekten av programmet og for å evaluere eventuelle varige atferdsendring.

Brukbarhetsstudien benyttet et brukersentrert design til å utvikle smarttelefon-baserte selv-tester av fysisk funksjon, hvor appene ble testet og videreutviklet gjennom tre iterasjoner. Resultatene viste at deltakerne gjorde en rekke feil under testingen når deltakerne selv-administrerte testene uten veiledning, men antall feil ble redusert gjennom forbedringene av appene. Den siste testingen ble gjennomført i deltakernes hjem, og viste en mindre økning i feil sammenlignet med forrige testing. Studien viste at appene har et potensiale for å brukes som selv-tester, men utfallsmålene må valideres før testene kan bli gjort tilgjengelig for yngre eldre.

Siste studie var en metodestudie som viste at ved å instrumentere den kliniske testen 'Timed Up and Go', kunne man med hjelp av en maskinlæringsprosedyre predikere scoren på CBMS-testen med høy nøyaktighet i et utvalg geriatriske pasienter og hjemmeboende eldre.

Denne avhandlingen har gitt ny kunnskap som viser at det er et behov for funksjonelle tester som er sensitive, tilgjengelige, valide og reliable for å måle fysisk funksjon hos yngre eldre. Selv-administrerte tester i app-format kan potensielt gjøre testing enklere og mer tilgjengelig. Resultatene fra avhandlingen viser også at trening og fysisk aktivitet integrert i daglige gjøremål og rutiner kan være et alternativ til tradisjonelle treningsprogram, som kan gi økt etterlevelse gjennom atferdsendring. Flere studier er nødvendig for å validere utfallsmålene fra de smarttelefon-baserte selv-testene presentert i denne avhandlingen, og en større klinisk kontrollert studie med lengre oppfølging er nødvendig for å evaluere effekten av programmet med trening og fysisk aktivitet integrert i daglige gjøremål og rutiner.

Acknowledgements

This PhD-project was carried out at the Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), by support from the PreventIT-project that was funded by European Union's Horizon 2020 framework-programme.

I would like to thank my main supervisor, Kristin Taraldsen, for your never-ending support, for always being available for questions and discussions, for always being so encouraging and motivating, and for always sharing your knowledge and experience from academia.

Thanks also to my co-supervisors, Beatrix Vereijken, Jorunn Helbostad, and Sabato Mellone, for introducing me to the fields of geriatrics and technology, and to conferences and researchers around the world. It has been inspiring and a pleasure to work alongside you for all these years.

I would also like to thank all the great colleagues I have worked with, including the entire PreventIT-consortium and all of you in the GeMS research group. Especially, I would like to thank Yngvild Gagnat, Phillipp Anders, Gemma Goodall, and Per Bendik Wik, for all the good times we have shared.

Thanks to all participants who volunteered in the studies. Your enthusiasm and curiosity have been highly appreciated and valuable in giving more meaning to it all.

To my family and friends, thank you for supporting me. Special thanks to my own little family, Iris and Bobby, for all your love and support.

Table of contents

Abbreviations and frequently used phrases

ADL	Activities of daily living
aLiFE	adapted Life-style Integrated Functional Exercises
App	Application
Feature	A parameter calculated from the inertial sensor signal(s), for example: ‘duration of turning’, ‘mean velocity during turning’, ‘maximum velocity during turning’.
IMU	Inertial measurement unit
Iteration	A cyclic process of prototyping, testing, analysing, and refining a system
iTUG	instrumented Timed Up and Go
mHealth	Mobile health
LiFE	Life-style Integrated Functional Exercises
PA	Physical activity
PLSR	Partial least square regression
RCT	Randomised controlled trial
Self-test	A test that can be administered by oneself
STS	sit-to-stand
SUS	System Usability Scale
UEQ	User Experience Questionnaire

Abstract

Decline in physical function in seniors leads to falls, decreased physical activity, dependence in activities of daily living, and reduced quality of life, constituting a major burden for the individual itself, and an economic burden for the health care system. In addition, the age composition of the worlds' population is changing fast; life expectancy is increasing, and fertility rates are decreasing, which is prospected to put stress on health care services. This thesis is based on the recognition that there is a need to find effective methods to prevent functional decline in younger seniors. By preventing functional decline, seniors can achieve more active years without disability and dependence in everyday activities. Research has indicated that sensor technology might increase the utility of clinical tests in measuring physical function. Furthermore, studies have shown that exercise programmes where activities for muscle strength and balance have been integrated in daily habits and routines can be more effective in falls prevention compared to traditional exercise programmes.

The thesis aims to provide new knowledge on how we can use instrumented performance-based tests to measure and monitor physical function, and how interventions can be designed to help to improve physical function and increase physical activity. The aim of this thesis, which includes one literature review, one feasibility study, one usability study and one method study, was fourfold:

1) to systematically review the literature on commonly used tests of balance and strength and to evaluate their measurement properties in young seniors;

2) to evaluate the feasibility of an intervention, the adapted Lifestyle-integrated Functional Exercise (aLiFE) programme, adapted to be more challenging and suitable for preventing functional decline in young seniors;

3) to describe the development and usability testing of three smartphone-based self-tests of physical function using an iterative design in a home-setting;

4) to evaluate whether an instrumented Timed Up and Go (iTUG)-test can predict the score on a advanced test of balance and mobility. Further, we wanted to evaluate whether the iTUG model was equally or more predictive compared to a model of traditional clinical measurements in geriatric outpatients and healthy community-dwelling seniors.

The systematic review identified 3454 papers with a total number of 120 performance-based tests of physical function that had been used in healthy young seniors. Of these, 30 tests were most commonly used. The second step of the review identified 9 studies which had evaluated the measurement properties of 6 different tests: the Timed Up and Go (TUG), Short Physical Performance Battery (SPPB), 10-second Tandem stance, Five Times sit-to-stand, and the Community

Balance and Mobility Scale (CBMS). Of the identified tests, the CBMS seemed like the one most suited for application in healthy young seniors as it has no ceiling or floor effects and was found valid and reliable in healthy young seniors in the included method studies.

The feasibility study of the aLiFE programme showed that overall it was feasible and well accepted by the young seniors. The participants implemented most of the activities during the 4-week intervention period, and perceived the programme to be helpful, adaptable to their lifestyle, appropriately challenging, and safe. Although the study was not designed to evaluate effectiveness, moderate improvements in physical function measured by the CBMS were found. A larger study with longer follow-up is needed to establish effectiveness of the programme, as well as its ability to induce long-term adherence through behavioural change. In the usability study, a human-centered design was used when developing and testing the applications (apps) for three clinical tests of physical function, through 3 iterations. Results showed that young seniors made several errors while self-administering the clinical tests with a smartphone in an unsupervised setting, and that rate of errors was affected by changes made across the iterations. Although smartphone-based self-testing of physical function is promising, validation of outcome measures is needed before being offered to end-users. Finally, we found that in a supervised lab-setting, the instrumented Timed Up and Go could predict the score on the Community Balance and Mobility Scale with a high and similar accuracy as compared to standard clinical tests, in a mixed sample of geriatric patients and community-dwelling seniors.

This thesis showed that there is a need for functional tests which are more sensitive, available, valid and reliable in young seniors. Important insight was provided about what usability problems may arise when developing smartphone-based functional self-tests, and how changes to the apps affect such problems. Furthermore, with a partial least squares regression analysis it was found that by instrumenting the TUG with a smartphone, a seemingly simple clinical test which otherwise is not sensitive for detecting functional decline in young seniors, could predict higher-level balance and mobility as measured with the CBMS. The findings indicate that smartphone apps can be feasible for home-based assessments and monitoring of physical functions, potentially motivating the end-user to become more physically active and prevent functional decline. Future work should aim to further develop and validate the self-tests and integrate these tests into a solution that combines testing with an exercise intervention tailored to the user based on the test results. A life-style integrated exercise-approach was evaluated in the thesis, and the results were promising, however more research is needed to verify its effectiveness.

List of publications

Paper I *Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review*
Bergquist R, Weber M, Schwenk M, Ulseth S, Helbostad JL, Vereijken B, et al.
BMC Geriatrics. 2019;19(1):9.

Paper II *The Adapted Lifestyle-Integrated Functional Exercise Program for Preventing Functional Decline in Young Seniors: Development and Initial Evaluation*
Schwenk M, Bergquist R, Boulton E, Van Ancum JM, Nerz C, Weber M, et al.
Gerontology. 2019;65(4):362-74.

Paper III *App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study*
Bergquist R, Vereijken B, Mellone S, Corzani M, Helbostad JL, Taraldsen K.
JMIR mHealth and uHealth. 2020;8(4):e16507-e.

Paper IV *Predicting advanced balance ability and mobility with an instrumented Timed Up and Go-test*
Bergquist R, Nerz C, Taraldsen K, Mellone S, Vereijken B, Helbostad JL, Ihlen EAF, Mikolaizak AS.
Manuscript submitted for publication.
[Published 3. September in Sensors 2020;20(17):4987]

1. Introduction

1.1 Aging and physical function

There is no common definition of physical function. In their Disablement Process-model from 1994, Verbrügge and Jette defined *functional limitation* as restrictions in basic mental and physical actions, and *disability* as a concept encompassing difficulty in performing activities of daily living (ADLs)(1). Functional disability can be viewed as the gap between personal capability and environmental demands. More recent definitions, however, have included the ability (or the lack thereof) to carry out ADLs and advanced physical activities (2). In this thesis, *physical function* is thus understood as the ability to perform basic and advanced physical actions, and *functional decline* is understood as the deterioration of physical function. The definition includes upper- and lower-extremity function, but this thesis focuses on lower-extremity function, which is important for mobility (3).

As we age, our physical function deteriorates. Age-related loss of physical function presents a major burden for the individuals and their dependents, as well as for the health care system economically. Due to increased life-expectancy and lower fertility rates globally, it is expected that the health care services will be put under substantial pressure (4). Moreover, the United Nations have claimed that the change in population age composition that follows, will be one of the most disruptive social transformations of the 21st century (4, 5).

The point in life at which our physical function starts to decline differs greatly between individuals, but chronological age is nonetheless an important risk factor for functional decline (6). Other risk factors with strong evidence for functional decline include cognitive impairment, depression, comorbidity, high and low body mass index (BMI), reduced observed lower extremity performance, low frequency of social contact, low level of physical activity, no alcohol use, poor self-perceived health, smoking and vision impairment (2). At time of retirement, between the age of 60-70, balance, gait and mobility typically start to decline at a higher rate than before (7, 8). For many people, retirement is a life event where they experience large changes in their physical activity levels (9, 10), which is an important risk factor for functional decline (2).

1.2 Assessment of physical function

Physical age-related changes typically include loss of muscle strength, proprioception, vision, vestibular sense, and reaction time (11). The decline in functioning of these systems impairs our balance ability, and also affects our performance in activities such as walking, chair rising, stair climbing, stepping and postural transfers (11, 12). Physical function can be measured by self-report or performance-based tests. Self-reported physical function is often measured with questionnaires

consisting of items that ask about the perceived ability to- or difficulty in- performing ADLs. What we measure with physical tests is *physical performance*, and it is commonly understood as the observable ability to perform tasks, such as standing up from a chair, walking, turning, or standing on one foot (13). The ability to perform such activities typically diminishes slowly, and the types of tasks people find challenging seem to follow a hierarchical pattern, with tasks requiring balance, strength and agility being impaired first, such as doing heavy housework, walking longer distances or using stairs (14).

With standardised tests, self-reported or performance-based, one can get an indication of a person's physical function at that point in time by comparing the result with normative values for the person's age or condition. By repeating the assessment later, we can get an indication whether there's been a change in the physical function, e.g. after an exercise intervention. The test results, if found valid for this purpose, can also predict future health status, and measures of physical function could thus be regarded as "vital signs" to screen for in clinical settings (15).

Physical function in seniors is commonly assessed with single tests or test batteries, that measure domains such as walking speed, grip strength, standing balance, and ability to make postural transfers (sit-to-stand) (16). What characterises the most commonly used performance-based tests is that they can be administered quickly and with little space and equipment. In addition, they often measure more than just one domain of function, and, most importantly, their measurement properties have been evaluated in method studies. One of the most commonly used tests is the Timed Up and Go (TUG) test (17), which is widely recognised and allows the analysis of sit-to-stand transitions, walking and turning in one test that is quick and easy to administer. The test is face valid to most populations and can be easily taught to health care professionals. However, the total duration of completing a single TUG test, measured in seconds, does e.g. not discriminate fallers from non-fallers (18), or accurately predict falls in higher-functioning seniors (19).

Other commonly used functional tests in older adults of have also been found inappropriate to use in younger or higher-functioning seniors, primarily due to ceiling effects (20, 21). Ceiling effects, or its opposite, floor effects, refer to the phenomena where scores on a test or scale cluster towards the ends of the scale. A criterion often used to evaluate whether a test is prone to ceiling/floor effects is if 15% of the participants achieves the highest (ceiling) or lowest (floor) score possible on a test (13). Ceiling/floor effects poses a serious problem in both research and clinical practice because comparison between subjects in the top/bottom end of the scale becomes impossible, and we may not be able to identify changes over time (22). There is consequently a need for new ways of

measuring functional decline in seniors that overcome the issues associated with ceiling effects, to be able to detect early changes in physical function.

1.3 Interventions for preventing functional decline in young seniors

Physical activity has been found to prevent both the onset and progression of functional decline (23), and reduce the risk of hypertension, type 2 diabetes, some cancers, dementia, depression, bone health, falls and fall-related injuries (24). Seniors are recommended to follow the same guidelines for physical activity and exercise as adults in general, as the principles for the effect of physical activity are similar (24, 25). Public guidelines for physical activity recommend 150-300 minutes of moderate-intensity, or 75-150 minutes per week of vigorous-intensity aerobic physical activity, or a combination (24). In addition, the guidelines recommend sitting less, and doing muscle-strengthening activities of moderate or greater intensity that involve all major muscle groups twice or more per week (24). Following the guidelines can help minimise the negative effects of an otherwise inactive lifestyle, and limit the development and progression of chronic disease and disabling conditions (25). Although these guidelines have been available for many years, only about one third of adults and seniors follow them in Norway (26). As such, there is a need for specific strategies that can foster regular physical activity and strength training during aging to remain active and independent for as long as possible.

Structured exercise programs can potentially make it easier for people to engage in regular physical activities, and have proved effective in the short term (27). There are, however, two important challenges with structured exercise programs. First, many seniors don't find structured exercise appealing, and report reasons for not engaging in structured exercise such as lack of transportation, limited access to facilities, time commitments, unwillingness to join a group, or aversion to exercise due to not regarding themselves as "sporty" (28). Second, structured exercise programs often fail to induce behavioural change, and consequentially authors have repeatedly discussed the lack of long-term adherence to structured exercise programmes (28). Lack of adherence to a programme is a problem both for individuals who would benefit from them, and in research, as sufficient adherence rates to the programme is required for establishing the effectiveness of a given exercise intervention. To achieve long-term adherence, exercise programmes for young seniors should thus incorporate strategies for inducing behavioural change.

An alternative way of designing exercise programmes to increase adherence, is to integrate exercises into the participants' daily routines (28). With this approach, the exercises are integrated into routines that people already have, so there is no need to dedicate extra time to perform them. The exercises can be performed while doing virtually any daily task, such as squatting while getting

something out from a low drawer or shelf, climbing the stairs two steps at a time, doing walking lunges or tandem walking while moving around the house, doing calf raises when picking something from a high shelf, or balancing on one foot while cleaning the teeth. In line with the principle of specificity, the exercises are considered functional as they enhance the basic everyday motor performances of common ADLs.

There is some evidence that intervention programmes integrating functional exercise into daily life are advantageous compared to traditional exercise programmes in older adults, in terms of effectiveness and adherence (28). One example of such a programme is the Lifestyle-integrated Functional Exercise (LiFE)-programme, which was developed and tested in seniors with the main aim being to reduce falls (29). In a randomized controlled trial (RCT), the LiFE-programme was also found to increase adherence, reduce falls, improve function, decrease disability, compared to a traditional exercise programme and a sham interventions (29). Few RCTs have, however, evaluated whether interventions with life-style integrated exercises are feasible and effective in younger seniors (28, 29). Given the need for exercise interventions that can foster behavioural change and long-term adherence for preventing functional decline, efforts should be made to investigate the usefulness of such a program tailored specifically to young seniors.

1.4 Technological advancements in functional assessment

Technological innovations over the years have had a large impact on the objective measurement of physical function. Sensor-based tests, so-called instrumented tests, let us analyse aspects of human movement with a higher precision as compared to traditional methods, such as human observation and manual time-taking. Early examples are the force plates introduced in the 60s and 70s, which can quantify the postural control of a person standing on it by measuring the applied ground reaction forces. Another example are instrumented mats, that when you walk on them, they can register the footsteps from which several temporal and spatial gait parameters can be calculated, such as step time, cycle time, step length and single/double support (30). Another example is the robotic dynamometer, a machine that allows you to perform controlled measurements of strength/power during isokinetic flexions/extensions in upper and lower extremities (31).

Devices such as force plates, instrumented mats and dynamometers are mostly restricted to hospitals and advanced research labs due to their size and price levels. With the introduction of inertial measurement units (IMUs), however, accurate and objective measurements of human movement have become more available also outside the lab (32). IMUs are relatively small devices equipped with sensors such as accelerometers, gyroscopes and magnetometers, that measure acceleration, angular acceleration and magnetism, respectively. With sophisticated algorithms, the

sensor signals recorded by the IMUs can be processed with computer software to quantify the movement of a person wearing the IMU. Modern IMUs come as small, lightweight pieces with on-board battery, inertial sensors, microprocessor and memory, which can be worn rather discretely on the body. They can withstand water immersion, and their batteries allow them to record movements continuously for several weeks.

The characteristics of IMUs have made them valuable in lab-based fall risk assessment, where they have improved predictive accuracy of commonly used clinical tests of sit-to-stand, walking and balance control (33, 34). The instrumentation of TUG (iTUG) with sensors allows measuring of spatial and temporal features from the different segments of an iTUG trial, such as sit-to-stand transitions, walking, and turning. These additional features have made the iTUG more sensitive for measuring physical function as compared to the original TUG in people with fall risk, Parkinson's disease, disability or cognitive impairment (19, 35-39). Especially the turning features from an iTUG have been found sensitive in populations with impaired motor control due to neurological conditions (37, 38, 40), fallers (41), and in persons with mild cognitive impairment (36), which could also be explained by motor control impairment (36, 42).

Static posturography assessments instrumented with wearable IMUs provide several features/parameters which have been shown to be reliable and valid for quantifying postural control (34, 43), and that are sensitive to differences in performance between healthy controls and people with mild neurological conditions (44), healthy young and senior people (34), and fallers and non-fallers (45).

The instrumented sit-to-stand has also been found valid for a number of important outcomes (46-51). One study showed that phase-specific durations were comparable with those measured by opto-electronic motion capture systems (47). Another study with seniors showed that movement duration measured with IMUs was more strongly associated with quality of life and self-reported physical function as compared to manually recorded measures of durations (51). Furthermore, with instrumented measurements of sit-to-stand tests one can analyse discrete segments of the movement, e.g. the dynamic part (rising or sitting down), duration and angular velocity (trunk flexion/extension), which are sensitive to age (50), fall risk (48) and functional status (49).

With modern smartphones becoming more and more sophisticated in terms of computational power, memory, size and weight, and the sensors they are equipped with, the smartphone could be regarded an IMU itself. In the EU-funded project FARSEEING (52), one of the tasks was to evaluate whether a smartphone could validly and reliably measure performance during a TUG test compared to an IMU designed for exactly that purpose, and the results were positive (53). Another study

demonstrated that a smartphone-based measurement of postural stability was comparable to a force plate, and it was able to discriminate between younger seniors at low and high risk of falling (54). The smartphone has a particular potential that ordinary IMUs don't have, namely that so many people around the world owns a smartphone - and the number is increasing, also among seniors (55). This has paved the way for integrating clinical tests into self-administrable smartphone apps that allow users to perform the tests unsupervised in their own homes. The app can provide instant, objective feedback on their functional performance compared to normative data or compared to their own previous scores. In other contexts, it could also be possible for end-users to have their physical function monitored remotely by clinicians or researchers.

Using smartphones to administer clinical tests of physical function is a relatively new field, and little literature exists on the measurement properties or the usability of app-based self-tests of physical function. A systematic review from 2017 identified 13 studies in which smartphone apps had been used to evaluate physical function (56). Of the 13 studies, only 5 evaluated the validity, and 3 the reliability of the app-based measurements. Furthermore, a majority of the applications were intended to be used in a clinical context, and no studies had evaluated the usability of the apps. The authors concluded that apps need to be designed with consideration of the end-users' level of function, and test for usability in the target user-group (56).

1.5 App design and development

Innovations in technology have introduced connected health system, which refers to health technology used to deliver health care services remotely, such as eHealth, telehealth, telemedicine and mHealth (57). Connected health systems often have a user interface which requires some degree of human-computer interaction. How the user interface is designed can have important implications for a product or service, as demonstrated by the observation that otherwise excellent products have failed in the market due to poor interface design, while well-designed mediocre products have flourished (58). Therefore, when developing an mHealth app for unsupervised self-assessment of physical function, the design and usability of the app must be carefully considered.

One approach to the design and development of apps is the human-centred design (HCD) process, which the International Organisation for Standardization (ISO) defines as "an approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics and usability knowledge and techniques." (59). The idea is that a system or product will be better if representatives from the target user group are involved throughout the design and development process, to ensure that their needs and preferences are truly reflected in the final product. HCD has become the norm, and end-user

involvement during the design process is even required documented for FDA approval (58). According to ISO, a human-centred design "enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of use on human health, safety and performance" (59). For a design process to be called human-centred, the ISO 9241-210 outline six requirements that need to be met: (1) The design is based upon an explicit understanding of users, tasks, and environments; (2) Users are involved throughout design and development; (3) The design is driven and refined by user-centered evaluation; (4) The process is iterative; (5) The design addresses the whole user experience; (6) The design team includes multidisciplinary skills and perspectives (58).

Requirements 3 and 4, as listed above, point to the need for usability testing. Usability is defined by ISO as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (59). When measuring aspects of one's health, the accuracy of the results relies on correct administration of the test. As one author phrased it: "If an application is valid and reliable, it will be of little use if the intended population cannot properly operate it." (56) Thus, any usability problem associated with the apps could directly affect the validity and should be identified and addressed before it is made available to end users.

Usability testing is usually done through several iterations of testing with target user groups, ideally until there exist no major usability problems associated with the app. Usability studies are most often carried out in a lab setting, which is convenient and offers a high degree of control, as opposed to field-based usability testing. However, field-based testing, which, in this context, would be a home setting, provides insight into how the system is used under more realistic situations. Depending on the system being tested and the phase of development, usability should ideally be tested in both lab and home settings.

Current literature suggests that there is little consensus on how usability testing of app-based functional tests in seniors should be performed. Among three studies identified (60-62), two used an iterative design in which the app was improved based on feedback and results from previous iterations (60, 61), while the third study describes only one session of usability testing (62). The studies also vary with respect to which usability data they collect, and how they collect it. User experience questionnaires and interviews were used in all three studies, a Think Aloud protocol was used in two of the studies (60, 61), and video-taped user-testing was used in one study (61). Whilst one of the studies conducted assessments in both a research lab facility and an unoccupied apartment to mimic a home setting (60), the others did not report the setting in which testing was conducted. Finally, all of these studies describe an app that can potentially be used unsupervised,

however none of the studies have reported the study participants' observed ability to perform the tests unsupervised. If end-users perform the test incorrectly, the results are invalid, and thus usability studies should verify that the app can be used correctly in an unsupervised setting.

1.6 Rationale for the thesis

There is a need for innovative strategies to prevent functional decline in the aging population. The adoption of digital technologies has been recognized as a key strategy for cutting costs in health care (57). App-based self-administrable tests could serve as a valuable tool for young seniors to monitor their physical function and detect subtle changes at an early stage where there is still time and opportunity to prevent it. However, it has been recognised that a significant gap in health technology exists between the clinical functionality on one side and the user experience on the other, which warrants attention (63). Several papers have described smartphone apps developed to measure balance and fall risk, but very few of the apps have been designed and tested for usability for unsupervised use in young seniors (56). The prevention of functional decline does not only require timely, user-friendly, valid, reliable and sensitive measurements, but also interventions that are tailored specifically to young seniors and their needs in order to effectively help them maintain or improve their physical function. There are many established tests and interventions developed to detect and prevent functional decline, but they are either not designed with young seniors in mind, and thus not suitable for use in this target group, or they are complex, time- and resource intensive, which limits their feasibility for use in large scale public health approaches. Thus, there is a need for structured and well-planned studies to gain insight on feasibility and validity of innovative ways for measuring and preventing functional decline in young seniors.

1.7 Aims of the thesis

The overall aim of this thesis was to provide new knowledge on how exercise interventions and functional assessments should be designed for preventing functional decline, and how smartphone-based apps can be used as a platform for delivering such interventions and assessments to young seniors. The work consists of four papers: one systematic review and three prospective studies comprising one feasibility study and two method studies focusing on the development of smartphone-based apps of self-assessment of physical function to evaluate the usability and the measurement properties.

The specific aims were as follows:

Paper I:

To 1) identify commonly used tests of balance and strength, and 2) to evaluate their measurement properties in young seniors.

Paper II:

To 1) develop the aLiFE programme to be more challenging and suitable for preventing functional decline in young seniors, and 2) to perform an initial feasibility evaluation of the program. We also assessed pre-post changes in physical function.

Paper III:

To describe the development and usability testing of three smartphone-based self-tests of physical function using an iterative design in a home-setting.

Paper IV:

To 1) evaluate how well the averaged inertial sensor features from 5 iTUG repetitions can predict the CBMS total score within a group of geriatric outpatients and healthy community-dwelling seniors, and 2) to investigate whether the iTUG, and which components of the iTUG can predict the CBMS total score more accurately, compared to standard clinical tests used in routine assessments today.

2. Methods

The work presented in this thesis was conducted as part of the EU-funded PreventIT project (the European Horizon 2020 research and innovation programme, grant agreement number: 689238). Data were collected in Trondheim, Norway, at the Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), and at two of the collaborating clinical sites in the PreventIT-project; Amsterdam, The Netherlands, and Stuttgart, Germany. The search for papers for the systematic review (Paper I) was conducted in November 2018; the data for Paper II were collected between May and July 2016; the data for the first part of Paper III were collected in Trondheim, Stuttgart, and Amsterdam during summer of 2017 and 2018, and the second part of Paper III in Trondheim between May and April 2019; the data for Paper IV was collected in Stuttgart from December 2018 to September 2019.

2.1 Design and data collection

The four papers in the thesis focus on development of exercise interventions and instrumented, supervised and non-supervised clinical tests of physical function for preventing functional decline, by using smartphone apps as a means for delivering it to young seniors. We conducted a systematic review (Paper I), a feasibility study describing development and pre-post changes of a novel exercise programme (Paper II), a usability study of smartphone-based self-test apps (Paper III), and, finally, a method study assessing the predictive accuracy of smartphone-based measurements of physical function (Paper IV). An overview of study design and the studies from which we collected data for the four papers, are presented in Table 1.

Table 1. Study designs and data source of the papers included in this thesis.

Paper	Design	Project
I	Systematic review	
II	Feasibility evaluation	The aLiFE feasibility study
III	Usability evaluation	The PreventIT feasibility RCT and the Usability study
IV	Method study	The iTUG study

The PreventIT project aimed to develop and test a personalised ICT-based intervention aimed at behavioural change in people who have recently retired, in order to decrease risk for age-related functional decline. The data used in paper II-IV were collected from three different studies carried out in collaboration with partner institutions of the PreventIT project.

Paper II is the main paper from the aLiFE feasibility study and describes the development and feasibility evaluation of the aLiFE programme, which was developed to constitute one of the study arms in the PreventIT feasibility RCT. The aLiFE study included participants from the three clinical sites, Trondheim, Stuttgart, and Amsterdam. Participants took part in an exercise intervention over 4-weeks. This was designed as an uncontrolled pre-post -test intervention study. The intervention included four home-visits by trained assessors. Data were collected from the pre- and post-test assessments, as well as during the home visits during the intervention, by six trainers (exercise scientists and physiotherapists) certified to deliver the aLiFE programme.

Paper III describes a usability evaluation of app-based self-tests of physical function, consisting of data from two separate studies. Data for the first part were collected from the The PreventIT feasibility RCT and included participants from the three clinical sites in Trondheim, Stuttgart, and Amsterdam. This feasibility RCT allocated participants to one of three groups: a control group receiving general advice on physical activity and two intervention groups receiving either the aLiFE-programme, delivered as in the conventional LiFE-programme through home-visits and paper manuals, or the same intervention but delivered through a connected system of smartphone, smartwatch and an integrated app. All participants attended a baseline assessment, a 6-month intervention period, post-test assessments, a 6-month unsupervised intervention period, and a follow-up assessment. As part of the baseline and follow-up assessment, participants performed unsupervised self-administered tests of physical function using smartphone apps. The trials were observed by assessors, and usability data were recorded and used for the first part of Paper III. For the second part of Paper III, we designed a home-based usability study to conduct a usability evaluation of how well target end-users of the apps would be able to self-administer clinical tests unsupervised in their own homes. This study was conducted in Trondheim, Norway. Data collected from the testing consisted of video recordings, questionnaires, interviews and inertial sensor signals recorded by the smartphone that the participants used to self-administer the tests.

Paper IV report the results from the iTUG study, conducted at the Robert-Bosch-Hospital in Stuttgart, Germany. This was a method study to evaluate the ability of the iTUG to predict the score on the CBMS (20). The CBMS is a test-battery that assesses advanced balance and mobility function. In addition to CBMS and the iTUG, outcomes consisted of self-reported and objectively measured standard clinical tests of physical function. The test order was randomised, with participants starting with either the iTUG or the standard clinical tests first.

2.2 Study sample characteristics

The primary target group of the work conducted in this thesis was healthy community-dwelling young seniors around the time of retirement, 61-70 years, as this age represents a timely opportunity to initiate measures for preventing functional decline.

For the purpose of identifying tests which have been used to measure physical function (paper I) and creating an effective exercise intervention (paper II), we exclusively recruited healthy seniors between 60 and 70 years old. For developing and testing the usability of app-based self-tests (paper III), we recruited participants from a wider age range, 60-80, in order to have a sample that represented a wider range of the target population. The proportion of smartphone users in the age of 60-70 in the western world is relatively high (55) and rising fast, and there is a lot of variation in how confident they are in their abilities to make full use of the features available in a smartphone. For the prediction model we built in the iTUG study (paper IV), there was a need to have heterogeneity in functional performance to ensure that the study results would have been generalisable. To this end, we included two groups in the age range 60 to 85; one of healthy community-dwelling seniors, and one group of geriatric rehabilitation patients. A detailed description of the participants and eligibility criteria are presented in the following.

Paper I (16) was a systematic review of a total of 295 articles with a target population either between the age of 60-70 or part of a group of which the mean age was within the 60-70-year age range. In 282 of the articles, the mean age was within the target range, whereas 13 studies included participants between the age of 60-70.

Paper II (64) included 30 community-dwelling persons aged 60–70 who were not frequently exercising. The ten participants at each of the three clinical sites were recruited via newspapers and flyers. Exclusion criteria were inability to walk 500 metres without walking aids, cognitive impairment (defined as a Montreal Cognitive Assessment (65) score <24 points), diseases where exercise is contraindicated, and frequently exercising (defined as attending organized exercise classes more than twice a week and/or exercising >2 hours alone each week).

Paper III (66) consists of data from participants from the baseline and follow-up assessments from the follow-up assessment of the PreventIT feasibility RCT (see section 2.1 for details). First, an invitation letter was sent out to 2000 random people between the age of 61-70 at each site, with an even distribution of men/women. Those who contacted us were screened for eligibility via telephone. If eligible, they were invited to the hospital where they were subjected to a final screening. In short, the target group was healthy, community-dwelling seniors between 61-70 years of age whom had been retired for more than 6 months, were able to read a newspaper or text on a

smartphone, spoke Norwegian, German or Dutch, and could walk 500 m without walking aid. A complete overview of inclusion/exclusion criteria for each step is described in the protocol paper for the PreventIT feasibility RCT (67). All 189 participants that were included at baseline, and 134 out of 180 intervention participants, completed testing of the first and second version of the self-tests, thus included in the first part of Paper I.

In addition, we recruited a separate group of 20 community-dwelling adults ranging from 60 to 80 years to test the third version of self-tests (second part of Paper III). Inclusion criteria were as follows: community-dwelling, aged between 60 and 80 years, able to walk 500 m independently, speak Norwegian, able to hear sound from a smartphone, and current users of a smartphone. Participants were excluded if they reported any severe cardiovascular, pulmonary, neurological, or mental diseases.

Paper IV (68) included in total 60 participants, of whom 40 were community-dwelling seniors between the age of 60 and 86 years, and 20 were geriatric outpatients between the age of 61-85 years.

An overview of the participants' characteristics in each paper is presented in Table 2, except for the systematic review, for which participants' characteristics are not available.

Table 2. Characteristics of participants in the different papers.

Paper	Population	Age range	Mean age (SD)	Male (%)	n included
1	Healthy community-dwelling seniors aged	60-70			
2	Healthy community-dwelling seniors aged	60-70	66.4 (2.7)	12 (40)	30
3	Iteration 1: Healthy community-dwelling seniors	61-70	66.3 (2.4)	90 (47.4)	189
	Iteration 2: Healthy community-dwelling seniors	61-70	66.3 (2.5)	64 (47.8)	134
	Iteration 3: Healthy community-dwelling seniors	60-79	68.7 (5.2)	11 (55)	20
4	Group 1: Geriatric rehabilitation patients	61-85	78.9 (5.9)	11 (55)	20
	Group 2: Healthy community-dwelling seniors	60-86	71.8 (7.3)	17 (42.5)	40

2.3 Ethical approvals and considerations

Paper II, III, and IV had ethical approval prior to study start. The Regional Committees for Medical and Health Research Ethics (REC), the Ethik-Kommission am Universitätsklinikum Tübingen, and the Medical Ethical Committee, VU University Medical Center, all approved the study protocols for Paper II (Trondheim: REC Central 2016/48, Stuttgart: registration number 033/2016BO2, Amsterdam: registration number NL56456.029.16) and Paper III (Trondheim: REK midt, 2016/1891, Stuttgart: registration number 770/2016BO1, Amsterdam: METc VUmc registration number 2016.539, NL59977.029.16). Both studies, The PreventIT aLiFE feasibility study and the PreventIT feasibility RCT, were registered in ClinicalTrials.gov (Trial registration: ISRCTN37750605 and registration number: NCT03065088). For the Usability-study in the second part of Paper III, The Norwegian Center for Research Data approved that the data protection was in accordance with current regulations (reference number 391684). Paper IV was approved by the local medical ethical committee in Germany (no: 850/2018BO1).

All participants that responded to the invitations, met the inclusion criteria and agreed to participate in each of the studies, were given written and oral information about the studies and all assessments prior to signing the written consents. Assessors (Papers II, III, and IV) and aLiFE trainers (Paper II) were all trained and experienced, and the well-being of participants was prioritised throughout all assessments, for example allowing for breaks between assessments. All studies were conducted in accordance with the Declaration of Helsinki.

2.4 Development of app-based self-tests of physical function

We based our development of the stand-alone smartphone apps 'Self-TUG' and 'Self-Tandem' on prototypes initially developed by researchers at UNIBO for the FARSEEING-study (52). The Self-TUG and Self-Tandem let the user self-administer a clinical test of mobility and balance, respectively. As part of this PhD project, The Self-TUG and Self-Tandem were further developed and tested, in addition to a new app-based version of the clinical test 'Five times sit-to-stand' (Self-STS); the apps use the inertial sensors embedded into the smartphone. A detailed description of the apps, including user interface and underlying algorithms, are presented in the following. All apps were developed with Android Studios.

2.4.1 Self-TUG

The Self-TUG prototype consisted of a simple user interface consisting of three screens. When opening the app, the main screen appears, with a green start-button centred on a white background (Figure 1, left panel). If pressed, a new screen appears with a time countdown, and a stop-button (Figure 1, right panel). When the test is completed, a result screen appears (Figure 2).

The algorithms embedded in the Self-TUG automatically detects the postural transfers in a TUG trial from the sensor signals; sit-to-walk (StW), turning, turn-to-sit (TtS). After the user has manually pressed the start-button in the app, it does not start until the algorithm recognises that the user is ready (i.e. sitting still on the chair), which is determined by thresholds in the sensor signals.

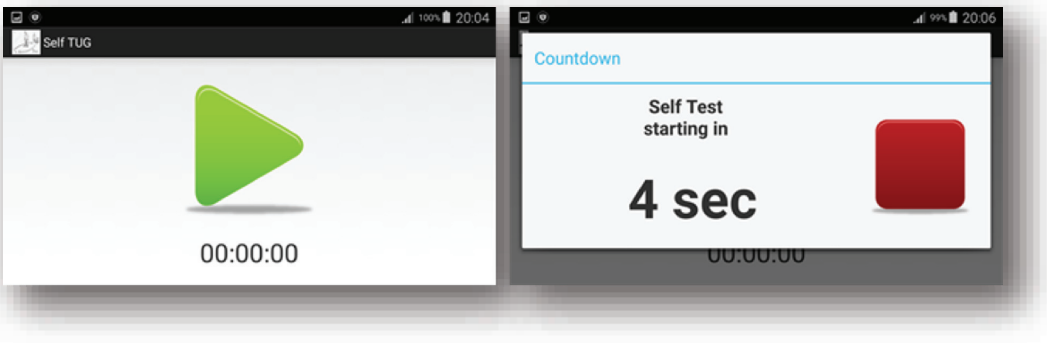

Figure 1. Screenshots from the prototype version of Self-TUG. Left: start button. Right: Countdown presented once start-button is pressed.

When the user is ready, i.e. no movement (above a certain threshold) is detected, a sound is generated by the app, indicating for the user that the test can be initiated. A selection mask is used to recognise when the user starts the task, i.e. starts to bend forward before rising from the chair, by processing the angular velocity around the medio-lateral axis and the anterior-posterior acceleration. This selection mask decides whether changes in the sensor signals corresponds to the postural transition or not with a binary "true" (one) or "false" (zero). When the final segment of a TUG is detected, i.e. sitting down again, another sound is generated, indicating that the test has been completed. When the test is completed, the app processes the recorded signals and present the results which are limited to the total duration of the test and the duration of the StW-phase in seconds (Figure 2). The algorithms for all iTUG phases have been described in detail by Sabato Mellone in his doctoral thesis (35).

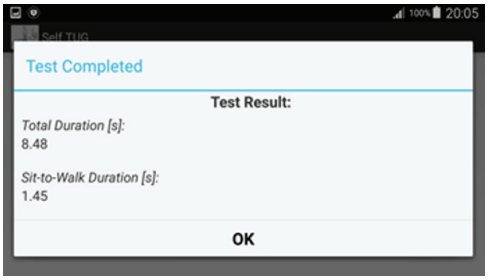

Figure 2. Results from a Self-TUG trial performed with the prototype version.

All sensor signals from a trial are automatically written to a text file which is saved on the internal memory of the phone, ready for off-line processing. Algorithm outputs for a representative iTUG trial is presented in Figure 3.

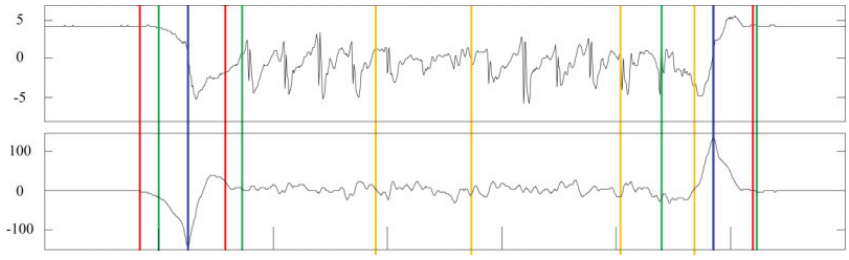

Figure 3. Sensor signals from a TUG trial with a wearable sensor attached to the lower back (35). The upper plot shows the antero-posterior component of the acceleration (m/s²) and the lower plot shows the angular velocity (°/s) around the medio-lateral axis. The green vertical lines define the selection masks. Blue vertical lines correspond to the absolute maximum/minimum of the angular velocity around the medio-lateral axis within the selection mask. Red vertical lines represent the start and beginning of StW and the end of TtS. The orange vertical lines are the beginning and end of the turning phases.

The prototype of Self-TUG was accompanied with a customised elastic belt with a transparent plastic pocket for inserting the smartphone into, and a paper manual describing step-by-step how to self-administer and perform the test. The procedures were to first insert the smartphone into the pocket, press the play-button, strap the belt around the waist so that the smartphone was located at the lower back with the screen facing out, and then to sit down on the chair and wait for the start-signal.

Version 1

The first version of the Self-TUG app was similar as the prototype, however the instruction manual was compressed and simplified into one page of instructions that explained only a few key points necessary to perform the test correctly.

Version 2

For the second version of the Self-TUG app, we implemented an instructional video with a voiceover into the app, demonstrating step-by-step how to perform the self-test. The algorithm was modified to decrease the time between the point where no movement is detected and the start signal. The placement of the smartphone was changed from lower back to the front pocket. The new placement eliminates the need for the custom waist belt, so that deployment is simplified for commercial applications too, and makes the self-administration procedure more user-friendly. The app interface was kept as it was in the prototype with two exceptions. One was the navigation flow for how to

watch the instruction video and/or start the test, which is illustrated in the flowchart, Figure 4. In addition, we added a short reminder of how to proceed with the task to the countdown panel, e.g. “Put the phone gently in your pocket”, presented in the lower right screenshot in Figure 4.

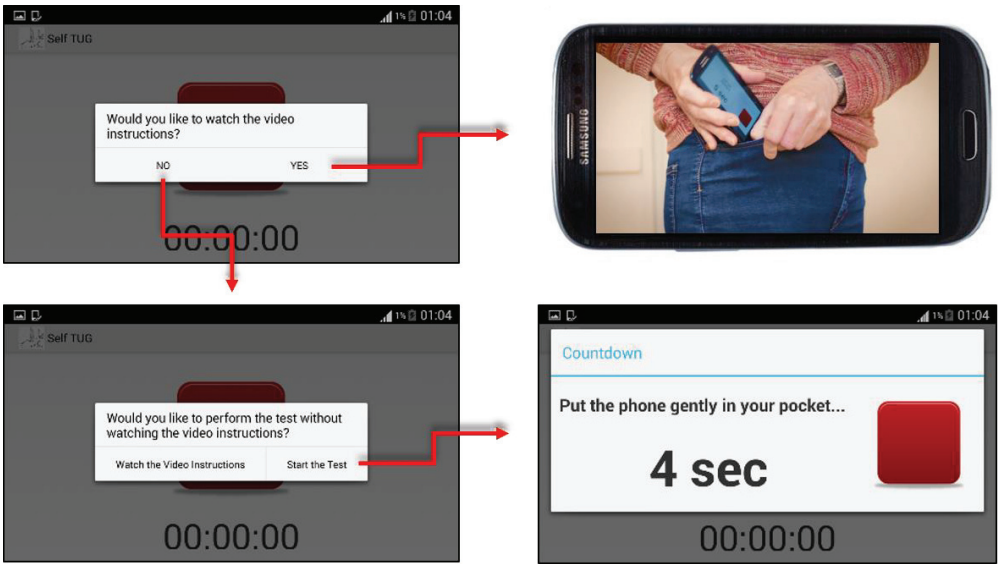

Figure 4. Flowchart illustrating the navigation options in version 2 of the Self-TUG app.

Version 3

The third version of the Self-TUG app was designed with the aim of preparing for a home-based user context. This meant the user had to make necessary preparations themselves, including choosing an appropriate chair, measuring a straight 3-meter walkway, placing a small object at the end of the walkway, and making sure there is enough space to perform the test unhindered. As such, instructions for how to prepare the tests were added to the apps in the form of a list of bullet points. Another hierarchy was thus added to the app interface, as illustrated in Figure 5. In addition, the countdown-screen was changed by replacing the text reminder with two images illustrating what to do. The timer countdown was replaced with a “throbber”, an animated graphical element which we added to symbolise that the test had started and further action was required, but without creating a feeling of time pressure, which some users felt when they were presented with the timer countdown in version 1 and 2. These changes are shown in the center right screenshot of Figure 5.

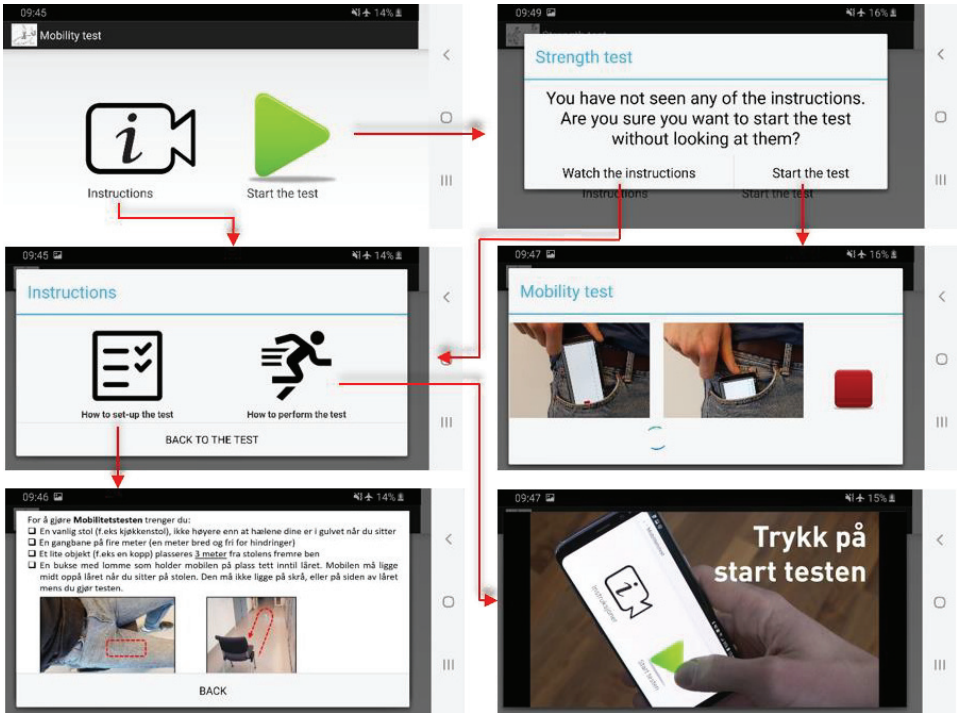

Figure 5. Flowchart illustrating the navigation options in version 3 of the Self-TUG app.

In addition, the instruction videos were edited to include the changes made to the navigation flow, text and graphical elements were added to emphasise key points of the task, see Figure 6. Due to time constraints in the development phase, the text and new voiceovers were only made in Norwegian language, as all participants in the home-based usability study were Norwegian.

Real-time verbal feedback based on the inertial signals from the smartphone, e.g. “sit down”, “place your arms on the armchair”, “get up from the chair”, and “proceed with the test”.

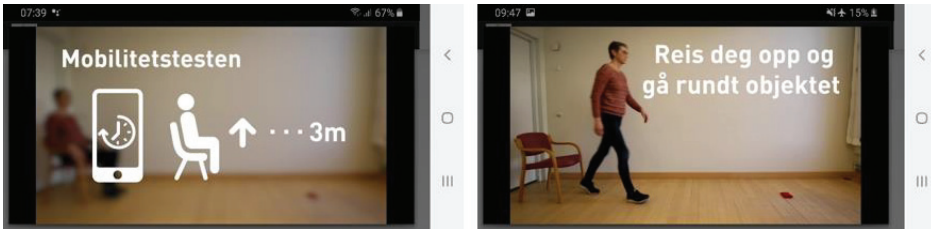

Figure 6. Text and graphical elements in instruction video for Self-TUG version 3.

2.4.2 Self-Tandem

The Self-Tandem is an app that lets the user self-administer a test of timed, static posturography in a tandem standing position. In this position, one foot is placed in front of the other, heel touching toe, without shoes. The user holds the smartphone against the chest with both hands while trying to stand as still as possible for 15 seconds.

The Self-Tandem prototype was similar to Self-TUG in user interface and functionality; the main difference was that after the user pressed "start", it did not rely on signal input for detecting that the user was ready. Instead, the test started immediately following the countdown, indicated by a start signal, and lasted for a duration of 15s, after which an end signal was played.

The Self-Tandem app records the inertial sensor signals during the task and quantifies the body sway. With a set-up in which the smartphone is placed against the chest, the assessment is quick and easy from a user's perspective, as nothing else but the smartphone is required to do the test. The type of features that can be reliably measured from this set-up is however limited, since neither the placement of the feet and smartphone, nor the correct orientation of the smartphone, can be verified. There are, however, some features that do not require a distinction between the anterior-posterior and medio-lateral axes, and could thus be considered more robust with this set-up. One is "Sway area", estimated as the sum of the triangles formed by two consecutive points on the sway trajectory on the horizontal plane and the mean point on the plane, and "Ellipse area", which is the area of the confidence ellipse enclosing 95% of the points on the sway trajectory (35). These parameters are defined by analogy with the COP displacement as measured with a force plate.

Version 1 and 2

Changes to the Self-Tandem app, between the prototype and version 1, were the same as for Self-TUG app, namely a shortened manual describing how to perform the self-test. For version 2 of the Self-Tandem app, we added an instruction video in the app and added a short text reminder to the countdown-screen saying "Hold the phone against your chest…", see Figure 7. In addition, we changed the test condition from one where the participant should keep their eyes closed while balancing, to one where they could keep their eyes open.

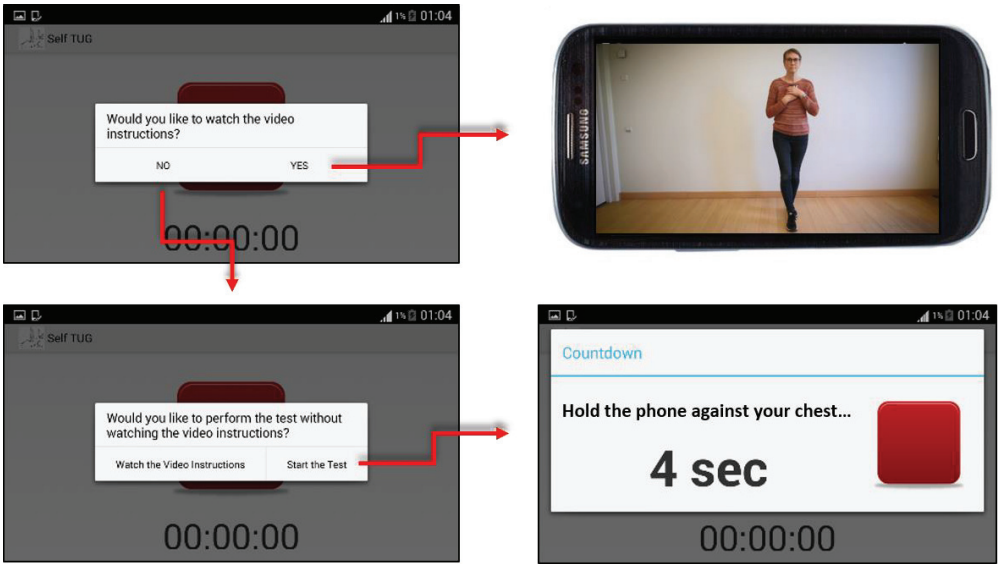

Figure 7. Flowchart illustrating the navigation options in version 2 of the Self-Tandem app.

Version 3

We made the same changes for the third version of Self-Tandem app as with the Self-TUG app, including new instruction videos and a set-up instruction menu.

2.4.3 Self-STS

As no prototype of an app-based test of functional strength was available, the Self-STS was developed as part of this PhD-project and included during the usability testing during the PreventIT follow-up assessment.

Version 1

The Self-STS app allows the user to self-administer a 'Five times sit-to-stand'-test. In terms of user interface and functionality, the Self-STS app was similar to version 2 of the Self-TUG and Self-Tandem apps, with a simple navigation flow (Figure 7) and a video instruction embedded within the app (Figure 8).

Figure 8. Screenshot of the Self-STS instruction video version 1.

The algorithms for detecting the user's readiness to perform the test was the same as in the Self-TUG app, as the task in both tests required starting and ending in a still, seated position on a chair. After starting the test, the user would insert the phone gently into the pocket, sit down and fold their arms across their chest, and wait for the start signal. After the start signal the user would stand up and sit down again repeatedly for five times. The end signal was generated when the participant was sitting still after performing the five repetitions. The outcome was the duration from when the sensors first detected that the participant rose from the chair until the participant sat down again for the last time. Since usability was in focus, we did not develop or validate specific algorithms and feature extraction procedures from individual repetitions of the sit-to-stand.

Version 2

Version 2 of the Self-STS app was tested in the home-based usability testing, corresponding to version 3 of the Self-TUG and Self-Tandem. Similar adjustments were therefore made to the Self-STS in the preparation for this usability study, including instruction for test set-up and new video instruction (Figure 9).

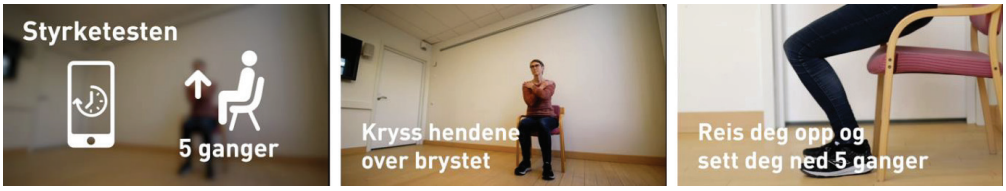

Figure 9. Screenshots of the Self-STS version 2 video instructions.

In addition we added a real-time verbal feedback based on the inertial sensor signals from the smartphone, e.g. "sit down", "cross your arms in front of your chest", and also a real-time counting of the repetitions, based on postural transfers detected by magneto-inertial sensor signals, to help the participant keep track of how many repetitions they have performed.

2.5 Usability testing

The usability testing conducted in Paper III took place both in a lab and in a home setting. Before the first versions were ready, three volunteers were invited to pilot the prototype of the Self-TUG and Self-Tandem at a workshop held in Trondheim. The volunteers received the smartphone and manual, and a chair and a TUG walkway had been prepared beforehand. The results indicated that there was a need to simplify the instructions, and this work was done before version 1 of the Self-TUG and Self-Tandem apps were ready to be tested.

The usability testing of version 1 and 2 of Self-TUG and Self Tandem, in addition to version 1 of Self-STS, was conducted as part of the clinical assessments in the PreventIT-study. The tests were performed in physiological test labs in Trondheim, Stuttgart and Amsterdam. The set-up of the tests, i.e. the chair, walkway and written instructions, was prepared beforehand by the assessors. Assessors verbally informed the participants about the test before starting, what the purpose of a usability test is, they could stop whenever they wanted and that the supervisor could not provide any help. Assessors observed while the participant attempted to self-administer the tests and took notes of what happened. The same procedures were used during the next usability testing, one year later.

For the usability testing conducted in the participants' homes, the participants wore a chest-mounted GoPro-camera, in addition to the GoPro camera recording the entire space where tests were performed. The participants were, in accordance with the instructions, supposed to prepare the test set-up themselves, including choosing where in their house they should perform the tests. The video recordings allowed for detailed analysis of how the participants prepared the test set-up, how they used the apps and how they performed the tests. Following testing, qualitative data was collected from questionnaires and interviews.

2.6 Outcome measures and statistical analyses

This chapter describes the most important outcome measures in each of the four papers included in this thesis. Statistical analyses are described below the outcomes for which it is relevant.

2.6.1 Paper I: Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review

The systematic review in Paper I included two separate searches in Medline and Embase; the first was conducted in June 2016 to identify performance-based tests of strength and balance, and the other in December 2017, with the aim being to identify studies in which the measurement

properties of the most used tests were evaluated. The abstracts were screened by two assessors, and eligible full-text articles were later analysed by two reviewers.

Most frequently identified performance-based tests of strength and/or balance in young seniors

To identify the most commonly used performance-based tests of strength and/or balance in young seniors, we counted how many times each of the 105 identified tests had been used in a study. Those with ≥3 citations were included in step two of the systematic review.

Consensus-based Standards for selection of health Measurement Instruments (COSMIN)

The quality of method studies included in step 2 of the review was evaluated by use of the COSMIN checklist (69). COSMIN describes how to rate the quality of the following nine categories of measurement properties: internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity, and responsiveness, with several items within each category.

2.6.2 Paper II: The Adapted Lifestyle-Integrated Functional Exercise Program for Preventing Functional Decline in Young Seniors: Development and Initial Evaluation

In Paper II we collected outcome measures of feasibility both during and after the intervention, as well as exploratory pre-post measures.

Feasibility measures obtained during the intervention:

Participants implemented up to four activities during each home visit, completing a total of maximum 16 activities during the 4-week intervention period. We used a activity planner and a counting sheet (70), where the participants reported on their performance. **Adherence** was defined as the number of activities implemented during the intervention, and "Implemented" was defined as reporting performing an activity at least once per week. We defined **frequency of practice** as weekly practice of each activity from the daily practice documented as activity episodes. For instance, tandem walking along the hallway would be one episode.

We defined **adverse events** in this study as self-reported pain, falls and injuries during the intervention period. Participants were asked to report all adverse events to the trainers.

The **acceptability of the activities** was defined as reported perceived helpfulness for improving strength, balance, and PA; adaptability to personal routine, and safety of practice documented by 7-point Likert-scale items developed for the study, administered by the trainers during each of the four home visits.

The observed **task challenge** of specific aLiFE activities was documented on a form used to record starting levels for each activity. The trainer would assess the participant's ability in each activity in order to set a proper starting level, ranging from 1-4, from which the participant could progress (see (64), online suppl. Table 2). Being unable to perform an activity at the lowest level indicated a floor effect, conversely performance at the highest level indicated ceiling effect.

Feasibility measures obtained after the intervention:

The **acceptability of aLiFE** was defined as overall reported acceptability, perceived helpfulness, adaptability, level of difficulty, and safety documented using 7-point Likert-scale (see (64), online suppl. Table 4). Participants were also asked the open-ended question "Please explain why you scored in this way and suggest any changes to the program" and answers were documented. Focus groups were conducted to collect further information about the aLiFE intervention with results reported elsewhere (71). We documented **activity preferences** with the question: "Please name your 3 favourite aLiFE activities."

Exploratory pre-post measures:

Physical function was measured with the CBMS. **Physical activity (PA) levels** was measured continuously in the week prior (week 0) and during (week 3) the intervention, using a sensor worn on the lower back (DynaPort MoveMonitor, McRoberts, Netherlands). Percentage of sedentary time (i.e., energy expenditure ≤1.5 MET) and walking time were extracted from raw data using validated algorithms (72). The pre-post changes in CBMS and PA-levels were explored with the Wilcoxon signed-rank test. Mean difference, confidence intervals, and Cohen's d effect sizes were calculated. Effects were interpreted as small (d = 0.2), medium (d = 0.5), and large (d = 0.8), p value ≤0.05 was accepted. The analysis was performed using SPSS 22.0 (IBM, Armonk, NY, USA).

2.6.3 Paper III: App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study

In paper III we collected outcome measures of usability during testing of the apps.

We defined **usability errors** as deviations from the test instructions and counted the number of errors from the clinical record forms in the first and second iterations and from video recordings in the third iteration. **Usability problem categories** were defined based on a subjective interpretation of what caused the errors. We defined the **proportion of correctly performed tests** as the first trial without any errors made by the participants.

User experience was collected as the participants' experience with using the app-based self-tests in the third iteration using a questionnaire consisting of 6 items relevant to the apps. Each item was

scored on a 5-point Likert scale. Frequencies of responses within each category across all items were calculated.

The **System Usability Scale (SUS)** (73) is a 5-point Likert-scale consisting of 10 items, providing a global view of subjective assessments of usability. The scale has been found valid for comparing the usability of two similar systems, or two different versions of the same system (74). Scores were added for each participant and multiplied by 2.5 to get a usability score ranging from 0 to 100 (with a higher score indicating better usability).

A semi-structured interview was conducted following the usability testing of the third iteration. The interview transcripts were analysed using **thematic analysis** (75) to identify relevant themes. Quotes were extracted for each subtheme and translated from Norwegian to English for analysis. The questions presented to the participants were "What did you think about using these apps to test your physical function?" and "Do you have any ideas for how the apps can be improved?"

2.6.4 Paper IV: Predicting advanced balance ability and mobility with an instrumented Timed Up and Go-test

In Paper IV, inertial sensor signal data recorded during the iTUG trials were obtained from the smartphone and processed in Matlab (Mathworks, MA, US). The algorithms used to separate the different segments of the TUG (sit-to-stand, walking, turning), and calculate signal features from these segments, have been described elsewhere (46, 53). Signal features were computed from the five separate iTUG trials from all participants and used as *predictor variables (X)* in model 1 of a Partial Least Squares regression (PLSR) analysis. The CBMS score of each participant was used as the *response variable (Y).* In model 2, the Standard clinical tests were used as predictor variables. The PLSR models were validated in a 7-step cross-validation procedure. The iTUG features (model 1) or Standard clinical tests (model 2) that were significantly ($p < 0.05$) correlated with Xtrain and Ytrain were selected for the PLSR model. Data were then cross-validated to identify the robust latent variables (corresponding to components in Principle Component Analysis) without overfitting the model, using a Monte Carlo simulation procedure with 100 repeated random repartitioning of Xtrain and Ytrain.

The **R squared (R^2)** was obtained from the PLSR analysis as a measure of how much of the variation in the CBMS score could be explained by the two models we wanted to compare (iTUG features versus Standard clinical tests), and thus how well the models could predict the CBMS score.

The **Root Mean Square Error of Prediction (RMSEP)** was calculated to find how much error was associated with the level of prediction found for the two models compared with PLSR. Z-scores were

obtained to test the difference in RMSEP between the model with iTUG features and Standard clinical tests.

3. Summary of results

A short description of results from Paper I-IV is presented in the following. The interpretation of the results will be discussed in detail in the discussion chapter of this thesis.

3.1 Paper I: Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review

This paper presents tests of balance and strength commonly used in young seniors aged 60–70 years and the measurement properties of the tests most frequently used. Abstracts from 3454 papers were evaluated, and 295 papers were included for the full text review. Results showed 120 unique tests or versions of tests, of which 69 were tests of balance and 51 were tests of muscle strength. We identified the following domains of balance tests: static steady-state balance tests (28 tests), dynamic steady-state balance tests (14 tests), proactive balance (8 tests), reactive balance (7 tests), and performance test batteries (9 batteries). For strength measurements, we identified the following domains: one-repetition-maximum tests (6), maximum isometric strength tests (9), and muscle power tests (36).

Out of the 120 strength and balance tests we identified, a total of 26 balance tests and 15 strength tests were cited in ≥3 papers and were thus included in the evaluation of measurement properties. Abstracts from 1880 papers were evaluated, and 9 papers were included for the full text review in step two of the systematic review. Results identified the measurement properties of the '10s Tandem stance', 'TUG', 'SPPB', 'CBMS' and the 'Five times sit-to-stand'.

An important finding was that very few of all tests available for measuring balance and/or strength have been assessed for their measurement properties in healthy young seniors. The quality of most of the method studies rated in this review ranged only from "poor" to "fair". Based on the findings in this review, the CBMS seems to be the only suitable scale for adequately assessing balance in healthy young and higher-functioning seniors, as it showed no ceiling effects and was found valid and reliable.

3.2 Paper II: The Adapted Lifestyle-Integrated Functional Exercise Program for Preventing Functional Decline in Young Seniors: Development and Initial Evaluation

Paper II presents a study in which the feasibility of the aLiFE programme was evaluated, when delivered to 30 young seniors in Trondheim, Stuttgart and Amsterdam. The measures of adherence and frequency showed that participants implemented on average 12 out of 16 activities during the 4-week intervention. Most frequently implemented were the 'sit-to-stand' for the strength module, 'one-leg stand' and 'stepping over objects' for the neuromotor module, and 'break up sitting' for the

PA module; while 'toe standing', 'stepping and changing direction', and 'walk faster' were least frequently implemented.

In measures of acceptability, the participants perceived activities from all three domains as helpful for achieving their intended outcomes. Most activities were reported as easy to implement, and all activities were perceived safe to perform. The participants seemed to like the program, and all except one reported that they would recommend the programme to a friend. We found that there was less consensus on how easy it was to adapt the activities to one's progression.

For the exploratory measures of pre-post changes, we found that the CBMS increased with a medium effect size (d = 0.72, p = 0.001). Time spent walking also increased (d = 0.36) and time spent sitting decreased (d = -0.10), albeit not significantly.

3.3 Paper III: App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study

Paper III describes the development and usability evaluation of the Self-TUG, Self-Tandem and Self-STS, through three iterations. In the first iteration, 42 (22%) of the participants performed the Self-TUG correctly, and 127 (67.2%) performed the Self-Tandem correctly. One or more errors were made in 120 of the 378 (32%) trials, in which forgetting or misunderstanding the written instructions were the most common causes of errors.

In the second iteration, 108 (83.1%) performed the Self-TUG correctly, 106 (79.1%) and 40 (30.1%) the Self-Tandem and Self-STS, respectively. Errors caused by the usability problems in the category labelled *'incorrect performance of test'* were made in 66 of 402 trials (16%). Percentage of errors caused by *'performs test without starting app'* and *'did not sit still and wait for start signal after test was started'* increased from the first usability test, while the frequency of *'incorrect placement of phone'*, *'did not hear/perceive instructions'*, and *'accidentally cancelled the test'* decreased.

In the third iteration, 14 (70%), 18 (90%) and 5 (25%) performed correctly the Self-TUG, Self-Tandem and Self-STS, respectively. Errors caused by the usability problem *'incorrect performance of test'* were made in 19 of 60 (32%) test trials. We completely reduced the usability problems *'did not sit still and wait for start signal after test was started'* and *'incorrect placement of phone'*, while the frequencies of *'performs test without starting app'*, *'did not hear/perceive instructions'*, and *'accidentally cancels the test'* remained as in the previous iteration.

Perceived user experience of the apps was rated "positive" or "very positive", and the mean score on the system usability scale was 77.63 (SD 16.1 points, range 42.5-97.4 points).

3.4 Paper IV: Predicting advanced balance ability and mobility with an instrumented Timed Up and Go-test

The PLSR analysis in paper 4 was based on data from a total of 60 community-dwelling seniors and geriatric outpatients. Results showed that a model with 3 components containing 23 iTUG features and two descriptive variables (age and education), explained the variation in CBMS scores with an R^2 of 0.852 (95% CI 0.849-0.855), and a RMSEP of 11.81 points. All except one iTUG feature had a VIP >0.83 and <1.21. Fourteen of the other 22 iTUG features had a VIP >1. Six out of the 10 features with highest R^2s were features computed from the turning segments of iTUG.

For the standard clinical tests, the PLSR analysis resulted in an explained variation in CBMS of R^2 = 0.825 (95% CI 0.82-0.83). A two-component model showed a RMSEP of 12.85 points, which was significantly higher compared to RMSEP of the iTUG-model (p=<0.0001). This model kept the following tests: TUG, gait speed (fast and habitual), SPPB, 30-CST, Short-FES-I, and the 8-LBS, in addition to the descriptive variables “age” and “education”. Weight, BMI, height, and the MoCA score were excluded in the PLSR procedure.

Results together showed that the iTUG features were slightly more accurate than standard clinical tests in predicting CBMS scores, suggesting that an iTUG, which is fast and easy to administer, can potentially be used to predict a person’s score on the CBMS in research and clinical care.

4. Discussion

4.1 Main results

In summary, the systematic review identified 120 performance-based tests of physical function used in younger seniors, of which 69 measured balance and 51 muscle strength. The TUG, SPPB and BBS were the most cited tests in the included papers. We identified only 9 method studies evaluating the measurement properties of 5 different tests, of which the CBMS seems most promising for use in young seniors, being valid, reliable and not prone to ceiling effects.

The aLiFE programme was found to be well accepted by the participants overall; they perceived activities from all three domains as helpful for achieving their intended outcomes and all but one reported they would have recommended the programme to a friend, some participants found it challenging to adapt the activities to one's progression. Moderate effect sizes were found for changes in physical function and levels of physical activity, measured before and after the 4-weeks study period.

The usability study found that measures of usability overall were high, and people reported having high levels of user experience. The usability problems identified were reduced throughout the iterations, and consequentially a higher proportion of participants performed the tests correctly. Conducting usability in a home setting introduced some new usability problems should be addressed in future research before the apps can be made available to end-users.

In the method study evaluating the predictive ability of the iTUG, we found that inertial sensor-signals from a smartphone worn during a 5-times repeated iTUG predicted the CBMS score in geriatric outpatients and community-dwellers with an accuracy of 85.2%. The accuracy was similar to that found with standard clinical tests, suggesting an important implication for saving time and resources when assessing physical function in seniors.

A general discussion of the main findings will be given below, followed by a discussion of the methodological issues with respect to the work conducted in this thesis, what implications the results could have, and future recommendations.

4.2 Discussion of results

The interpretation of the most important results of the individual papers included in the thesis will be discussed in the following sections.

4.2.1 Performance-based tests of strength and balance in young seniors

With increased focus on prevention of functional decline there is a need for functional tests that are suited to the target group of young seniors. Before conducting the systematic review, it was not clear whether such tests existed. Overall, there seems to be very little consensus on how to best assess physical function in young seniors, indicated both by the high number of tests identified in general (120 tests), and the high number of variations for many of the tests. As an example, we identified 13 different versions of the one-legged standing balance test (54, 76-86), and 6 different versions of the sit-to-stand test (87-92). The most important finding however was that among all the tests which were identified, only 9 studies had evaluated the measurement properties of them in young seniors (93-101). The findings suggest that the tests probably weren't intended for the target group to begin with. The tests assessed in method studies included in our review, we also know were originally designed for other groups or specific conditions, e.g. the SPPB for assessing lower-extremity functioning of older adults (90), the TUG for assessing older adults at risk of falling (17), the CBMS for assessing higher-level balance in children with traumatic brain injury (102), and the sharpened Romberg (10s Tandem standing) for assessing ataxia (103).

Healthy people aged 60-70 years are a relatively new group to study physical performance in, and the findings confirm that, accordingly, there is a lack of tests designed specifically for them. However, that does not imply that already existing tests are inappropriate. As the review showed, the CBMS is a test designed for a completely different purpose, but still seems like a promising test battery to use in this group, due to its more challenging tasks. The TUG was the most used test in younger seniors, and 3 of the 9 method studies included in the review had evaluated measurement properties of the TUG. The results of these method studies did not, however, suggest that the original TUG is a suitable test of physical function in young seniors. The iTUG has been suggested in previously in the literature (19, 38, 39, 53, 104) as a more sensitive test than the TUG, but no method studies evaluating the iTUG were included in the review, probably due to not meeting our inclusion criteria.

4.2.2 Feasibility of the aLiFE exercise intervention

Although effective in the short-term, traditional exercise programmes have failed to demonstrate long-term adherence, and they are also not found sufficiently appealing by some seniors (28, 71). Regular physical activity, strength and balance exercises are important for preventing functional decline; therefore, we adapted the LiFE-programme so that it would be more relevant and useful for a younger target group by adding new activities, added more challenging levels to new and existing activities, and a physical activity module to increase walking and reduce sedentariness. We found that the participants liked the concept of integrating activities in their daily tasks and activities, and

they liked most of the activities that they could choose from. In line with the original LiFE programme (29), the participants were offered up to 4 activities per visit but decided themselves how many to implement. Our aLiFE programme consisted of 16 activities in total, and the participants integrated on average 12.1 ± 1.8 activities during the intervention. Participants reported that the short duration of the programme made it challenging to integrate more activities. Trying out many activities throughout the 4-weeks period was encouraged, so that we would get feedback on as many activities as possible. With a longer intervention period, the participants could perhaps have had more time to get familiar with the activities and get a better opportunity to find tasks and routines that the activities would integrate more seamlessly with. However, weekly frequency of practice in the aLiFE study were comparable to previous research in an older cohort (105), suggesting that this type of exercise programme can be applied in young seniors.

Whilst it was not the aim of the study, moderate effect sizes were found for changes in CBMS (d=0.72), which is in line with findings of change in balance performance in interventions with similar length (106). For pre- and post-measurements of daily walking duration, a nonsignificant, moderate change was found (d=0.36). Why it was nonsignificant could possibly be ascribed to several causes. One reason could be that the physical activity (PA)- module which we developed for aLiFE was inadequate for inducing a noticeable increase in PA. As an example, the activities in the 'Reduce sedentariness'-principle, like 'Sit less', might be more challenging to stay conscious of throughout the day and thus more difficult to integrate because they are more unspecific as compared to the some of the balance and strength activities, like standing on one foot or doing calf raises which can be linked to specific situations, such as brushing teeth and preparing food by the kitchen bench. Furthermore, the sample was quite active to begin with, as shown by a 9% proportion of time spent walking at baseline (64, 107), and increasing PA beyond this might require either adding more activities for increasing PA or modifying the existing activities.

Finally, the study was not designed to establish the effecteffectiveness of aLiFE. As such the duration of the programme could have been too short for the participants to properly integrate the activities in their lives. Studies have indicated that up to 6 months may be required to induce behavioural change (108, 109). In addition, the sample size might be too small. To establish the effectiveness of aLiFE on physical function and PA, a larger and longer-lasting study designed to establish effectiveness is needed.

4.2.3 Development and usability-testing of self-test apps

Paper III described the development and usability testing of the self-test apps. We identified several usability problems with the apps that caused participants to perform the tests incorrectly. The

solutions we implemented to improve the apps seemed to work for most of the usability problems, as they decreased from the first to the second usability testing, of version 1 and version 2. New changes were made to further improve and prepare the apps to be feasible for use in a home setting. This meant that instructions needed to be clearer, not only for how to perform the test, but also how to prepare it, as the set-up had been prepared by the assessors beforehand in the lab-based test sessions. The real-time counting of repetitions that we implemented in the Self-STS turned out to confuse more than help participants to perform the tests correctly, and if the incorrect trials caused by time counting are ignored, the results indicate high completion rates. An important point to add in this context is that in the usability testing, we only considered the first trial when calculating the completion rates. What the results show are thus the usability of the tests when performed for the first time.

The response on the UEQ and on the relevant questions in the interview also points to an overall satisfactory experience with the apps among the participants. In addition, the mean score on the SUS scale was 77.63 (16.1), which is above average and correspond to an adjective rating between "good" and "excellent" (110). Interestingly, the score is almost identical to those reported in another usability study of self-tests of balance and fall risk, assessed with a smartphone (SUS score 79.17) and tablet (77.92) in older adults (60). SUS and UEQ data were however not collected during iteration 1 and 2, preventing us from analysing the change in these outcomes throughout the entire development, and from studying the relationship between SUS and UEQ and completion rates. Previous research have demonstrated that while the number of usability problems encountered by participants are reflected by the SUS scores, the completion rates might not (110). The cause of this observation is unclear, but it nonetheless points to the importance of assessing both subjectively experiences usability and user experience (SUS and UEQ), as well as objectively measured completion rates.

4.2.4 The iTUG as a test of advanced balance ability test

The systematic review described in Paper I identified the CBMS as a valid and reliable test of advanced balance ability in healthy young seniors (16, 20, 101), and it can thus be considered a current gold standard. The applicability of the CBMS is, however, limited for several reasons, including the large space it requires, being time consuming, the equipment needed, and training required for assessors. For these reasons, we wanted to investigate how well the smartphone-based iTUG could predict the CBMS score. We found that with a machine-learning analysis of iTUG features and descriptive variables, the CBMS could be predicted with a high accuracy. The performance of the iTUG prediction model (R^2 0.852) was comparable to a model of standard clinical tests (R^2 0.825), although slightly higher with the iTUG, and significantly less error of prediction. What is most

interesting with the results is the high prediction level achieved with iTUG despite it being such a simple test. In contrast, the standard clinical tests were expected to achieve a high prediction level, because of the inclusion of several tests, such as the SPPB, Eight level balance scale, 7-meter fast walking, and because these tests consist of elements that resemble the tasks in the CBMS. The iTUG however, is from a broad perspective a simpler motor and balance task where you get up, walk, turn around, walk back and sit down again, in a comfortable pace. Still, the features made available from the sensors and the algorithms processing their signals, let you accurately quantify the subtle differences people express when performing these sub-tasks. In another study with a similar aim and set-up as in our study, but using the MiniBESTest (111) as response variable and not the CBMS, it was found that the iTUG explained only 41% of the variation in the Mini-BESTest scores (40). The Mini-BESTest is a test battery consisting of 14 tasks, and the aim is to target and identify 6 different balance control systems (111). Many of the balance control systems measured with the Mini-BESTest are similar to those in CBMS, such as static and reactive postural control, dynamic gait, and dual tasking. This study included twice as many participants as in our study, and they were older and diagnosed with a neurological condition, making the studies not entirely comparable. Nonetheless, it is somewhat surprising that such a large difference was observed in the explained variation of the outcome measure by the iTUG between these two studies. The observed difference warrants further investigation to verify our findings.

In summary, it was interesting to find that the iTUG could predict a challenging and advanced test battery such as the CBMS, especially given the possible implications associated with it. A valid and reliable PLSR model of iTUG integrated into a smartphone app could serve as a simple yet valuable part of a clinicians' toolkit, or within a Self-TUG app where people can perform the iTUG in their own house, with or without remote evaluation from a clinician.

4.3 Methodological considerations

Methodological considerations of the work conducted in this thesis will be discussed in the following sections, including the evaluation of measurement properties in the systematic review, a discussion of the presence of bias in the participant sampling for this thesis, the validity of the smartphone-based self-tests, and, finally, the methods used to develop and test the apps.

4.3.1 Systematic review for identifying appropriate functional assessments

The most commonly used performance-based tests of physical function in older adults have been found inappropriate for use in young seniors (18, 20, 21). With the systematic review we wanted to investigate whether there were any performance-based tests cited in the literature, whose measurement properties had been reported in young seniors. Using a systematic review to achieve

this goal was probably adequate. In contrast to other study designs and other literature studies, the systematic review can provide a complete, exhaustive summary of current evidence, that is "methodical, comprehensive, transparent and replicable" (112).

The study quality was also strengthened by using the COSMIN checklist (69) to evaluate the quality of the studies reporting the measurement properties, as the quality rating enables the reader to know which results should be interpreted with caution. The COSMIN checklist does not, however, account for the results reported in the method studies, which can contribute to some confusion in the interpretation of the results of the systematic review. As an example, the original SPPB is a test battery known to have ceiling effects when administered in higher-functioning seniors, even in those up to the ages of 70-80 years (20, 21, 113). In our review, we identified one method study that had evaluated the measurement properties of the SPPB in young seniors, and according to the COSMIN checklist, the study was rated "Excellent" for its evaluation of construct and concurrent validity (95). The rating is based on the way in which the study was designed and carried out, and not the results, because there was a risk of ceiling effects even among the participants in that study. The same applies to the TUG, which was evaluated in 3 out of 9 method studies (93, 97, 98). This is briefly discussed in the systematic review. As an alternative, we could have compiled all test results from the target group reported in the studies identified in step 1 of the review. Compiling all test results would have provided a large pool of data on which we could have performed a meta-analysis to assess risk of ceiling/floor effects ourselves, although, calculating ceiling/floor effects would require the scores of each subject to calculate how many of the total sample achieved minimum or maximum scores, and papers rarely report the individual scores of participants, but rather describe the central tendency and dispersion of the group(s). Our systematic review did however draw on previous research to point out that although the method studies evaluating tests such as TUG and SPPB received high rating on the COSMIN scale, they were not necessarily appropriate in our target group (16).

Despite the limitations discussed here, the systematic review provided an exhaustive summary of the current literature indicating which performance-based tests that have been used in young seniors, and review highlighted that few have been evaluated for their measurement properties in this target group.

4.3.2 Participants and sampling bias

In this thesis, the main target population was young older seniors at time of retirement. The participants recruited for the studies were however recruited for each particular study, with different goals, and hence the recruitment strategies differed. For study results to be generalisable,

study participants need to represent the general population for which you aim to generalise the results. If not, there's a risk sampling bias, which can affect the validity and reliability of the study findings (114). The samples included in our studies likely differ in how representative they are of the target populations for the respective studies.

In the PreventIT-study, in which the first two usability tests of the apps were performed, the participants were community-dwelling seniors from Trondheim, Stuttgart, and Amsterdam, who had been randomly picked from registries to receive an invitation letter, which by itself is a good strategy to ensure a representative sample. However, privacy regulations required that we could not contact the participants until they have contacted us, thus there is chance of a non-response bias having occurred. A non-response bias is related to the sampling bias, and is observed when participants in voluntary samples are different with respect to gender, age, education levels and health status, compared to mandatory samples (115).

The aim of the usability tests conducted with this sample was to identify any usability issues associated with the first and second version of the apps. Whether sampling bias affected the validity of this study or not is uncertain. Sampling bias in this case would imply that the usability issues identified when testing this sample are systematically different from those we would have found in the target end-user group, for example that our sample consisted of participants that in general were more experienced and proficient users of smartphone apps as compared to the target group at large. A strong argument in this context is however that the aim was to identify usability issues, for which studies have reported that a sample size of 3-20 participants typically is valid (116). As we included 189 and 134 participants for the testing in iteration 1 and 2, respectively, it is therefore unlikely that we would have identified different usability issues with other participants from the target group.

In the home-based usability-study we used convenience sampling to recruit 20 participants from Trondheim. A convenience sample is a non-probability sampling method, where participants are invited because they fit some criteria, such as being available, in geographical proximity, and are willing to participate (117). The participants' age in the home-based usability study was more spread compared to the usability tests in PreventIT, as we included 20 persons aged between 60 and 80 years. By including participants at a higher age, we get participants where reduced visual acuity, hearing sensitivity and short-term memory may be more prevalent (118). These are factors which may impact how a person interacts with the smartphone, and in a usability testing context, it strengthens the study as we gain insight of usability problems which we otherwise may not have acquired. As this was a convenience sample, there is always a risk that these participants are

different from those people who would not have volunteered to participate, which would limit the generalisability of the results. However, it is likely that the two samples in which the apps were tested, together represent a wider variety with regards to the user target group for which the apps are intended. Future user-testing cycles should, however, increase the chance that any remaining (or new) usability problems are identified.

A convenience sample was included also in the aLiFE study consisting of 30 participants from Trondheim, Stuttgart and Amsterdam. The scores on the functional tests and the activity monitoring recorded at baseline indicated that it was a relatively fit group of seniors that had volunteered (64, 107). One participant had dropped out after baseline assessment due to the programme focusing too much on lower-body activities, and some participants in the Norwegian cohort reported that when volunteering to the study, they had expected the programme to contain some high-intensity aerobic exercises, confirming that some were possibly more fit and motivated to exercise than the average representative from the target population. As such, sampling bias may have played a role in this study, which could have some consequences for how the programme was designed in the end. People with lower levels of physical function and less motivation to engage in an exercise programme, might have different preferences for which type of activities they like and when they would like to perform them. It should also be acknowledged that we included participants from three different countries, and thus cultural differences could also have influenced the results to some extent. As an example, performing balance exercises out in the public could be more or less uncomfortable in one country as opposed to another. An idea could be to have different versions of the programme tailored to specific countries or populations, e.g. taking into consideration that more than half the year can be snowy in Norway, but not in southern Germany. Cultural adaptations of the programme could make it easier to change behaviour and form habits, which is an important for long-term adherence.

For the iTUG-study, participants represented the whole spectrum of physical functioning, as indicated by their scores on the CBMS being spread along the entire scale, which ranges from 0-96, with a higher score indicating higher functional ability. A clustering of scores was, however, observed in the lower and upper ends of the scale by the geriatric patients and community-dwelling healthy seniors, respectively. Having a heterogenous sample with respect to the response variable is an advantage when training machine learning algorithms, as the model can get more information for making accurate, robust and valid predictor models. Another important requirement is to have a large enough dataset. The sample size in our study, with 60 participants, was relatively small. It is generally recognized that machine learning-based prediction models trained on small sample-sizes are vulnerable to biased performance estimates (119). This study was however intended to provide

a ground truth model, and to achieve more certainty of its external validity, the PLSR model should first be trained on a larger dataset, and then validated in a separate sample representing the target population.

4.3.3 Validity of app-based self-administrable clinical tests

We aimed to develop self-test apps to detect and monitor functional decline in young seniors, and a timely question is thus whether they can do that, i.e., are the tests valid and reliable?

Several studies over the past decade have suggested that instrumented versions of the TUG may be more useful for identifying people at risk of falling as compared to the original TUG (19, 120), due to the possibility of deriving more accurate and detailed outcomes (features) derived from body-worn sensors. What has not been known is, however, how well the instrumented versions of TUG, which the Self-TUG is based on, can detect functional decline in young seniors. In our iTUG validation study, we showed that the iTUG could predict scores on the CBMS with an 85% accuracy. This is an interesting finding, because the CBMS is considered a promising performance-based test battery due to its motor task complexity; it is a challenging test, not prone to ceiling or floor effects, and has been found valid and reliable in healthy young seniors (20, 96, 101, 121). An accurate prediction of a CBMS score with the iTUG is, however, only useful if the score of the CBMS is useful, and currently, the CBMS has not been investigated extensively in young seniors. Thus, there is little reference data or established cut-off scores for predicting or discriminating persons at risk of e.g. functional decline or falling. One study has identified cut-offs for discriminating people at risk of falls, but the findings remain to be verified in larger studies with young seniors (20).

The Self-Tandem is a test of static balance, or postural control, which is widely used as part of balance assessments both in research and in the clinic (16). Postural control is measured using a range of different set-ups; with eyes open or closed, variations in how feet are positioned, where hands are placed, shoes on/off, soft/firm surface, and duration (16). The app is based on the Sharpened Romberg test (122), which adds a sensory integration taxing condition by timing the person while standing with one foot directly in front of the other foot, heel touching toe, without shoes.

Instrumented measurements of static posturography have been found reliable and valid for quantifying postural control (34, 43), and is sensitive to differences in performance between healthy controls people with mild neurological conditions (44), healthy young and senior people (34), and fallers and non-fallers (45). Sensor-based measurement of static posturography thus seem reasonable to include in an app-based battery of functional self-tests for young seniors. However, most studies that have validated sensor-based posturography have done so in controlled lab-

settings, where IMUs are fixated on standardised locations, e.g. at the L3 (34) or L5 (43, 44) on the lower back. This setting is very different from the set-up intended for the Self-Tandem, where the user performs the test unsupervised in a home environment, by holding the smartphone against their chest during the test. One study showed that some, but not all, sensor signals obtained from a smartphone held against the chest, such as in the Self-Tandem, were comparable with those obtained from a force plate, and that the features measured with the smartphone could discriminate between seniors at high and low risk of falling (54). However, in this study the participants were observed by researchers, verifying that they performed the tests correctly and held the smartphone such that it was oriented properly aligned with AP, ML and V axes. This is not possible with the current version of Self-Tandem. This issue could perhaps be mitigated by restricting the outcome measures to be based on features which do not require distinguishing between directions, but possibly at the expense of the validity due to less information to base the outcome measure on. Alternatively, all sensor signals are used to achieve the most valid outcomes from an instrumented Tandem, and we design the instructions to stress the correct orientation of the smartphone and provide a more valid but less robust outcome measure.

In another app-based self-test of balance control and fall risk, named Steady (60), the developers implemented a function in the app where the user after an attempt, has to tick of one of three options, indicating whether they performed the test correctly, incorrectly, or didn't attempt to. In contrast to Self-Tandem, Steady does not utilise the sensor signals for generating an outcome measure, but rather uses the self-reported ability to complete the balance tasks included in the app, as well as health history data. A similar solution could potentially be used in the Self-Tandem, where the user self-reported whether they performed the Self-Tandem exactly as described in the instructions, or not. If the user reports doing so, the app could provide an outcome measures that makes specific assumptions of device orientation. Nevertheless, a method study is required to gain knowledge on the validity and reliability of an unsupervised, instrumented test of postural control.

Sit-to-stand tests have been widely used to assess physical function in seniors (16, 88-92, 123). The time people use to perform five repetitions of sit-to-stand correlates with gait speed, and is a predictor of further disability (123). As with the Tandem, sit-to-stand tests are administered in many different variations (87-92), such as duration, repetitions and hands and feet placement, making comparisons of results difficult (16). The Self-STS developed as part of this thesis is an app-based version of the Five times sit-to-stand, originally used as one out of three items in the SPPB (90). We included this test as a stand-alone test as it is a commonly used variation of the sit-to-stand (16). In the literature, the instrumented sit-to-stand have been shown valid for a number of outcomes (46-51). However, in these studies the sensors were attached to the lower back, unlike our set-up for the

Self-STS, where the smartphone is supposed to be inserted into the front pocket of the trouser. This set-up has not yet been validated, but as discussed in Paper III, we believe this offers the best trade-off between motion detection ability on one side and ease of use on the other side. Another aspect to consider is that the abovementioned studies (46-51), as well as the studies included in the systematic review (16), instrumented self-tests were not validated in the target population of young seniors.

In summary, we selected self-tests based on well-known and commonly used clinical tests, which have been validated by various methods and across different populations. However, as shown in our review, few studies have been conducted to assess the measurement properties of the original version of the tests in young seniors, and no instrumented tests were identified. Instrumented versions of the tests that our self-tests are based on, have been validated in the literature, albeit not when administered as intended and described in this thesis, i.e. unsupervised with the smartphone worn in the trouser pocket. As part of further investigations, a wider method study would thus be required that is designed to evaluate not only the usability, but also the validity and reliability of the Self-TUG, Self-Tandem and Self-STS in young seniors.

4.3.4 App development and user testing

The development and usability testing conducted as part of this thesis were partly guided by principles from HCD, including: (1) The design is based upon an explicit understanding of users, tasks, and environments; (2) Users are involved throughout design and development; (3) The design is driven and refined by user-centered evaluation; (4) The process is iterative; (5) The design addresses the whole user experience; (6) The design team includes multidisciplinary skills and perspectives (58). Our approach meets most of these requirements, however we did not include end-users in the initial planning of how the apps would be designed.

Although the apps were tested in potential end-users during the first and second iteration, it was not until the home-based usability testing that we collected their experiences through questionnaires and interviews. An argument could be made that they should have been included from the beginning, to ensure that their perspectives and preferences were considered in every step of the design and development. This is a goal of design processes such as co-creation (124), where the end-user is viewed more as “partner” as compared to a “subject”, and end-users are thus more engaged together with the designer in an act of collective creativity during the design process. Co-creation has been applied in development of another smartphone-based self-test of balance by a research group from Umeå, Sweden, and in a paper describing this process, they conclude that co-creation was a feasible and valuable method for developing self-test apps for seniors (61). However, by

testing the apps in such a large sample as we did in iteration 1 and 2, we would have identified most usability issues that possibly could lead to an incorrect trial, and thus directly affect the validity of the test. By addressing them and re-iterating, any usability problems that caused people to make errors would eventually have been eliminated, and in our approach, we therefore prioritised identifying and eliminating as many usability problems as possible before involving end-users in the design process.

For the third iteration of usability testing that was conducted in the participants' own homes, end-users were more involved as we collected their opinions of the apps with questionnaires and an interview. Furthermore, testing in this setting provided valuable insight and introduced several issues that were not observed in the lab, increasing the ecological validity of the results.

4.4 Future work

The work in this thesis have sought to develop and evaluate methods for identifying and targeting functional decline in young seniors. The thesis has not itself resulted in implementation-ready methods, but it has driven the research further and achieved information necessary for formulating relevant questions to answer in future work.

The feasibility of the aLiFE programme has already been evaluated in the PreventIT-study, but results have not been published yet. The study was not designed to prove its effectiveness as compared to the other arms in the intervention; control and ICT-delivered aLiFE, but the study results should be a good indication of its feasibility to be delivered in the target group, including participants' adherence to the programme, which is an important condition for it to be considered as a means to prevent functional decline.

For the self-tests, a first step would be to validate the algorithms for identifying the sub-phases in Self-TUG and Self-STS with the smartphone placed in the pocket. Secondly, the features derived from all the self-tests should be further evaluated for their ability to detect or predict functional decline in young seniors. The latter was attempted in Paper IV, with promising results, but the findings need to be validated in a separate sample of participants for external validity. Should the results from such an investigation be positive, then further work should also be conducted to evaluate the predictive/diagnostic ability of the CBMS in young seniors.

5. Conclusions

Tests of physical function are being increasingly implemented in smartphones by utilising the built-in sensors to increase the accuracy and sensitivity of the outcome measures. Such solutions could be especially relevant for healthy young seniors, who are using smartphones at an increasing rate, and for whom traditional clinical tests are not sufficiently sensitive for detecting early signs of functional decline. The field of research focusing on healthy younger seniors is new, but given the rate at which this population is growing, together with the evidence for physical activity and exercise, it is important to both develop tests that can accurately identify functional decline, as well as interventions that can effectively prevent or delay further functional decline in healthy young seniors.

Results from this thesis provide knowledge of the tests used in healthy young seniors over the years. It was found that many tests of balance and strength have been applied in this population, but for many of these tests, the procedures for administration vary between studies, and there is a lack of studies evaluating their measurement properties in young seniors. The original TUG is the most used test also in healthy young seniors, but studies of its measurement properties in young seniors do not indicate that it is an appropriate test for this group. The CBMS seems like a promising clinical tool for assessing physical function in young seniors due to its challenging tasks and ability to measure advanced balance and mobility.

We found that healthy young seniors were positive towards engaging in a life-style integrated exercise programme, where activities for increasing balance, muscle strength, general physical activity and reducing sedentariness, were integrated in daily tasks and routines. However, an RCT with longer follow-up which is designed to establish its effectiveness compared to other intervention approaches is required. Based on the findings of this thesis, the CBMS should be included to measure change in physical function before and after the intervention.

The smartphone apps presented in this thesis, were shown to have potential for being offered as a solution for self-testing and -monitoring of physical function in healthy young seniors in an un-supervised home-setting. Some improvements remain to improve the usability, and both the algorithms and outcomes need to be validated before being made available to end-users.

In a lab-setting, when standardised and supervised by trained assessors, we found that features derived from a five-times repeated iTUG could predict the score on the CBMS with a high accuracy in a mixed sample of healthy community-dwellers and geriatric patients. The prediction accuracy was similar to that of a battery of standard clinical tests. Because the repeated iTUG requires less than

five minutes to administer, little space and nothing but a chair and a smartphone, the findings can potentially save time and resources in research and clinical care. As with many prediction models built from machine-learning methods, an external validation should be performed to verify the results of this work from the thesis.

6. References

1. Verbrugge LM, Jette AM. The disablement process. Soc Sci Med. 1994;38(1):1-14.
2. Stuck AE, Walthert JM, Nikolaus T, Büla CJ, Hohmann C, Beck JC. Risk factors for functional status decline in community-living elderly people: a systematic literature review. Soc Sci Med. 1999;48(4):445-69.
3. Guralnik JM, Ferrucci L, Pieper CF, Leveille SG, Markides KS, Ostir GV, et al. Lower Extremity Function and Subsequent Disability: Consistency Across Studies, Predictive Models, and Value of Gait Speed Alone Compared With the Short Physical Performance Battery. The Journals of Gerontology: Series A. 2000;55(4):M221-M31.
4. Chen C, Lim JT, Chia NC, Wang L, Tysinger B, Zissimopoulos J, et al. The long-term impact of functional disability on hospitalization spending in Singapore. The Journal of the Economics of Ageing. 2019;14:100193.
5. United Nations. Department of Economic and Social Affairs, Population Division. World population prospects: The 2017 Revision, Key Findings and Advance Tables. New york; 2017.
6. Guralnik JM, LaCroix AZ, Abbott RD, Berkman LF, Satterfield S, Evans DA, et al. Maintaining Mobility in Late Life. I. Demographic Characteristics and Chronic Conditions. Am J Epidemiol. 1993;137(8):845-57.
7. Choy NL, Brauer S, Nitz J. Changes in postural stability in women aged 20 to 80 years. J Gerontol A Biol Sci Med Sci. 2003;58(6):525-30.
8. Isles RC, Choy NL, Steer M, Nitz JC. Normal values of balance tests in women aged 20-80. J Am Geriatr Soc. 2004;52(8):1367-72.
9. Engberg E, Alen M, Kukkonen-Harjula K, Peltonen JE, Tikkanen HO, Pekkarinen H. Life events and change in leisure time physical activity. Sports Med. 2012;42(5):433-47.
10. Van Dyck D, Cardon G, De Bourdeaudhuij I. Longitudinal changes in physical activity and sedentary time in adults around retirement age: what is the moderating role of retirement status, gender and educational level? BMC Public Health. 2016;16(1):1125.
11. L. Sturnieks D, St George R, R. Lord S. Balance disorders in the elderly. Neurophysiologie Clinique/Clinical Neurophysiology. 2008;38(6):467-78.
12. Liu CJ, Latham NK. Progressive resistance strength training for improving physical function in older adults. Cochrane Database Syst Rev. 2009(3).

13. Freiberger E, de Vreede P, Schoene D, Rydwik E, Mueller V, Frändin K, et al. Performance-based physical function in older community-dwelling persons: a systematic review of instruments. Age Ageing. 2012;41(6):712-21.
14. Bendayan R, Cooper R, Wloch EG, Hofer SM, Piccinin AM, Muniz-Terrera G. Hierarchy and Speed of Loss in Physical Functioning: A Comparison Across Older U.S. and English Men and Women. The journals of gerontology Series A, Biological sciences and medical sciences. 2017;72(8):1117-22.
15. Studenski S, Perera S, Wallace D, Chandler JM, Duncan PW, Rooney E, et al. Physical Performance Measures in the Clinical Setting. J Am Geriatr Soc. 2003;51(3):314-22.
16. Bergquist R, Weber M, Schwenk M, Ulseth S, Helbostad JL, Vereijken B, et al. Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review. BMC Geriatr. 2019;19(1):9.
17. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142-8.
18. Schoene D, Wu SM, Mikolaizak AS, Menant JC, Smith ST, Delbaere K, et al. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. J Am Geriatr Soc. 2013;61(2):202-8.
19. Sprint G, Cook DJ, Weeks DL. Toward Automating Clinical Assessments: A Survey of the Timed Up and Go. IEEE Rev Biomed Eng. 2015;8:64-77.
20. Balasubramanian CK. The community balance and mobility scale alleviates the ceiling effects observed in the currently used gait and balance assessments for the community-dwelling older adults. J Geriatr Phys Ther. 2015;38(2):78-89.
21. Bergland A, Strand BH. Norwegian reference values for the Short Physical Performance Battery (SPPB): the Tromsø Study. BMC Geriatr. 2019;19(1):216.
22. Garin O. Ceiling Effect. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer Netherlands; 2014. p. 631-3.
23. Tak E, Kuiper R, Chorus A, Hopman-Rock M. Prevention of onset and progression of basic ADL disability by physical activity in community dwelling older adults: A meta-analysis. Ageing Research Reviews. 2013;12(1):329-38.
24. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. JAMA. 2018;320(19):2020-8.

25. Chodzko-Zajko WJ, Proctor DN, Fiatarone Singh MA, Minson CT, Nigg CR, Salem GJ, et al. Exercise and Physical Activity for Older Adults. Med Sci Sports Exerc. 2009;41(7):1510-30.
26. Hansen B, Anderssen S, Steene-Johannessen J, Ekelund U, Nilsen A, Andersen ID, et al. Fysisk aktivitet og sedat tid blant voksne og eldre i Norge—Nasjonal Kartlegging 2014–2015. Norwegian Directorate of Health: Oslo, Norway. 2015;154.
27. Hill KD, Hunter SW, Batchelor FA, Cavalheri V, Burton E. Individualized home-based exercise programs for older people to reduce falls and improve physical performance: A systematic review and meta-analysis. Maturitas. 2015;82(1):72-84.
28. Weber M, Belala N, Clemson L, Boulton E, Hawley-Hague H, Becker C, et al. Feasibility and Effectiveness of Intervention Programmes Integrating Functional Exercise into Daily Life of Older Adults: A Systematic Review. Gerontology. 2018;64(2):172-87.
29. Clemson L, Fiatarone Singh MA, Bundy A, Cumming RG, Manollaras K, O'Loughlin P, et al. Integration of balance and strength training into daily life activity to reduce rate of falls in older people (the LiFE study): randomised parallel trial. BMJ. 2012;345:e4547.
30. Cutlip RG, Mancinelli C, Huber F, DiPasquale J. Evaluation of an instrumented walkway for measurement of the kinematic parameters of gait. Gait Posture. 2000;12(2):134-8.
31. Feiring DC, Ellenbecker TS, Derscheid GL. Test-Retest Reliability of the Biodex Isokinetic Dynamometer. J Orthop Sports Phys Ther. 1990;11(7):298-300.
32. Bourke AK, Ihlen EAF, Bergquist R, Wik PB, Vereijken B, Helbostad JL. A Physical Activity Reference Data-Set Recorded from Older Adults Using Body-Worn Inertial Sensors and Video Technology—The ADAPT Study Data-Set. Sensors. 2017;17(3):559.
33. Ejupi A, Lord SR, Delbaere K. New methods for fall risk prediction. Curr Opin Clin Nutr Metab Care. 2014;17(5):407-11.
34. Moe-Nilssen R, Helbostad JL. Trunk accelerometry as a measure of balance control during quiet standing. Gait Posture. 2002;16(1):60-8.
35. Mellone S. Analisi del movimento tramite sensori inerziali: dal laboratorio al letto del paziente [dissertation]. Bologna: University of Bologna; 2013.
36. Mirelman A, Weiss A, Buchman AS, Bennett DA, Giladi N, Hausdorff JM. Association between performance on Timed Up and Go subtasks and mild cognitive impairment: further insights into the links between cognitive and motor function. J Am Geriatr Soc. 2014;62(4):673-8.

37. Palmerini L, Mellone S, Avanzolini G, Valzania F, Chiari L. Quantification of motor impairment in Parkinson's disease using an instrumented timed up and go test. IEEE Trans Neural Syst Rehabil Eng. 2013;21(4):664-73.
38. Salarian A, Horak FB, Zampieri C, Carlson-Kuhta P, Nutt JG, Aminian K. iTUG, a sensitive and reliable measure of mobility. IEEE Trans Neural Syst Rehabil Eng. 2010;18(3):303-10.
39. Weiss A, Herman T, Plotnik M, Brozgol M, Giladi N, Hausdorff JM. An instrumented timed up and go: the added value of an accelerometer for identifying fall risk in idiopathic fallers. Physiol Meas. 2011;32(12):2003-18.
40. Caronni A, Sterpi I, Antoniotti P, Aristidou E, Nicolaci F, Picardi M, et al. Criterion validity of the instrumented Timed Up and Go test: A partial least square regression study. Gait Posture. 2018;61:287-93.
41. Greene BR, O'Donovan A, Romero-Ortuno R, Cogan L, Scanaill CN, Kenny RA. Quantitative Falls Risk Assessment Using the Timed Up and Go Test. IEEE Trans Biomed Eng. 2010;57(12):2918-26.
42. Buchman AS, Boyle PA, Leurgans SE, Barnes LL, Bennett DA. Cognitive function is associated with the development of mobility impairments in community-dwelling elders. Am J Geriatr Psychiatry. 2011;19(6):571-80.
43. Godfrey A, Lara J, Del Din S, Hickey A, Munro CA, Wiuff C, et al. iCap: Instrumented assessment of physical capability. Maturitas. 2015;82(1):116-22.
44. Mancini M, Salarian A, Carlson-Kuhta P, Zampieri C, King L, Chiari L, et al. ISway: a sensitive, valid and reliable measure of postural control. J Neuroeng Rehabil. 2012;9(1):59.
45. Greene BR, McGrath D, Walsh L, Doheny EP, McKeown D, Garattini C, et al. Quantitative falls risk estimation through multi-sensor assessment of standing balance. Physiol Meas. 2012;33(12):2049.
46. Coni A, Van Ancum JM, Bergquist R, Mikolaizak AS, Mellone S, Chiari L, et al. Comparison of Standard Clinical and Instrumented Physical Performance Tests in Discriminating Functional Status of High-Functioning People Aged 61–70 Years Old. Sensors. 2019;19(3):449.
47. Janssen WGM, Bussmann JBJ, Horemans HLD, Stam HJ. Validity of accelerometry in assessing the duration of the sit-to-stand movement. Med Biol Eng Comput. 2008;46(9):879-87.

48. Najafi B, Aminian K, Loew F, Blanc Y, Robert PA. Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly. IEEE Trans Biomed Eng. 2002;49(8):843-51.
49. Regterschot GR, Zhang W, Baldus H, Stevens M, Zijlstra W. Sensor-based monitoring of sit-to-stand performance is indicative of objective and self-reported aspects of functional status in older adults. Gait Posture. 2015;41(4):935-40.
50. Van Lummel RC, Ainsworth E, Lindemann U, Zijlstra W, Chiari L, Van Campen P, et al. Automated approach for quantifying the repeated sit-to-stand using one body fixed sensor in young and older adults. Gait Posture. 2013;38(1):153-6.
51. van Lummel RC, Walgaard S, Maier AB, Ainsworth E, Beek PJ, van Dieen JH. The Instrumented Sit-to-Stand Test (iSTS) has greater clinical relevance than the manually recorded sit-to-stand test in older adults. PLoS One. 2016;11(7).
52. Mellone S, Tacconi C, Schwickert L, Klenk J, Becker C, Chiari L. Smartphone-based solutions for fall detection and prevention: the FARSEEING approach. Z Gerontol Geriatr. 2012;45(8):722-7.
53. Mellone S, Tacconi C, Chiari L. Validity of a Smartphone-based instrumented Timed Up and Go. Gait Posture. 2012;36(1):163-5.
54. Hsieh KL, Roach KL, Wajda DA, Sosnoff JJ. Smartphone technology can measure postural stability and discriminate fall risk in older adults. Gait Posture. 2019;67:160-5.
55. Anderson M, Perrin A. Technology use among seniors [Internet]: Pew Research Center; 2017 [cited 2019 Feb 15]. Available from: https://www.silvergroup.asia/wp-content/uploads/2017/07/Technology-use-among-seniors-_-Pew-Research-Center.pdf
56. Roeing KL, Hsieh KL, Sosnoff JJ. A systematic review of balance and fall risk assessments with mobile phone technology. Arch Gerontol Geriatr. 2017;73:222-6.
57. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff. 2014;33(2):194-9.
58. Harte R, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, Quinlan LR, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR human factors. 2017;4(1):e8.
59. International Organization for Standardization. Standard No. 9241-210. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems

[Internet] Geneva: ISO; 2019 [cited 2020 Jul 9]. Available from: https://www.standard.no/no/Nettbutikk/produktkatalogen/Produktpresentasjon/?ProductID=1046413.

60. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. JMIR Aging. 2018;1(2):e11569.
61. Mansson L, Wiklund M, Öhberg F, Danielsson K, Sandlund M. Co-Creation with Older Adults to Improve User-Experience of a Smartphone Self-Test Application to Assess Balance Function. Int J Environ Res Public Health. 2020;17(11):3768.
62. Zhong R, Rau P-LP. A Mobile Phone–Based Gait Assessment App for the Elderly: Development and Evaluation. JMIR Mhealth Uhealth. 2020;8(5):e14453.
63. Sun R, Sosnoff JJ. Novel sensing technology in fall risk assessment in older adults: a systematic review. BMC Geriatr. 2018;18(1):14.
64. Schwenk M, Bergquist R, Boulton E, Van Ancum JM, Nerz C, Weber M, et al. The Adapted Lifestyle-Integrated Functional Exercise Program for Preventing Functional Decline in Young Seniors: Development and Initial Evaluation. Gerontology. 2019;65(4):362-74.
65. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695-9.
66. Bergquist R, Vereijken B, Mellone S, Corzani M, Helbostad JL, Taraldsen K. App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study. JMIR mHealth and uHealth. 2020;8(4):e16507-e.
67. Taraldsen K, Mikolaizak AS, Maier AB, Boulton E, Aminian K, van Ancum J, et al. Protocol for the PreventIT feasibility randomised controlled trial of a lifestyle-integrated exercise intervention in young older adults. BMJ Open. 2019;9(3):e023526.
68. Bergquist R, Nerz C, Taraldsen K, Mellone S, Ihlen EAF, Vereijken B, et al. Predicting advanced balance ability and mobility with an instrumented Timed Up and Go-test. Manuscript submitted for publication. 2020.
69. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49.

70. Clemson L, Munro J, Singh MF. Lifestyle-integrated Functional Exercise (LiFE) program to prevent falls: trainer's manual: Sydney University Press; 2014.
71. Boulton E, Weber M, Hawley-Hague H, Bergquist R, Van Ancum J, Jonkman NH, et al. Attitudes Towards Adapted Lifestyle-Integrated Functional Exercise Developed for 60–70-Year-Olds: Perceptions of Participants and Trainers. Gerontology. 2019;65(6):599-609.
72. Zhang W, Schwenk M, Mellone S, Paraschiv-Ionescu A, Vereijken B, Pijnappels M, et al. Complexity of daily physical activity is more sensitive than conventional metrics to assess functional change in younger older adults. Sensors. 2018;18(7):2032.
73. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability evaluation in industry. London: Taylor & Francis; 1996. p. 189-94.
74. Peres SC, Pham T, Phillips R. Validation of the System Usability Scale (SUS):SUS in the Wild. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2013;57(1):192-6.
75. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.
76. Abe T, Ogawa M, Loenneke JP, Thiebaud RS, Loftin M, Mitsukawa N. Association between site-specific muscle loss of lower body and one-leg standing balance in active women: The HIREGASAKI study. Geriatrics & Gerontology International. 2014;14(2):381-7.
77. Benavent-Caballer V, Rosado-Calatayud P, Segura-Ortí E, Amer-Cuenca J, Lisón J. The effectiveness of a video-supported group-based Otago exercise programme on physical performance in community-dwelling older adults: a preliminary study. Physiotherapy. 2016;102(3):280-6.
78. Brown M, Holloszy JO. Effects of walking, jogging and cycling on strength, flexibility, speed and balance in 60-to 72-year olds. Aging Clin Exp Res. 1993;5(6):427-34.
79. Chun SH, Cho B, Yang H-K, Ahn E, Han MK, Oh B, et al. Performance on physical function tests and the risk of fractures and admissions: findings from a national health screening of 557,648 community-dwelling older adults. Arch Gerontol Geriatr. 2017;68:174-80.
80. Hulya TD, Sevi YSS, Serap A, Ayse OE. Factors affecting the benefits of a six-month supervised exercise program on community-dwelling older adults: interactions among age, gender, and participation. Journal of physical therapy science. 2015;27(5):1421-7.

81. Jalali MM, Gerami H, Heidarzadeh A, Soleimani R. Balance performance in older adults and its relationship with falling. Aging Clin Exp Res. 2015;27(3):287-96.
82. Kuo Y-L, Huang K-Y, Chiang P-T, Lee P-Y, Tsai Y-J. Steadiness of spinal regions during single-leg standing in older adults with and without chronic low back pain. PLoS One. 2015;10(5):e0128318.
83. Moriya S, Notani K, Murata A, Inoue N, Miura H. Analysis of moment structures for assessing relationships among perceived chewing ability, dentition status, muscle strength, and balance in community-dwelling older adults. Gerodontology. 2014;31(4):281-7.
84. Schilling BK, Falvo MJ, Karlage RE, Weiss LW, Lohnes CA, Chiu LZ. Effects of unstable surface training on measures of balance in older adults. The Journal of Strength & Conditioning Research. 2009;23(4):1211-6.
85. Yıldırım P, Ofluoglu D, Aydogan S, Akyuz G. Tai Chi vs. combined exercise prescription: A comparison of their effects on factors related to falls. J Back Musculoskelet Rehabil. 2016;29(3):493-501.
86. yun Liu X, Gao J, xiang Yin B, yu Yang X, xi Bai D. Efficacy of Ba Duan Jin in improving balance: a study in Chinese community-dwelling older adults. J Gerontol Nurs. 2016;42(5):38-46.
87. Brotherton SS, Williams HG, Gossard JL, Hussey JR, McClenaghan BA, Eleazer P. Are measures employed in the assessment of balance useful for detecting differences among groups that vary by age and disease state? J Geriatr Phys Ther. 2005;30(1):14-9.
88. Cooper AJ, Simmons RK, Kuh D, Brage S, Cooper R. Physical activity, sedentary time and physical capability in early old age: British birth cohort study. PLoS One. 2015;10(5):e0126465.
89. Gray SL, Aragaki AK, LaMonte MJ, Cochrane BB, Kooperberg C, Robinson JG, et al. Statins, angiotensin-converting enzyme inhibitors, and physical performance in older women. J Am Geriatr Soc. 2012;60(12):2206-14.
90. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol. 1994;49(2):M85-94.
91. King AC, Pruitt LA, Phillips W, Oka R, Rodenburg A, Haskell WL. Comparative effects of two physical activity programs on measured and perceived physical functioning and other

health-related quality of life outcomes in older adults. J Gerontol A Biol Sci Med Sci. 2000;55(2):M74-83.

92. Rikli RE, Jones CJ. Development and validation of a functional fitness test for community-residing older adults. J Aging Phys Act. 1999;7(2):129-61.

93. Alexandre TS, Meira DM, Rico NC, Mizuta SK. Accuracy of Timed Up and Go Test for screening risk of falls among community-dwelling elderly. Brazilian Journal of Physical Therapy. 2012;16(5):381-8.

94. Cani KC, Silva IJCS, Karloh M, Gulart AA, Matte DL, Mayer AF. Reliability of the five-repetition sit-to-stand test in patients with chronic obstructive pulmonary disease on domiciliary oxygen therapy. Physiother Theory Pract. 2020;36(1):219-25.

95. Gómez JF, Curcio C-L, Alvarado B, Zunzunegui MV, Guralnik J. Validity and reliability of the Short Physical Performance Battery (SPPB): a pilot study on mobility in the Colombian Andes. Colomb Med. 2013;44(3):165-71.

96. Gordt K, Paraschiv-Ionescu A, Mikolaizak AS, Taraldsen K, Mellone S, Bergquist R, et al. The association of basic and challenging motor capacity with mobility performance and falls in young seniors. Arch Gerontol Geriatr. 2020;90:104134.

97. Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the Timed “Up & Go” test in people with Parkinson disease. Phys Ther. 2001;81(2):810-8.

98. Ng SS, Hui-Chan CW. The timed up & go test: its reliability and association with lower-limb impairments and locomotor capacities in people with chronic stroke. Arch Phys Med Rehabil. 2005;86(8):1641-7.

99. Scaglioni-Solano P, Aragón-Vargas LF. Validity and reliability of the Nintendo Wii Balance Board to assess standing balance and sensory integration in highly functional older adults. Int J Rehabil Res. 2014;37(2):138-43.

100. Takacs J, Garland SJ, Carpenter MG, Hunt MA. Validity and reliability of the community balance and mobility scale in individuals with knee osteoarthritis. Phys Ther. 2014;94(6):866-74.

101. Weber M, Van Ancum J, Bergquist R, Taraldsen K, Gordt K, Mikolaizak AS, et al. Concurrent validity and reliability of the Community Balance and Mobility scale in young-older adults. BMC Geriatr. 2018;18(1):156.

102. Howe JA, Inness EL, Venturini A, Williams JI, Verrier MC. The Community Balance and Mobility Scale--a balance measure for individuals with traumatic brain injury. Clin Rehabil. 2006;20(10):885-95.
103. Johnson BG, Wright AD, Beazley MF, Harvey TC, Hillenbrand P, Imray CHE. The Sharpened Romberg Test for Assessing Ataxia in Mild Acute Mountain Sickness. Wilderness Environ Med. 2005;16(2):62-6.
104. Palmerini L, Mellone S, Rocchi L, Chiari L. Dimensionality reduction for the quantitative evaluation of a smartphone-based Timed Up and Go test. Conf Proc IEEE Eng Med Biol Soc. 2011;2011:7179-82.
105. Burton E, Lewin G, Clemson L, Boldy D. Effectiveness of a lifestyle exercise program for older people receiving a restorative home care service: a pragmatic randomized controlled trial. Clin Interv Aging. 2013;8:1591-601.
106. Lesinski M, Hortobágyi T, Muehlbauer T, Gollhofer A, Granacher U. Effects of Balance Training on Balance Performance in Healthy Older Adults: A Systematic Review and Meta-analysis. Sports Med. 2015;45(12):1721-38.
107. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I, et al. How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act. 2011;8:80.
108. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol. 1983;51(3):390.
109. Stiggelbout M, Hopman-Rock M, Crone M, Lechner L, Van Mechelen W. Predicting older adults' maintenance in exercise participation using an integrated social psychological model. Health Educ Res. 2006;21(1):1-14.
110. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human–Computer Interaction. 2008;24(6):574-94.
111. Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. J Rehabil Med. 2010;42(4):323-31.
112. Siddaway AP, Wood AM, Hedges LV. How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. Annu Rev Psychol. 2019;70:747-70.

113. Sayers SP, Guralnik JM, Newman AB, Brach JS, Fielding RA. Concordance and discordance between two measures of lower extremity function: 400 meter self-paced walk and SPPB. Aging Clin Exp Res. 2006;18(2):100-6.
114. Smith J, Noble H. Bias in research. Evidence Based Nursing. 2014;17(4):100.
115. Cheung KL, ten Klooster PM, Smit C, de Vries H, Pieterse ME. The impact of non-response bias due to sampling in public health studies: A comparison of voluntary versus mandatory recruitment in a Dutch national survey on adolescent health. BMC Public Health. 2017;17(1):276.
116. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. Journal of Usability Studies. 2009;5(1):34-45.
117. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. American journal of theoretical and applied statistics. 2016;5(1):1-4.
118. Harte RP, Glynn LG, Broderick BJ, Rodriguez-Molinero A, Baker PM, McGuiness B, et al. Human centred design considerations for connected health devices for the older adult. J Pers Med. 2014;4(2):245-81.
119. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One. 2019;14(11):e0224365-e.
120. Greene BR, Doheny EP, Walsh C, Cunningham C, Crosby L, Kenny RA. Evaluation of Falls Risk in Community-Dwelling Older Adults Using Body-Worn Sensors. Gerontology. 2012;58(5):472-80.
121. Gordt K, Mikolaizak AS, Taraldsen K, Bergquist R, Van Ancum JM, Nerz C, et al. Creating and Validating a Shortened Version of the Community Balance and Mobility Scale for Application in People Who Are 61 to 70 Years of Age. Phys Ther. 2019.
122. Franchignoni F, Tesio L, Martino MT, Ricupero C. Reliability of four simple, quantitative tests of balance and mobility in healthy elderly females. Aging Clin Exp Res. 1998;10(1):26-31.
123. Whitney SL, Wrisley DM, Marchetti GF, Gee MA, Redfern MS, Furman JM. Clinical Measurement of Sit-to-Stand Performance in People With Balance Disorders: Validity of Data for the Five-Times-Sit-to-Stand Test. Phys Ther. 2005;85(10):1034-45.
124. Sanders EB-N, Stappers PJ. Co-creation and the new landscapes of design. Co-design. 2008;4(1):5-18.

Paper I

Bergquist et al. BMC Geriatrics (2019) 19:9
https://doi.org/10.1186/s12877-018-1011-0

BMC Geriatrics

RESEARCH ARTICLE Open Access

Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review

Ronny Bergquist[1*†], Michaela Weber[2†], Michael Schwenk[2,3], Synnøve Ulseth[1], Jorunn L. Helbostad[1], Beatrix Vereijken[1] and Kristin Taraldsen[1]

Abstract

Background: Many balance and strength tests exist that have been designed for older seniors, often aged ≥70 years. To guide strategies for preventing functional decline, valid and reliable tests are needed to detect early signs of functional decline in young seniors. Currently, little is known about which tests are being used in young seniors and their methodological quality. This two-step review aims to 1) identify commonly used tests of balance and strength, and 2) evaluate their measurement properties in young seniors.

Methods: First, a systematic literature search was conducted in MEDLINE to identify primary studies that employed performance-based tests of balance and muscle strength, and which aspects of balance and strength these tests assess in young seniors aged 60–70. Subsequently, for tests used in ≥3 studies, a second search was performed to identify method studies evaluating their measurement properties. The quality of included method studies was evaluated using the Consensus-based Standards for selection of health Measurement Instruments (COSMIN) checklist.

Results: Of 3454 articles identified, 295 met the inclusion criteria. For the first objective, 69 balance and 51 muscle strength tests were identified, with variations in administration mode and outcome reporting. Twenty-six balance tests and 15 muscle strength tests were used in ≥3 studies, with proactive balance tests and functional muscle power tests used most often. For the second objective, the search revealed 1880 method studies, of which nine studies (using 5 balance tests and 1 strength test) were included for quality assessment. The Timed Up and Go test was evaluated the most (4 studies), while the Community Balance and Mobility (CBM) scale was the second most assessed test (3 studies). For strength, one study assessed the reliability of the Five times sit-to-stand.

Conclusion: Commonly used balance and muscle strength tests in young seniors vary greatly with regards to administration mode and outcome reporting. Few studies have evaluated measurement properties of these tests when used in young seniors. There is a need for standardisation of existing tests to improve their informative value and comparability. For measuring balance, the CBM is a new and promising tool to detect even small balance deficits in balance in young seniors.

Keywords: Systematic review, Performance-based tests, Measurement properties, Older adults, Balance, Muscle strength

* Correspondence: ronny.bergquist@ntnu.no
† Ronny Bergquist and Michaela Weber contributed equally to this work.
[1]Department of Neuromedicine and Movement Science, Norwegian Univerity of Science and Technology, Trondheim, Norway
Full list of author information is available at the end of the article

Background

Numerous studies have demonstrated that impairments in balance and decreased muscle strength in lower extremity muscles are important risk factors for early age-related decline in physical function [1–5], falls [3–6], future disabilities [7], hospitalization [5], and death [6–8]. Early declines in balance and muscle strength are already apparent in the third decade of life [9–12], with an accelerated decline occurring from the decade of young seniors aged 60 to 70 years [9, 13–15]. Especially age-related impairments in vision and the vestibular and proprioceptive systems, most obvious from 50 years and older [9, 16, 17], contribute to the acceleration of balance decline. For muscle strength, especially age-related changes in lean muscle mass greatly increase the risk for physical inactivity, mobility deficits, functional limitations and falls [2, 15, 18].

Balance and muscle strength tests can be used to assess and monitor individual's health over time, and predict multi-morbidity, dependence in basic activities of daily living (ADLs) and early mortality [18–22]. Such tests also are of substantial value in predicting future health status and functional performance in older adults [22].

Numerous performance-based clinical tests assessing balance and/or muscle strength exist. Tests of grip strength, walking speed, sit-to-stand, and standing balance are shown to be markers of both current and future health [1, 18–21]. As a result, there is an increased interest in these tests and their potential use as simple screening tools in the general population to identify people who may benefit from targeted interventions aimed at preventing functional decline [1, 18, 23, 24].

However, in order to test balance and muscle strength adequately, it is important that the tests are sufficiently challenging since an early detection of loss of balance and muscle strength is important to prevent age-related functional decline in young seniors [25–29]. For young seniors, generally functioning at a higher level, it is questionable whether existing balance and muscle strength tests are sensitive enough to detect early subtle balance declines [1, 23]. Balance is a complex composite of multiple body systems including the ability to align different body segments and to generate multi-joint movements to effectively control body position and movement [30]. Since balance is highly task-specific, several aspects need to be assessed which can be categorized into static steady-state balance (i.e., maintaining a steady position in sitting or standing), dynamic steady-state balance (i.e., walking), proactive balance (i.e., anticipating a predicted disturbance such as crossing or walking around an obstacle), and reactive balance (i.e., compensating for a disturbance) [30]. Recent systematic reviews of the literature on balance tests have shown that widely used assessment tools such as the Berg Balance Scale (BBS) or Short Physical Performance Battery (SPPB) show ceiling effects in community-dwelling, healthy older adults aged 60 years and over [23, 31]. Ceiling effects of these instruments in higher functioning older adults will hamper the detection of early balance deficits, and thus intervention-related changes over time may not be detected [32, 33]. Although some balance tests such as the Fullerton Advanced Balance (FAB) scale [34], are developed for use in higher functioning older adults, these tests typically do not include tasks that challenge balance for the specific population of healthy, higher functioning older adults [35, 36].

For muscle strength, commonly used tests such as the Five times sit-to-stand (5STS) are not challenging enough in order to detect risk factors in higher functioning older adults [37]. Especially with regard to confirming the effects of an intervention, such tests have ceiling effects as most older adults can perform the test effortlessly and therefore do not show changes in performance level [37].

At present, no systematic literature review has examined which balance and muscle strength tests are used for the population of young seniors. The aim of this systematic review was to 1) identify any performance-based clinical tests used to measure balance and/or muscle strength in young seniors aged 60–70 years, and 2) evaluate the measurement properties of the most commonly used performance-based clinical balance and muscle strength tests.

Methods

Study design

The study is a two-step systematic literature review with two separate literature searches. The first step included the search and systematic review of performance-based clinical tests used for measuring balance or muscle strength in young seniors.

The second step included a search and a systematic review of methodological studies evaluating the measurement properties of performance-based clinical tests that have been used in ≥3 studies identified in step one.

Search strategy

The search in step one was performed in MEDLINE to identify relevant studies published until June 1st 2016, with an update made to identify also newer studies published until November 5th 2018 (Fig. 1). A combination of free-text and MeSH-terms was used that represents the following concepts: 'postural balance', 'muscle strength', 'movement', motor activity', 'physical exertion', 'physical endurance', 'exercise tolerance', and 'physical fitness'. Additional search terms aimed to exclude animal studies, participants outside our target age group, and non-English studies (see Additional file 1). The search in

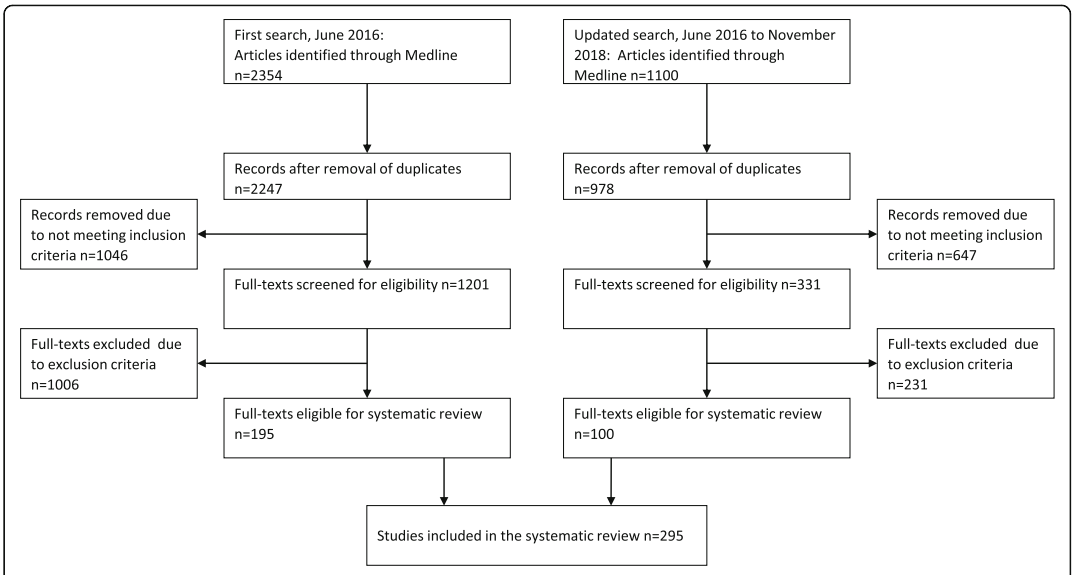

Fig. 1 Study selection of performance based tests through the different phases (first search)

step two was performed in MEDLINE and EMBASE to identify relevant method studies published until December 19th 2017, and also updated to include newer studies published until November 23rd 2018 (Fig. 2). We combined a search on the most commonly identified tests (≥3 articles) with a search on measurement properties, including validity, reliability, sensitivity, accuracy, responsiveness, and specificity (see Additional file 1).

Inclusion/exclusion criteria

In the first step, articles were included if they (1) described a performance-based clinical test that measured

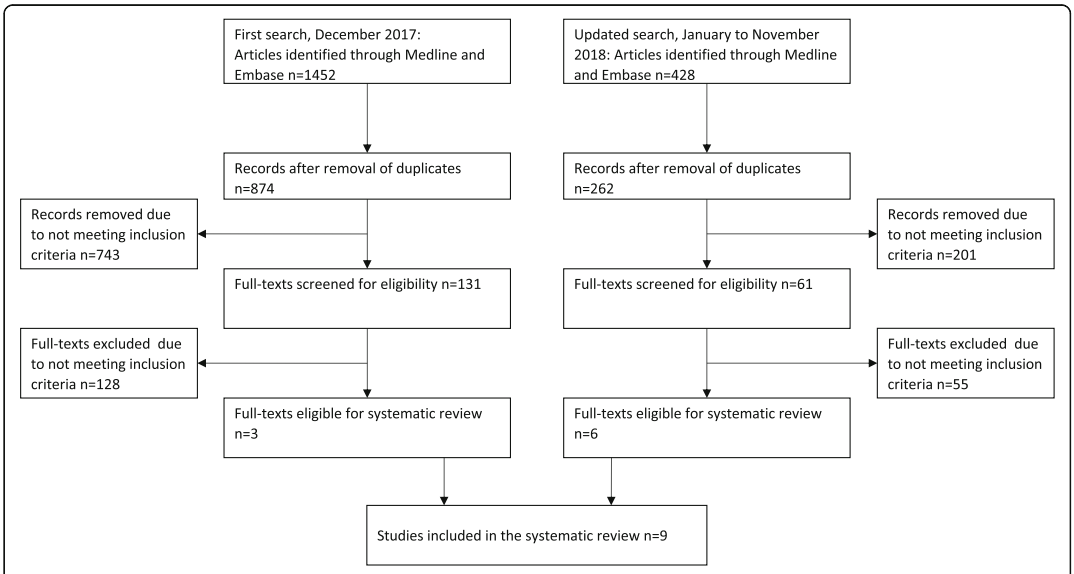

Fig. 2 Study selection of method studies through the different phases (second search)

aspects of balance and/or muscle strength, (2) included participants with an age or mean age between 60 and 70 years, and (3) were written in English. Articles were excluded if (1) in principal the test could not be completed without fixed laboratory equipment, (2) all groups were included on the basis of having a clinical condition (i.e., no healthy and/or control groups), and (3) manuscripts were reviews, books, posters, or conference proceedings. In the second step, articles were included if they (1) described a performance-based clinical test that was used in at least 3 studies identified in the first search, (2) evaluated one or more measurement properties in one or more of the tests described, and (3) included participants with an age or mean age between 60 and 70 years.

For the selection of articles in the first part of the study, two authors performed independent reviews of article abstracts. Discrepancies were discussed until agreement was achieved; if not, a third reviewer made the final decision. The tests detected were labelled "in-lab" when they required advanced, fixed lab equipment, or "out-of-lab", if in principal they could be performed in a home setting. Despite gait speed being a very common measure of physical performance in older adults, it is not a specific measure of balance or muscle strength, but rather considered to be a general measure of health and function [38, 39]. Therefore we included only articles with tests of gait speed if the test included one or more additional test elements that challenge the sensory system beyond that of normal or fast walking and thus require a balance reaction (i.e. dynamic, proactive or reactive). Test batteries were included if one or more of the tests in the battery was in accordance with our definition of a performance-based test of balance and/or strength.

The review of full-texts was completed by three of the authors where one reviewed all articles and two reviewed one-half each. Discrepancies were discussed with one of the other reviewers and a decision was made based on consensus. For the second part of the study, two authors each screened one-half of the abstracts and full-texts of the methodological studies.

Data extraction

Information from each full-text article was extracted into an excel sheet, containing information about the performance-based clinical tests (name of the test, measurement unit, scoring, and sample characteristics).

Results were categorized into sections representing balance or muscle strength measures. Since balance tests are task-specific, balance tests were categorized according to the framework of Shumway-Cook and Woollacoot [30, 1) static steady-state balance (i.e., maintaining a steady position in sitting or standing), including measures of postural sway obtained during quite standing (e.g. CoM sway); (2) dynamic steady-state balance (i.e., walking); (3) proactive balance (i.e., anticipating predicted disturbances such as crossing or walking around an obstacle); (4) reactive balance (i.e., compensating disturbances); and (5) results of balance test batteries. Muscle strength tests were categorized according to a previous published qualitative review [10], resulting in the following categories: (1) 1 Repetition Maximum (1RM); (2) Maximum Isometric Strength (MIS); and (3) Muscle Power.

Assessment of measurement properties

The quality of the method studies included in the second step was evaluated by three independent reviewers using the COSMIN checklist [40]. COSMIN describes how to rate the quality of the following nine categories of measurement properties: internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing, cross-cultural validity, criterion validity, and responsiveness, with several items within each category [40]. Each category is rated as "poor", "fair", "good" or "excellent", with a "worse-score-count"-approach, meaning that each category will get the lowest rating achieved for any of the items within that category [40]. As the criteria of each rating score can be different between categories, the method studies receive a rating for each measurement property assessed. Thus the quality of a study evaluating validity and reliability of a test can be rated "poor" for its assessment of validity, and "fair" for its assessment of reliability. Two amendments were made to the COSMIN guidelines. The first refers to the handling of missing cases. Because missing cases largely is an issue with questionnaires and not tests of physical performance, it was not considered relevant for the quality assessment, and thus articles were not given negative ratings for not addressing it. The second refers to sample sizes. Articles with sample sizes between 21 and 30 were rated as "fair" instead of "poor", as the sample size affects the precision of estimates rather than the quality of the methodological study itself [41].

Results

Study selection

Out of 3454 articles identified, 295 articles were included in the full-text review (Fig. 1). In total, 69 balance tests and 51 muscle strength tests were identified (Table 1; Additional file 2). Out of these tests, 26 balance tests and 15 muscle strength tests were used in ≥3 articles. These tests were included in the second search on measurement properties, and revealed only three method studies from reviewing 874 abstracts and 131 full-text articles (Fig. 2).

All studies included young seniors, where 282 studies had a sample with a mean age between 60 and 70 years

Table 1 Summarized description of balance and strength tests

Balance test	N[a]	Age
Static steady-state balance		
Side-by-side, eyes open, 10 s (8 studies)	21,419	40–87 (62.6–70.4)
Side-by-side, eyes closed, 10 s (1 study)	37	60–81 (67.7 ± 5.3)
Side-by-side, eyes open, 30 s (10 studies)	14,003	52–90 (62.7–71.6)
Side-by-Side, on foam, eyes open, 30 s (1 study)	122	69.7–71.6
Side-by-side, eyes closed, 30 s (7 studies)	364	57–75 (64.7–71.6)
Side-by-side, 60 s (1 study)	54	60+ (66.0 ± 5.0)
Semi-tandem, 10 s (6 studies)	16,926	40–87 (62.6–70.0)
Semi-tandem, 30 s (4 studies)	13,416	52–90 (62.7–65.0)
Tandem, 10 s (8 studies)	17,100	40–87 (62.6–71.6)
Tandem, 30 s (3 studies)	13,410	52–90 (64.8–65.0)
Tandem, 60 s (1 study)	12	69.0 ± 3.0
OLS (5 studies)	2266	52–84 (64.0–69.1)
OLS, no time limit (3 studies)	718	50–79 (53.9–73.1)
OLS, eyes closed, no time limit (4 studies)	391	50–79 (60.0–67.1)
OLS, 15 s (1 study)	19	60–68
OLS, 25 s (1 study)	26	59.7–60.5
OLS, 30 s (10 studies)	4773	55–84 (62.0–69.0)
OLS, eyes closed, 30 s (2 studies)	1812	60–84 (63.2–69.0)
OLS, eyes open, 45 s (1 study)	60	62.9–64.4
OLS, eyes closed, 45 s (1 study)	60	62.9–64.4
OLS, alternating eyes open and eyes closed (1 study)	557,648	66.0
OLS, 60 s (19 studies)	39,736	34–90+ (61.8–77.0)
OLS, 60 s, eyes closed (6 studies)	536	60–84 (66.3–69.4)
OLS, 120 s (1 study)	501	65–74 (69.3–69.7)
Romberg Test (5 studies)	1262	50–80 (50.8–69.0)
Sharpened Romberg (2 studies)	76	62.5–72.8
Romberg with Jendrassik maneuver (1 study)	266	65–74 (69.5 ± 3.0)
Equi Test (1 study)	55	61–83 (69.3 ± 5.5)
SOT (1 study)	23	60–78 (66.2–71.3)
CTSIB (2 studies)	61	64.0–69.0
Dynamic steady-state balance		
Tandem walk (8 studies)	260	55–85 (65.5–77.0)
Step test (2 studies)	67	53–83 (65.7–66.9)
Four Square step test (6 studies)	470	55–81 (62.0–71.5)
Step width & length, eyes open and eyes closed (1 study)	56	66.7–72.8
MSL test (2 studies)	59	60–81 (67.7–77.0)
360° turn (1 study)	282	60–74
180° turn (2 studies)	99	55+ (61.8–68.5)
6 m backwards walk (3 studies)	77	65–84 (68.9–69.7)
10-m walk under single- and dual-task condition (1 study)	54	65–80
Floor Transfer Task (1 study)	102	61.2–67.0
SEBT (2 studies)	212	65.4–68.9
Dynamic balance/agility (2 studies)	120	60–84 (66.1–69.8)

Table 1 Summarized description of balance and strength tests *(Continued)*

Narrow corridor walk (1 study)	40	69.8 ± 7.5 (60+)
Sideway walk test (1 study)	32	61.8 ± 4.6
Proactive balance		
TUG (92 studies)	61,826	46–99 (61.4–77.0)
Chair rise and walk (1 study)	39	65–85
8-ft Up and Go (27 studies)	4724	51–89 (62.1–70.1)
FRT (30 studies)	13,679	50–99 (61.5–71.3)
LRT (1 study)	28	57–73 (65.9–66.0)
7 m obstacle walk (1 study)	134	69.6–70.3
Curved walking (1 study)	1054	65.0 ± 7.0
Zigzag walking (1 study)	81	50–74 (59.0–61.0)
Reactive balance		
Reactive balance test (1 study)	102	65–80 (69.8–70.0)
Push and release test (2 studies)	54	65–80
Adaptive gait test (1 study)	20	61–81
Step Execution Test (2 studies)	72	60–88 (67.7–69.6)
Backwards stepping test (1 study)	36	65–75 (66.2–68.3)
Crossover stepping test (1 study)	36	65–75 (66.2–68.3)
Limits of Stability test (1 study)	30	64.2 ± 7.3
Performance batteries		
BBS (35 studies)	2324	56–88 (61.4–74.0)
SPPB (34 studies)	17,687	60–89 (65.4–72.3)
Tinetti Test / Performance Oriented Mobility Assessment (7 studies)	8166	55.0–97.6 (62.5–66.8)
PPT (2 studies)	91	60–83 (67.4–68.8)
FAB scale (7 studies)	308	52–89 (61.8–69.5)
CS-PFP-10 (1 study)	26	60+ (68.6–72.3)
PPB (4 studies)	2149	64.0–69.9
CBM (3 studies)	132	55–70 (66.4–69.9)
8-level balance scale (2 studies)	102	55–70 (66.4–69.9)
FMM (1 study)	90	65.3 ± 4.6
Strength test	N[a]	Age
One repetition maximum		
Handgrip strength (81 studies)	130,821	34–89 (60.4–70.5)
Shoulder flexor strength (1 study)	85	65–84 (69.0 ± 0.4)
Hip muscle strength (2 studies)	45	55–75 (63.7–68.4)
Knee extensor strength (1 study)	85	65–84 (69.0 ± 0.4)
Leg strength (6 studies)	272	55–75 (61.1–69.3)
Toe grasping strength (2 studies)	722	52–78 (66.3–67.6)
Maximal Isometric Strength (MIS)		
Elbow extensor strength (1 study)	26	69.2–70.0
Hip extensor strength (1 study)	39	60–78 (68.5–69.7)
Hip flexor strength (2 studies)	818	60–78 (68.5–69.7)
Hip abductor strength (2 studies)	744	61.8–68.7
Knee extensor strength (11 studies)	1595	60.78 (61.1–71.6)
Knee flexor strength (1 study)	39	60–78 (68.5–69.7)

Table 1 Summarized description of balance and strength tests *(Continued)*

Leg strength (6 studies)	2544	50–79 (61.4–69.0)
Ankle dorsiflexor strength (7 studies)	357	60–78 (61.8–69.7)
Ankle plantar flexor strength (5 studies)	832	50–80 (61.8–68.5)
Functional muscle power		
Upper body functional muscle power		
30 s arm curl (20 studies)	5768	51–89 (61.9–69.9)
Abdominal Strength (2 studies)	252	59–60+ (63.0–66.9)
Single forearm contractions (1 study)	32	59–85 (66.0 ± 2)
Seated medicine ball throw (1 study)	36	68.8–68.9
Lower body functional muscle power		
Five times Sit-to-Stand (61 studies)	81,289	40–90+ (58.7–71.0)
One time sit-to-stand (7 studies)	414	60–74 (61.6–69.9)
Ten times sit-to-stand (6 studies)	73,283	50–81 (62.6–69.0)
15 s Sit-to-stand (1 study)	5777	65–79 (69.8–70.1)
30 s sit-to-stand (51 studies)	7493	51–91 (61.2–71.6)
1 min sit-to-stand (2 studies)	123	55–70 (62.2–70.7)
One time kneel-to-stand (1 study)	259	60+ (67.6 ± 7.0)
Floor rise to standing (7 studies)	172	65–84 (67.0–69.3)
Five Step Test (1 study)	621	50+ (66.8–69.4)
Stair climbing (2 studies)	1143	55–79 (63.8–67.5)
Stair climbing (8 steps) (2 studies)	111	65.6–67.8
Stair climbing (10 steps) (3 studies)	212	50–75 (62.7–71.5)
Stair climbing (11 steps) (3 studies)	77	65–84 (68.9–69.3)
Stair climbing (12 steps) (2 studies)	337	45–80 (58.7–64.8)
Stair climbing (14 steps) (1 study)	30	68.5 ± 5.1
Stair climbing (15 steps) (1 study)	134	69.6–70.3
Stair ascent (23 steps) (1 study)	62	60–83 (66.6–71.0)
Stair ascent (16 steps) (1 study)	48	60–80 (68.6 ± 6.1)
Stair ascent (10 steps) (4 studies)	158	62–80 (66.0–70.0)
Stair ascent (9 steps) (2 studies)	71	62.7–70.0
Stair ascent (4 steps) (1 study)	33	60–74 (64.4–65.7)
Stair ascent (one time) (1 study)	259	60+ (67.6 ± 7.0)
Stair descent (16 steps) (1 study)	48	60–80 (68.6 ± 6.1)
Stair descent (14 steps) (1 study)	33	67.0 ± 4.5
Stair descent (10 steps) (1 study)	19	66.0 ± 1.0
Stair descent (9 steps) (1 study)	48	69.8–70.0
Stair descent (one time) (1 study)	259	60+ (67.6 ± 7.0)
Functional leg extensor strength (1 study)	1133	55–79 (63.8–64.1)
Lift and reach (1 min) (2 studies)	123	55–70 (62.6–70.7)
Standing long jump (2 studies)	98	50–79 (63.7 ± 1.1)
Squat jump (1 study)	63	65–70 (67.5 ± 0.4)
Single knee extension contractions (1 study)	32	59–85 (66.0 ± 2.0)

[a]The total number included was the total number of participants in all studies per balance/strength test; *OLS* One-leg standing balance, *SEBT* Star Excursion Balance Test, *TUG* Timed Up and Go, *FRT* Functional Reach Test, *LRT* Lateral Reach Test, *SOT* Sensory Organization Test, *BBS* Berg Balance Scale, *SPPB* Short Physical Performance Battery, *PPT* Physical Performance Test, *FAB* Fullerton Advanced Balance, *CS-PFP-10* Continuous Scale-Physical Functional Performance-10 item test, *PPB* Physical Performance Battery, *CBM* Community Balance & Mobility scale, *FFM* Functional Movement Measurement

and 13 studies [42–55] included participants with an age between 60 and 70 years exclusively.

Balance performance tests

Static steady-state balance tests

A total of 28 tests assessing static steady-state balance were identified. Single-activity measures (24 tests) were grouped into four main activity domains: (1) Side-by-side, (2) Semi tandem, (3) Tandem, and (4) One-leg-stand. Variations were found in performance within each category regarding (1) time (range 10–120 s), (2) vision (eyes open; eyes closed), (3) surface (firm; foam), and (4) number of trials (range 1–6 trials). The method of scoring included (1) total time (s), (2) category of time intervals (categorized according to the total time), (3) percentage of participants able to hold the position, and (4) body sway measures (e.g., displacement of the Center of Pressure, CoP; sway velocity).

Three Romberg tests were identified, with variations in (1) time (range 10–60s), (2) standing positions (Side-by-Side; Side-by-Side and Tandem; Side-by-Side, Semi-tandem, and Tandem), (3) vision (eyes open; eyes closed), and (4) incorporated muscle strength element (i.e., abduction of the upper limbs). The method of scoring included (1) total time (s), (2), scoring (categorized according to the total time), and (3) percentage (ability to hold the position for a pre-determined time). Four other tests identified were the Equi test, the Sensory Organization Test (SOT), the modified Clinical Test of Sensory Interaction in Balance (mCTSIB), assessing measures of body sway (e.g., CoP displacement), and the 8-level balance scale, scoring balance performance according to the ability to perform progressively challenging standing positions.

Dynamic steady-state balance tests

A total of 14 tests assessing dynamic steady-state balance were identified: (1) the tandem walk, with variations in the distance walked (9.14 m; 10 m), (2) the Step test, with variations in the demand of the activity (using the worse leg), (3) The Four Square Step Test (FSST), (4) a step width and length measuring walking test, (5) the Maximum Step Length (MSL) test, (6) the 360° turn, (7) the 180° turn, (8) the 6 m backwards walk test, (9) the 10 m walk under single- and dual-task conditions, (10) the floor transfer task, (11) the Star Excursion Balance Test (SEBT), (12) a walking test measuring dynamic balance and agility, (13) the narrow corridor walk, and (14) the sideways walk test. The method of scoring included (1) total time (s), (2) distance (step width and length), (3) number of steps, (4) number of missteps, (5) percentage (inability to complete the test), and (6) scoring (categorized according to the total time for completion of test).

Proactive balance tests

Eight tests for assessing proactive balance control were identified. The Timed Up and Go (TUG) test was used in 92 studies, with variations in (1) set pace (self-paced; fast paced), (2) distance walked (range 2.44–3.05 m), (3) turn (walk to a line on the floor and return; walk to a cone, turn around the cone and return), (4) chair (with/without armrests; with/without backrest; height range 40–46 cm), (5) number of trials (range 1–4), (6) incorporated cognitive (counting backwards; saying animal names) and motor (carrying a cup of water) tasks, and (7) outcome measure (s; m/s; step-related variables; phase-related movement analyses; accelerations). One study investigated the chair rise and walk test, and 27 studies the 8-ft Up-and-Go test, both tests evaluated by time (s). Another 30 studies investigated the Functional Reach Test (FRT), with variations in (1) number of trials (range 1–5), (2) arms (extending the right or left arm forward; raising both arms in front), (3) hands (making a fist; with fingers extended), and (4) distance (tip of the middle finger; position of the third metacarpal). The method of scoring included (1) maximum distance reached (cm; inches), and (2) percentage (maximum distance reached normalized to height). Four other tests were the Lateral Reach Test (LAT), evaluated by the maximum distance reached (cm), and the 7 m obstacle walk, the Zigzag walking test, and the Curved walking test, all three evaluated by the total time (s) [109].

Reactive balance tests

Seven tests for assessing reactive balance control were identified: (1) the Reactive Balance Test, measuring oscillations in medio-lateral and anterior-posterior directions, (2) the Push and Release Test, measuring the amount of steps needed to regain balance, (3) the adaptive gait test, measuring gait speed (m/s) and the number of step errors, (4) the Step Execution Test, measuring reaction time (ms), (5) the Backwards Stepping Test, measuring ground reaction forces (N/kg),(6) the Crossover Stepping Test, measuring ground reaction forces (N/kg), and (7) the Limits of stability test, measuring reaction time (s), movement velocity (m/s), and maximum excursion (%).

Performance test batteries/scales

Nine performance test batteries that included different balance tasks were identified: (1) the Berg Balance Scale (BBS) which was used in 35 studies, (2) the Short Physical Performance Battery (SPPB), which was investigated in 34 studies, (3) the Tinetti Performance Oriented Mobility Assessment (POMA), which was investigated in seven studies, (4) the Fullerton Advanced Balance (FAB) scale, which was investigated in seven studies, (5) the Physical Performance Test (PPT) with variations in the

number of included items (range 7–9), (6) the Continuous Scale-Physical Functional Performance-10 item (CS-PFP-10) test, (7) the Physical Performance Battery (PPB), (8) the Community Balance & Mobility (CBM) scale, and (9) the Functional Movement Measurement (FMM). All performance test batteries used a scoring scheme (e.g., 0 'unable to perform' up to 4 'able to perform the task safely') for the assessment of the performance.

Muscle strength performance

One repetition maximum tests

We identified six tests measuring the One Repetition Maximum (1 RM) of upper- and lower-body extremities. Eighty-one studies investigated handgrip strength, with variations in (1) the measurement instrument (electronic; hydraulic; bulb hand dynamometer), (2) testing position (sitting; standing), (3) demand (both hands; dominant hand; preferred hand; adjusted size for men and women), and (4) number of trials (1–3). The method of scoring included (1) force (kg; pounds; kg/bodyweight; pounds/square; Newton; kilopascal), (2) percentage (force scores, i.e., kg classified as weakness), and (5) outcome (mean of trials; best trial). Other studies used 1 RM of shoulder flexors, hip muscles, knee extensors, legs, or toes, either assessed by force (kg) or torques.

Maximum isometric strength tests

There were nine tests measuring Maximum Isometric Strength (MIS). Eleven studies used MIS tests of knee extensors, with variations in (1) outcome (mean of trials; best trial), and (2) outcome dimension (kg; N/k; percentage, i.e., muscle strength/bodyweight). Six studies evaluated leg muscle strength, assessed by force (kg). Ankle dorsiflexor MIS tests were used in seven studies, either evaluated by force (kg, N/kg) or percentage (muscle strength/bodyweight). Five studies assessed ankle plantar flexor strength by force (kg). One study included MIS tests of hip extensors, two of hip flexors and hip abductors, evaluated by force (kg) or percentage (i.e., muscle strength in relation to total bodyweight). Elbow extensor strength was measured in one study by force (kg), as well as knee flexor strength, measured by percentage (muscle strength/bodyweight).

Muscle power tests

We identified 36 muscle power tests. For upper-body extremities, four tests were identified. The 30 s Arm Curl Test was used in 20 studies, with variations in the weight used (2.0 kg for all participants; 2.27 kg for women and 3.63 kg for men). The test recorded the number of repetitions in 30 s. Abdominal muscle power was investigated in two studies and the number of repetitions in 30 s was recorded. Single forearm contractions, evaluated by Maximum Voluntary Contraction (MVC, in kg), and seated medicinal ball throws, measured by maximum distance reached (cm), were investigated in one study each.

For lower-body extremities, six versions of sit-to-stand (STS) were used in 128 studies, with variations in (1) method of measurement (time to perform one repetition; time to perform five repetitions (5STS); time to perform ten repetitions (10STS); number of repetitions in 15 s (15 s STS); 30 s (30s STS); 60 s (60s STS)), (2) chair (height: standard; adjusted; range 30–60 cm; with backrest; without backrest; without armrests), (3) position (back at the back of the chair; sitting in the middle of the chair; sitting in the front half of the chair; sitting on the edge of the chair), (4) time of measurement (starting/finishing in a sitting or standing position), (5) pace (self-paced; fast paced), (6) number of trials (range 1–3), and (7) outcome (mean of trials; best trial). The method of scoring included (1) total time (s), (2) repetitions, (3) scoring, (4) force (N/s in kg; W in kg), and (5) speed (stands per minute).

There were seven different types of stair climbing tests investigated in 11 studies with variations in (1) number of steps (standard flight of stairs; range 8–15 steps), and (2) method of measurement (time; stair climbing power; W).

Six studies investigated stair ascent, and 4 studies investigated stair descent. Tests varied in (1) number of stair steps (range 1–23) and (2) method of measurement (time; score).

Eight other tests for measuring muscle power of lower-body extremities were identified: (1) Lift and Reach, assessed by repetitions over 1 min, (2) Floor rise to standing, assessed by time (s), (3) Five Step Test, assessed by time (s), (4) One-Time Kneel-to-Stand, assessed by time (s), (5) Functional Leg Extensor Muscle Strength, assessed by the maximum weight in relation to bodyweight, (6) Standing Long Jump, assessed by distance (cm), (7) Squat jump, assessed by maximum ground reaction force (N*kg-1), rate of force development (N*kg-1), and force (N), and (8) Single Knee Extension Contractions, assessed by maximum work rate.

Assessment of measurement properties

Thirty-nine tests were used in ≥3 articles that were identified through step 1. In step 2, nine studies were identified that assessed measurement properties of four balance tests/scales (10s Tandem stance, TUG, SPPB, CBM) and one strength test (5STS). The quality assessment of these nine included method studies [42, 52, 56–63] are shown in an additional file (see Additional file 3). The quality of the study that assessed validity and reliability of the 10s Tandem stance [61] was rated "poor" according to the COSMIN checklist [40]. Four studies

assessed the measurement properties of the TUG, with their study quality rated "good" [42, 59] for measures of validity, and "poor" for measures of reliability [59, 60]. Three studies assessed measurement properties of the CBM, and for measures of validity, the quality of these studies were rated as "fair" [52, 58, 62], for internal consistency as "poor" [52], and for reliability as "good" [52, 62]. The quality of the study assessing the SPPB was rated "excellent" for validity and "good" for reliability [57] in younger seniors. For strength, the study assessing reliability of the 5STS was rated as "fair" [56].

Discussion

In the first step, this systematic review identified 120 performance-based clinical tests used to measure balance and/or muscle strength in young seniors, of which 69 measured balance and 51 measured muscle strength. The TUG (92 articles), BBS (35 articles), and SPPB (34 articles) were the most used balance tests in our sample. Different variations of STS (e.g. 5STS, 30s STS) were most often used to assess muscle strength (128 articles), with the 5STS as the most commonly used test (51 articles), followed by the 30s STS (51 studies). In the second step, ten method studies were identified for the 39 performance-based clinical tests which were most commonly used. The method studies evaluated measurement properties of the 10s Tandem stance, TUG, SPPB, CBM, and 5STS n samples of young seniors.

Proactive balance was the aspect of balance that was tested most frequently, with TUG as the most frequently used test (92 articles; 61,826 participants). This finding aligns with an earlier review that found TUG to be the most used test to predict falls in healthy community-dwelling older adults aged ≥60 years [31]. TUG is fast to perform and easy to administer, and cut-offs between 12 and 13 s have shown moderate to high sensitivity and specificity in predicting falls in older adults [42, 64]. However, the TUG is a general test of mobility that provides little or no information on underlying balance deficits [30]. Performance of TUG is a relatively complex task in terms of motor performance, including a 'sit-to-stand'-movement, walking, turning and a 'turn-to-sit'-movement, but for young seniors, the score of total duration may not be sensitive enough to reveal early signs of functional decline [20]. The instrumented version of TUG could potentially be a more useful test of balance and mobility in higher functioning groups, as more details of the quality and quantity of the performance can be obtained objectively than merely the total duration [65].

For balance performance test batteries, BBS was the most commonly used test (35 articles; 2324 participants), closely followed by the SPPB (34 articles; 17,687 participants). BBS is widely used and has been coined the "gold standard" of balance assessment tools [66]. BBS is a significant predictor for ADL disability onset in older adults aged 80 and over [67], but in samples with a mean age in the mid-seventies it suffers from ceiling effects [68–70], even in older adults with a falls history [31]. A previous systematic review recommended the SPPB as the best performance-based tool for measuring physical function in older adults due to superior qualities related to validity, reliability, and responsiveness compared to other tests [71]. This review generally reported little ceiling effects for the SPPB in the "general (mixed) population" of community-dwelling older adults. However, when applied in higher-functioning community-dwelling older adults, the SPPB also showed ceiling effects [32, 72]. Despite being extensively used in older people in general and receiving appraisals for its measurement properties, the BBS and SPPB do not appear to be good enough for assessing physical performance in well-functioning young seniors due to ceiling effects. In this review, the method study assessing the measurement properties of the SPPB was rated "excellent" for its measure of validity and "good" for its measure of reliability [57]. However, the result of the method studies are not considered in this quality rating, but relatively high mean scores on the SPPB in this study (9.7 ± 2.0) align with the findings of other studies in healthy young seniors [32, 72].

The most frequently used muscle strength test across all categories were those including some variation of the 'sit-to-stand'-movement (128 studies), with the 5STS (61 articles; 81,289 participants) and the 30s STS (51 articles; 7493 participants) being the most popular among them.

The 5STS is commonly used as a test of physical performance in clinical assessments [73], and is also part of the SPPB test battery. We found a large variety in how this test was administered, thus making comparisons between versions a challenge. In the original and most applied protocol, the subject is "timed from the initial sitting position to the final standing position at the end of the fifth stand" [74]. In an earlier meta-analysis, the mean score on 5STS from 4184 participants between 60 and 69 years was 11.4 s [75]. This is relatively fast compared to identified cut-offs of 13.6 s for indication of increased disability and morbidity [76], and 15 s for predicting recurrent fallers [77]. However, as also this test lacks validation in young seniors, we have no basis for recommending this performance-based clinical test as a good measure for this specific population.

The second most used tool with a STS-variation was the 30s STS, originally developed to overcome floor effects of the 5STS [78]. We did not identify any method study that assessed the measurement properties of 30s STS, but in community-dwelling adults with a mean age of 70.5 ± 5.5 years, the test-retest reliability (ICC .89) and

concurrent validity was moderate, with associations with weight-adjusted 1 RM leg-press of $r = .71$ (women) and .78 (men) [78]. Therefore, the 30s STS could be suitable to measure physical performance in young seniors, but further studies are warranted to confirm this.

In the second step, nine method studies were identified, with only four out of 26 balance tests and one out of 13 strength tests having been used in ≥3 articles. It is apparent that very few of all available tests for measuring balance and/or strength have been assessed for their measurement properties in healthy young seniors. The quality of most of the method studies rated in this review ranged only from "poor" to "fair". However, there seems to be a shift in focus towards the current target group in the literature, as indicated by the high number of new studies that was identified in the updated literature search (Figs. 1 and 2).

The CBM and the 10s Tandem Stance were two of the tests that emerged as being used in ≥3 studies in the updated search. Therefore, these tests were added to the updated search of method studies. In two of three method studies assessing the CBM [52, 58], the measures of reliability were all high (>.97) and validity good to excellent in young seniors [52, 58]. However, study quality was rated "poor" with regard to validity measures with the COSMIN checklist. The studies assessing the CBM reported no ceiling effects in young seniors due to its challenging, higher level tasks [52, 58], and the CBM could be considered a feasible tool to adequately assess balance performance in healthy, higher functioning young seniors. The study assessing the 10s Tandem Stance found that valid and reliable measures of the Centre of Pressure (COP) can be obtained from a Wii Balance Board (WBB), compared to a laboratory force plate [61]. Such a device could be a suitable tool for a home-based assessment of balance/posture measures. However, COP measures as assessed by the WBB have not been evaluated in younger seniors so far.

New method studies of tests that were already included before the updated search, such as TUG, SPPB, and 5STS, indicate that not only new tests, but also well-established tests are evaluated for their potential suitability in measuring balance and/or strength in young seniors. The TUG showed excellent reliability, but both studies were rated as "poor" regarding their overall methodological quality [59, 60]. Another study, rated "good" according to COSMIN, found cut-off scores of 12.47 s on the TUG to be an accurate measure for screening of fall risk [42], while another study reported low discriminative ability of the TUG for healthy older adults vs. older adults with a history of falls [63], which is in line with previous findings concluding that the TUG is able to discriminate between fallers and multiple fallers, but not between non-fallers and fallers [79].

Based on the findings in this review, there seems to be only one promising scale for adequately assessing balance in healthy young seniors, i.e. showing no ceiling effects and having measures of high validity and reliability, namely the CBM, However, important measures such as responsiveness to identify intervention-related changes are currently lacking for this balance scale.

A limitation of this systematic review is the restriction to English written articles which might have influenced the final number of identified tests. However, this review was based on a broadly designed literature search which aimed at getting a broad overview of existing performance-based clinical tests used for measuring balance and/or muscle strength in young seniors. Due to the large number of identified and included articles, our search is unlikely to have missed any frequently used tests.

Conclusion

This systematic review identified a large number of performance-based clinical tests that have been used to measure balance and/or muscle strength in young seniors. The most commonly used balance tests suffer from ceiling effects in young seniors. Additionally, there is a wide variety and hence lack of consensus on how to administer balance and muscle strength tests, and how to report their outcomes. There is a need for guidance on how to administer and conduct balance and strength tests to improve their informative value and comparability of outcomes. Only nine method studies were identified that assessed the measurement properties of tests used in young seniors, indicating that more studies are required to identify suitable tests for assessing balance and strength in young seniors. Only in the last 2 years, three studies assessing the measurement properties of the CBM in healthy young seniors have been identified, indicating that it could be a promising tool to adequately measure balance. The CBM has a standardised assessment procedure and studies show that it is the only scale applied in young seniors not showing ceiling effects [52, 58], being more challenging and thus more sensitive to detect changes in balance performance in healthy younger seniors. However, more research is needed to further analyse its measurement properties, especially in terms of responsiveness and sensitivity to change [52, 58, 62].

In general, more challenging tests are needed to adequately assess young senior's physical performance, especially when aiming to identify early declines in function so that preventive strategies can be initiated in a timely manner.

Additional files

Additional file 1: Database search. Brief description: includes all search strings for MEDLINE and EMBASE for both, part 1, i.e., identifying existing tests and part 2, i.e., identifying methodological studies for identified tests which have been used in ≥3 studies (identified thorugh part 1). (DOCX 15 kb)

Additional file 2: Description of balance and strength tests. Brief description: Large table which contains all identified balance and strength tests with detailed description of test administration, scale design, and study population. (DOCX 1691 kb)

Additional file 3: The quality of studies assessing validity and/or reliability of included balance and strength tools and the rating of the reported results. Brief description: Overview of the identified methodological studies. (DOCX 28 kb)

Abbreviations

ADL: Activities of Daily Living; BBS: Berg Balance Scale; CBM: Community Balance and Mobility Scale; COM: Center of Mobility; COP: Center of Pressure; COSMIN: Consensus-based Standards for selection of health Measurement Instruments; CS-PFP-10: Continous Scale-Physical-Functional-Performance-10 item test; FAB: Fullerton Advanced Balance Scale; FMM: Functional Mobility Measurement; FRT: Functional Reach Test; FSST: Four Scare Step Test; LRT: Lateral Reach Test; mCTSIB: modified Clinical Test of Sensory Interaction in Balance; MIS: Maximum Isometric Contraction; MSL: Maximal Step Length; MVC: Maximum Voluntary Contraction; POMA: Performance Oriented Mobility Assessment; PPB: Physical Performance Battery; PPT: Physical Performance Test; RM: Repetition Maximum; SEBT: Star Excursion Balance Test; SOT: Sensory Organization Test; SPPB : Short Physical Performance Battery; STS: Sit to Stand; TUG: Timed Up and Go; WBB: Wii Balance Board

Acknowledgements

The authors wish to thank Nini Jonkman for her contribution to the design of this review. We thank Ingrid I. Riphagen for her assistance in conducting the literature search.

Funding

This study was supported by a doctoral scholarship from the Klaus Tschira Foundation (KTS), and PreventIT which received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 689238. The content is solely the responsibility of the authors and does not necessarily represent the official views of the KTS or the European Commission.

Availability of data and materials

Not applicable.

Authors' contributions

RB and MW performed the review of abstracts. RB, MW, SU, and KT performed the review of full-texts and data extraction. KT assisted in decision making. RB and MW drafted the manuscript. KT, MS, JLH, and BV helped to revise the manuscript. All authors have read and approved the final version of the manuscript, and agree with the order of presentation of the authors.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway. [2]Network Aging Research (NAR), Heidelberg University, Heidelberg, Germany. [3]Institute of Sports and Sports Sciences, Heidelberg University, Heidelberg, Germany.

Received: 6 August 2018 Accepted: 11 December 2018
Published online: 09 January 2019

References

1. Ferrucci L, Cooper R, Shardell M, Simonsick EM, Schrack JA, Kuh D. Age-related change in mobility: perspectives from life course epidemiology and geroscience. J Gerontol A. 2016;71(9):1184–94.
2. Hardy R, Cooper R, Sayer AA, Ben-Shlomo Y, Cooper C, Deary IJ, Demakakos P, Gallacher J, Martin RM, McNeill G. Body mass index, muscle strength and physical performance in older adults from eight cohort studies: the HALCyon programme. PLoS One. 2013;8(2):e56483.
3. Sousa LM, Marques-Vieira CM, Caldevilla MN, Henriques CM, Severino SS, Caldeira SM. Risk for falls among community-dwelling older people: systematic literature review. Rev Gaucha Enferm. 2016;37(4):e55030.
4. Sturnieks DL, St George R, Lord SR. Balance disorders in the elderly. Neurophysiol Clin. 2008;38(6):467–78.
5. Terroso M, Rosa N, Marques AT, Simoes R. Physical consequences of falls in the elderly: a literature review from 1995 to 2010. Eur Rev Aging Phys Act. 2014;11(1):51–9.
6. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: a review of the literature. Maturitas. 2013;75(1):51–61.
7. Idland G, Pettersen R, Avlund K, Bergland A. Physical performance as long-term predictor of onset of activities of daily living (ADL) disability: a 9-year longitudinal study among community-dwelling older women. Arch Gerontol Geriatr. 2013;56(3):501–6.
8. Williams ME, Gaylord SA, Gerrity MS. The timed manual performance test as a predictor of hospitalization and death in a community-based elderly population. J Am Geriatr Soc. 1994;42(1):21–7.
9. Era P, Sainio P, Koskinen S, Haavisto P, Vaara M, Aromaa A. Postural balance in a random sample of 7,979 subjects aged 30 years and over. Gerontology. 2006;52(4):204–13.
10. Granacher U, Mühlbauer T, Gruber M. A qualitative review of balance and strength performance in healthy older adults: impact fortesting and training. J Aging Res. 2012;2012(2012):708905.
11. Perna FM, Coa K, Troiano RP, Lawman HG, Wang C-Y, Li Y, Moser RP, Ciccolo JT, Comstock BA, Kraemer WJ. Muscular grip strength estimates of the US population from the national health and nutrition examination survey 2011–2012. J Strength Cond Res. 2016;30(3):867–74.
12. Peterson MD, Krishnan C. Growth charts for muscular strength capacity with quantile regression. Am J Prev Med. 2015;49(6):935–8.
13. Choy NL, Brauer S, Nitz J. Changes in postural stability in women aged 20 to 80 years. J Gerontol A Biol Sci Med Sci. 2003;58:525–30.
14. Isles RC, Choy NL, Steer M, Nitz JC. Normal values of balance tests in women aged 20–80. J Am Geriatr Soc. 2004;52:1367–72.
15. Landi F, Calvani R, Tosato M, Martone AM, Fusco D, Sisto A, Ortolani E, Savera G, Salini S, Marzetti E. Age-related variations of muscle mass, strength, and physical performance in community-dwellers: results from the Milan EXPO survey. J Am Med Dir Assoc. 2017;18(1):1.e1–8.
16. Ekdahl C, Jarnlo GB, Andersson SI. Standing balance in healthy subjects. Evaluation of a quantitative test battery on a force platform. Scand J Rehabil Med. 1989;21(4):187–95.
17. Hytönen M, Pyykkö I, Aalto H, Starck J. Postural control and age. Acta Otolaryngol. 1993;113(2):119–22.
18. Justice JN, Cesari M, Seals DR, Shively CA, Carter CS. Comparative approaches to understanding the relation between aging and physical function. J Gerontol A Biol Sci Med Sci. 2016;71(10):1243–53.
19. Cooper R, Kuh D, Hardy R. Objectively measured physical capability levels and mortality: systematic review and meta-analysis. BMJ. 2010;341:c4467.
20. Perera S, Patel KV, Rosano C, Rubin SM, Satterfield S, Harris T, Ensrud K, Orwoll E, Lee CG, Chandler JM. Gait speed predicts incident disability: a pooled analysis. J Gerontol A Biol Sci Med Sci. 2015;71(1):63–71.
21. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, Brach J, Chandler J, Cawthon P, Connor EB, et al. Gait speed and survival in older adults. JAMA. 2011;305(1):50–8.

22. Studenski S, Perera S, Wallace D, Chandler JM, Duncan PW, Rooney E, Fox M, Guralnik JM. Physical performance measures in the clinical setting. J Am Geriatr Soc. 2003;51(3):314–22.
23. Langley FA, Mackintosh SFH. Functional balance assessment of older community dwelling adults: a systematic review of the literature. Internet J Allied Health Sci Pract. 2007;5(4):13.
24. Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med. 2010;46(2):239–48.
25. Cooper R, Hardy R, Sayer AA, Kuh D. A life course approach to physical capability. In: Kuh D, Cooper R, Hardy R, Richards M, Ben-Shlomo Y, editors. A life course approach to healthy aging. 1st ed. London: Oxford University Press; 2014. p. 16–31.
26. den Ouden MEM, Schuurmans MJ, Arts IE, van der Schouw YT. Physical performance characteristics related to disability in older persons: a systematic review. Maturitas. 2011;69(3):208–19.
27. Granacher U, Hortobágyi T. Exercise to improve mobility in healthy aging. Sports Med. 2015;45(12):1625–6.
28. Tak E, Kuiper R, Chorus A, Hopman-Rock M. Prevention of onset and progression of basic ADL disability by physical activity in community dwelling older adults: a meta-analysis. Ageing Res Rev. 2013;12(1):329–38.
29. Wang BWE, Ramey DR, Schettler JD, Hubert HB, Fries JF. Postponed development of disability in elderly runners: a 13-year longitudinal study. Arch Intern Med. 2002;162(20):2285–94.
30. Shumway-Cook A, Woollacott MH. Motor control: translating research into clinical practice. 3rd ed. Philadelphia: Lippincott, Williams & Wilkins; 2007.
31. Power V, Van De Ven P, Nelson J, Clifford AM. Predicting falls in community-dwelling older adults: a systematic review of task performance-based assessment tools. Psychother Pract Res. 2014;35(1):3–15.
32. Fleig L, McAllister MM, Chen P, Iverson J, Milne K, McKay HA, Clemson L, Ashe MC. Health behaviour change theory meets falls prevention: feasibility of a habit-based balance and strength exercise intervention for older adults. Psychol Sport Exerc. 2016;22:114–22.
33. Hackney ME, Earhart GM. Effects of dance on gait and balance in Parkinson's disease: a comparison of partnered and nonpartnered dance movement. Neurorehabil Neural Rep. 2010;24(4):384–92.
34. Rose DJ, Lucchese N, Wiersma LD. Development of a multidimensional balance scale for use with functionally independent older adults. Arch Phys Med Rehabil. 2006;87(11):1478–85.
35. Balasubramanian CK. The community balance and mobility scale alleviates the ceiling effects observed in the currently used gait and balance assessments for the community-dwelling older adults. J Ger Phys Ther. 2015;38(2):78–89.
36. Howe J, Inness E, Venturini A, Williams JI, Verrier MC. The community balance and mobility scale - a balance measure for individuals with traumatic brain injury. Clin Rehabil. 2006;20:885–95.
37. Yamada T, Demura S. Effectiveness of sit-to-stand tests for evaluating physical functioning and fall risk in community-dwelling elderly. Hum Perform Meas. 2015;12:1–7.
38. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, Brach J, Chandler J, Cawthon P, Connor EB. Gait speed and survival in older adults. JAMA. 2011;305(1):50–8.
39. Viccaro LJ, Perera S, Studenski SA. Is timed up and go better than gait speed in predicting health, function, and falls in older adults? J Am Geriatr Soc. 2011;59(5):887–92.
40. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, De Vet HCW. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539–49.
41. Benfer KA, Weir KA, Boyd RN. Clinimetrics of measures of oropharyngeal dysphagia for preschool children with cerebral palsy and neurodevelopmental disabilities: a systematic review. Dev Med Child Neurol. 2012;54(9):784–95.
42. Alexandre TS, Meira DM, Rico NC, Mizuta SK. Accuracy of timed up and go test for screening risk of falls among community-dwelling elderly. Braz J Phys Ther. 2012;16(5):381–8.
43. Chun SH, Cho B, Yang H-K, Ahn E, Han MK, Oh B, Shin DW, Son KY. Performance on physical function tests and the risk of fractures and admissions: findings from a national health screening of 557,648 community-dwelling older adults. Arch Gerontol Geriatr. 2017;68:174–80.
44. Ćwirlej-Sozańska A, Wiśniowska-Szurlej A, Wilmowska-Pietruszyńska A, Sozański B, Wołoszyn N. Assessment of psychophysical capacities for professional work in late middle age and at the beginning of old age. Med Pr. 2018;69(4):375–81.
45. Fife E, Kostka J, Kroc Ł, Guligowska A, Pigłowska M, Sołtysik B, Kaufman-Szymczyk A, Fabianowska-Majewska K, Kostka T. Relationship of muscle function to circulating myostatin, follistatin and GDF11 in older women and men. BMC Geriatr. 2018;18(1):200.
46. Hoang OTT, Jullamate P, Piphatvanitcha N, Rosenberg E. Factors related to fear of falling among community-dwelling older adults. J Clin Nurs. 2017;26(1–2):68–76.
47. Paes T, Belo LF, da Silva DR, Morita AA, Donaria L, Furlanetto KC, Sant'Anna T, Pitta F, Hernandes NA. Londrina Activities of Daily Living Protocol: Reproducibility, Validity, and Reference Values in Physically Independent Adults Age 50 Years and Older. Respir Care. 2017;62(3):298-306.
48. Palmer RC, Batra A, Anderson C, Page T, Vieira E, Seff L. Implementation of an Evidence-Based Exercise Program for Older Adults in South Florida. Journal of Aging Research. 2016;2016:7. https://doi.org/10.1155/2016/9630241.
49. Ryan A, Murphy C, Boland F, Galvin R, Smith SM. What is the impact of physical activity and physical function on the development of multimorbidity in older adults over time? A population-based cohort study. J Gerontol A Biol Sci Med Sci. 2018;73(11):1538–44.
50. Santoni G, Angleman SB, Ek S, Heiland EG, Lagergren M, Fratiglioni L, Welmer A-K. Temporal trends in impairments of physical function among older adults during 2001-16 in Sweden: towards a healthier ageing. Age and Ageing. 2018;47(5):698-704.
51. Schilling BK, Falvo MJ, Karlage RE, Weiss LW, Lohnes CA, Chiu LZ. Effects of unstable surface training on measures of balance in older adults. J Strength Cond Res. 2009;23(4):1211–6.
52. Weber M, Van Ancum J, Bergquist R, Taraldsen K, Gordt K, Mikolaizak AS, Nerz C, Pijnappels M, Jonkman NH, Maier AB. Concurrent validity and reliability of the community balance and mobility scale in young-older adults. BMC Geriatr. 2018;18(1):156.
53. Xie YJ, Liu EY, Anson ER, Agrawal Y. Age-related imbalance is associated with slower walking speed: an analysis from the National Health and nutrition examination survey. J Geriatr Phys Ther. 2017;40(4):183–9.
54. Yiengprugsawan V, Steptoe A. Impacts of persistent general and site-specific pain on activities of daily living and physical performance: A prospective analysis of the English Longitudinal Study of Ageing. Geriatrics & Gerontology International. 2018;18(7):1051-1057.
55. Zhang W, Schwenk M, Mellone S, Paraschiv-Ionescu A, Vereijken B, Pijnappels M, Mikolaizak A, Boulton E, Jonkman N, Maier A. Complexity of daily physical activity is more sensitive than conventional metrics to assess functional change in younger older adults. Sensors. 2018;18(7):2032.
56. Cani KC, Silva I, Karloh M, Gulart AA, Matte DL, Mayer AF. Reliability of the five-repetition sit-to-stand test in patients with chronic obstructive pulmonary disease on domiciliary oxygen therapy. Physiother Theory Pract. 2018;1:1-7.
57. Gómez JF, Curcio C-L, Alvarado B, Zunzunegui MV, Guralnik J. Validity and reliability of the short physical performance battery (SPPB): a pilot study on mobility in the Colombian Andes. Colombia Medica. 2013;44(3):165–71.
58. Gordt K, Mikolaizak AS, Nerz C, Barz C, Gerhardy T, Weber M, Becker C, Schwenk M. German version of the community balance and mobility scale. Z Gerontol Geriatr. 2018:1–9. https://doi.org/10.1007/s00391-018-1374-z.
59. Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the timed "up & go" test in people with Parkinson disease. Phys Ther. 2001;81(2):810–8.
60. Ng SS, Hui-Chan CW. The timed up & go test: its reliability and association with lower-limb impairments and locomotor capacities in people with chronic stroke. Arch Phys Med Rehabil. 2005;86(8):1641–7.
61. Scaglioni-Solano P, Aragón-Vargas LF. Validity and reliability of the Nintendo Wii balance board to assess standing balance and sensory integration in highly functional older adults. Int J Rehabil Res. 2014;37(2):138–43.
62. Takacs J, Garland SJ, Carpenter MG, Hunt MA. Validity and reliability of the community balance and mobility scale in individuals with knee osteoarthritis. Phys Ther. 2014;94(6):866–74.
63. Yingyongyudha A, Saengsirisuwan V, Panichaporn W, Boonsinsukh R. The mini-balance evaluation systems test (mini-BESTest) demonstrates higher accuracy in identifying older adult participants with history of falls than do the BESTest, berg balance scale, or timed up and go test. J Geriatr Phys Ther. 2016;39(2):64–70.
64. Olsson Möller U, Kristensson J, Midlöv P, Ekdahl C, Jakobsson U. Predictive validity and cut-off scores in four diagnostic tests for falls–a study in frail older people at home. Phys Occup Ther Geriatr. 2012;30(3):189–201.

65. Mellone S, Tacconi C, Chiari L. Validity of a smartphone-based instrumented timed up and go. Gait Posture. 2012;36(1):163–5.
66. Southard V, Dave M, Davis MG, Blanco J, Hofferber A. The multiple tasks test as a predictor of falls in older adults. Gait Posture. 2005;22(4):351–5.
67. Huang S-C, Lu T-W, Chen H-L, Wang T-M, Chou L-S. Age and height effects on the center of mass and center of pressure inclination angles during obstacle-crossing. Med Eng Phys. 2008;30(8):968–75.
68. Boulgarides LK, McGinty SM, Willett JA, Barnes CW. Use of clinical and impairment-based tests to predict falls by community-dwelling older adults. Phys Ther. 2003;83(4):328–39.
69. Brauer SG, Burns YR, Galley P. A prospective study of laboratory and clinical measures of postural stability to predict community-dwelling fallers. J Gerontol Ser A Biol Med Sci. 2000;55(8):M469–76.
70. Herman T, Giladi N, Hausdorff JM. Properties of the 'timed up and go'test: more than meets the eye. Gerontology. 2011;57(3):203–10.
71. Freiberger E, De Vreede P, Schoene D, Rydwik E, Mueller V, Frändin K, Hopman-Rock M. Performance-based physical function in older community-dwelling persons: a systematic review of instruments. Age Ageing. 2012; 41(6):712–21.
72. Pardasaney PK, Latham NK, Jette AM, Wagenaar RC, Ni P, Slavin MD, Bean JF. Sensitivity to change and responsiveness of four balance measures for community-dwelling older adults. Phys Ther. 2012;92(3):388–97.
73. Goldberg A, Chavis M, Watkins J, Wilson T. The five-times-sit-to-stand test: validity, reliability and detectable change in older females. Aging Clin Exp Res. 2012;24(4):339–44.
74. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol. 1994;49(2):M85–94.
75. Bohannon RW. Reference values for the five-repetition sit-to-stand test: a descriptive meta-analysis of data from elders. Percept Mot Skills. 2006;103(1):215–22.
76. Guralnik JM, Ferrucci L, Pieper CF, Leveille SG, Markides KS, Ostir GV, Studenski S, Berkman LF, Wallace RB. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. J Gerontol A. 2000;55(4):M221–31.
77. Buatois S, Miljkovic D, Manckoundia P, Gueguen R, Miget P, Vançon G, Perrin P, Benetos A. Five times sit to stand test is a predictor of recurrent falls in healthy community-living subjects aged 65 and older. J Am Geriatr Soc. 2008;56(8):1575–7.
78. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Res Q Exerc Sport. 1999; 70(2):113–9.
79. Schoene D, Wu SMS, Mikolaizak AS, Menant JC, Smith ST, Delbaere K, Lord SR. Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. J Am Geriatr Soc. 2013;61(2):202–8.

Additional file 1

Database searches

Search for performance-based tests in MEDLINE (from 1946) to 5 November 2018 (last update).

1. ((young or younger or early) adj2 (retired or retirement or elderly or senior*1 or elder*1)).ti,ab.
2. (older adult*1 or older healthy adult*1 or older active adult*1 or older healthy individual*1 or older active individual*1 or older active men or older active women or older healthy men or older healthy women).ti,ab.
3. ((year*1 or age or aged) and ("50-70" or "50-65" or "51-69" or "55-70" or "55-69" or "60-70" or "60-65" or "61-69")).ab.
4. 1 or 2 or 3
5. muscle strength/ or movement/ or motor activity/ or physical exertion/ or physical endurance/ or exercise tolerance/ or physical fitness/ or postural balance/
6. (measured or measurement* or measuring or assess* or test*1 or scale*1).ti,ab. or Geriatric assessment/ or Anthropometry/ or outcome*.mp.
7. (fitness or physical function or physical performance or balance or strength).ti,ab.
8. limit 7 to medline
9. 7 not 8
10. (gait or leg*1 or walking or walk or knee or knees or postural sway or stand or standing or lower extremit* or lower limb*1).mp. or (go or step or steps or stepping).ti,ab.
11. (4 and (5 or 9) and 6 and 10) not animals/

Search for methodological studies of identified performance-based tests in MEDLINE (from 1946) and EMBASE (from 1974) to 23 November 2018 (last update).

1. (tandem walk* or tandem stand* or (side-by-side and (stand or feet or standing)) or feet together or semi-tandem or one leg* stand* or step test or timed up go or "8 foot up" or eight foot or functional reach or (grip strength and (measur* or test* or assess*)) or arm curl or sit to stand or chair stand or chair rise or stair climbing or stair ascent or isometric strength or handheld dynamomet* or performance oriented mobility scale or tinetti or fullerton advanced balance scale or berg balance or short physical performance battery).m_titl.
2. Observer variation/ or "Predictive value of tests"/ or Psychometrics/ or psychometr*.ti. or Reference Values/ or exp "Reproducibility of Results"/ or "Sensitivity and Specificity"/ or Validation studies.pt. or Evaluation Studies.pt. or accura*.ti. or clinimetr*.ti. or consisten*.ti. or develop*.ti. or discrimina*.ti. or feasib*.ti. or predictiv*.ti. or propert*.ti. or psychometr*.ti. or reliab*.ti. or repeatab*.ti. or reproducib*.ti. or responsive*.ti. or sensitiv*.ti. or specificity*.ti. or subscale*.ti. or suitab*.ti. or test-retest.ti,ab. or useful*.ti. or utility.ti. or valid*.ti. or varia*.ti.
3. 1 and 2

Search for methodological studies of newly identified performance-based tests in MEDLINE (from 1946) and EMBASE (from 1974) to 23 November 2018 (last update).

1. tandem stan* or standing balance or short physical performance battery or SPPB or ankle dorsiflexor or floor transfer or sit* ris* test or "community balance and mobility scale" or "timed up and go" or "instrumented timed up and go" or itug
2. Observer variation/ or "Predictive value of tests"/ or Psychometrics/ or psychometr*.ti. or Reference Values/ or exp "Reproducibility of Results"/ or "Sensitivity and Specificity"/ or Validation studies.pt. or Evaluation Studies.pt. or accura*.ti. or clinimetr*.ti. or consisten*.ti. or develop*.ti. or discrimina*.ti. or feasib*.ti. or predictiv*.ti. or propert*.ti. or psychometr*.ti. or reliab*.ti. or repeatab*.ti. or reproducib*.ti. or responsive*.ti. or sensitiv*.ti. or specificity*.ti. or subscale*.ti. or suitab*.ti. or test-retest.ti. or useful*.ti. or utility.ti. or valid*.ti. or varia*.ti.
3. 1 and 2

Additional file 2 Description of included balance and strength tests

Test characteristics			Scale		Study population		
Balance test	**Detailed Description**	**Unit**	**Level**	**Items**	**N[a]**	**Age**	**Sex[b]**
Static steady-state balance							
Side-by-side, eyes open, 10 s (8 studies)	Holding the position [1-7] Three trials [8]	Time (s) [3, 4, 6] Score [1, 5] Sway velocity CoG (degrees/s) [8] % of participants able to hold the position [2, 7]	N; O; R	1	21419	40-87 (62.6-70.4)	12005 F, 9294 M
Side-by-side, eyes closed, 10 s (1 study)	Three trials [8]	Sway velocity CoG (degrees/s)	R	1	37	60-81 (67.7±5.3)	28 F, 9 M
Side-by-side, eyes open, 30 s (10 studies)	Holding the position, eyes open [9-11]; with and without cognitive task, two trials [12] Comfortable foot position with their arms by their side [13] Habitual standing posture with arms by their side, feet hip-width apart, in their usual toe-out angle, looking straight ahead at a dot positioned at eye level on a plain background, approximately 1 m away [14] Two trials (with audio-biofeedback; without audio-biofeedback) [15] Two trials (comfortable stance with eyes open; narrow stance with eyes open) [16, 17] Three trials (eyes open; visual feedback about the performance from the computer screen; following visual cues on the computer screen to move the body to hit targets identified on the screen) [18] Twelve 30-s trials, randomly completed: three with no postural threat, nine with a possible perturbation (a push forward or pull backward to the upper trunk by the examiner; three of nine	Time (s) [9, 11] CoP displacement (mm [16, 17]; cm [10, 13, 14]) % of time that the trunk tilt within specified angle limits; RMS and MPF of trunk tilt [15] Postural control, movement time, path length [18] Trunk roll/trunk pitch angle and velocity [19] Velocity (cm/s) [14] ML, AP, vertical acceleration and RMS [12]	R	1-12	14003	52-90 (62.7-71.6)	6793 F, 7142 M

	with a perturbation after 30s, six of nine with a perturbation at 1s, 5s, 10s, 15s, 20s or 25s [19]						
Side-by-Side, on foam, eyes open, 30 s (1 study)	Holding the position [10]	CoP displacement (cm)	R	1	122	69.7-71.6	90 F, 32 M
Side-by-side, eyes closed, 30 s (7 studies)	Holding the position [10, 18]; with and without cognitive task, two trials [12] Two trials (comfortable and narrow stance) [16, 17] Comfortable foot position with their arms by their side [13]	CoP displacement (mm [16, 17]; cm [10, 13]) Postural control, movement time, and path length [18] ML, AP, vertical acceleration and RMS [12]	R	1-2	364	57-75 (64.7-71.6)	258 F, 138 M
Side-by-side, 60 s (1 study)	Two trials [20]	CoP displacement (cm)	R	1	54	60+ (66.0±5.0)	30 F, 24 M
Semi-tandem, 10 s (6 studies)	Holding the position [2-7]	Time (s) [3, 4, 6] Score [5] % of participants able to hold the position [2, 7]	N; O; R	1	16926	40-87 (62.6-70.0)	9482 F, 7378 M
Semi-tandem, 30 s (4 studies)	Holding the position [9, 12, 21] Two trials (with audio-biofeedback; without audio-feedback) [15]	Time (s) [9, 12, 21] % of time that the trunk tilt within specified angle limits; RMS and MPF of trunk tilt [15]	R	1-2	13416	52-90 (62.7-65.0)	6444 F, 6964 M
Tandem, 10 s (8 studies)	Holding the position [2-4, 6, 7, 10, 22, 23]	Time (s) [3, 6] CoP displacement (mm [4]; cm [10, 22]) % of participants able to hold the position [2, 7, 23]	N; R	1	17100	40-87 (62.6-71.6)	9494 F, 7503 M

Tandem, 30 s (3 studies)	Holding the position [9, 12, 21]	Time (s)	R	1	13410	52-90 (64.8-65.0)	6444 F, 6964 M
Tandem, 60 s (1 study)	Three trials [24]	Time (s)	R	1	12	69.0±3.0	12 F, 0 M
OLS (5 studies)	N/A [25-29]	Time (s) [25-29] Score [28] n (%) balance lost <5s [27]	N; O; R	1	2266	52-84 (64.0-69.1)	1197 F, 1069 M
OLS, no time limit (3 studies)	On the dominant leg [30] On the right leg [31] Three trials [32]	Time (s)	R	1	718	50-79 (53.9-73.1)	409 F, 309 M
OLS, eyes closed, no time limit (4 studies)	Holding the position as long as possible on the dominant leg with eyes closed [33, 34] Three trials with eyes closed [35] Lifting one leg from the floor as long as possible [36]	Time (s)	R	1	391	50-79 (60.0-67.1)	176 F, 38 M
OLS, 15 s (1 study)	Three trials on each leg with eyes open and eyes closed [37]	CoP displacement (cm)	R	2	19	60-68	9 F, 10 M
OLS, 25 s (1 study)	Two trials on each leg [38]	n (%) able to hold 20 s	N	2	26	59.7-60.5	18 F, 8 M
OLS, 30 s (10 studies)	Holding the position [12, 39-42]; average of three trials [43]; best of two trials [44] Holding the position on each leg [45]; two trials each leg, best of all four trials [46] Three trials on each leg, eyes open [47] Dominant limb, contralateral knee remaining at 90°, arms folded across the chest, head straight [48]	Time (s)	R	1-2	4773	55-84 (62.0-69.0)	1677 F, 1844 M
OLS, eyes closed, 30 s (2 studies)	Three trials on each leg [47, 49]	Time (s)	R	1-2	1812	60-84 (63.2-69.0)	927 F, 885 M
OLS, eyes open, 45 s (1 study)	Arms crossed over the chest [50]	Time (s)	R	1	60	62.9-64.4	53 F, 7 M

OLS, eyes closed, 45 s (1 study)	Arms crossed over the chest [50]	Time (s)	R	1	60	62.9-64.4	53 F, 7 M
OLS, alternating eyes open and eyes closed, 45 s (1 study)	Arms crossed over the chest; mean and best of three trials [51]	Time (s); categorization as normal (>9s eyes open; >5s eyes closed)	N; R	2	557648	66	N/A
OLS, 60 s (19 studies)	Holding the position on the dominant leg [52-55] Holding the position, eyes open [56-58] One trial on each leg [59, 60] Two trials on the preferred leg [8, 61] Two trials on each leg [16, 17, 62-64] Three trials (two trials on the preferred leg; one trial on the opposite leg) [24] On the preferred leg, opposite knee flexed, arms at their sides; best of two trials [52] Standing on solid surface (floor) or compliant surface (double-folded 2 cm thick exercise mat) [65]	Time (s)	R	1-2	39736	34-90+ (61.8-77.0)	21542 F, 18239 M
OLS, 60 s, eyes closed (6 studies)	Holding the position [56, 57, 66, 67] Two trials on each leg [68] Standing on solid surface (floor) or compliant surface (double-folded 2 cm thick exercise mat) [65]	Time (s)	R	1-2	536	60-84 (66.3-69.4)	211 F, 291 M
OLS, 120 s (1 study)	One trial on each leg [69]	Time (s)	R	2	501	65-74 (69.3-69.7)	279 F, 222 M
Romberg Test (5 studies)	N/A [39] Feet together and tandem stand, both stances with eyes open and eyes closed, 10 s [70] Parallel, semi-tandem and tandem stand, 10 s [4] Standing with both feet together and eyes closed, 60 s [71] Four progressively challenging conditions; in the fourth condition, participants had to maintain balance on a foam-padded surface with their	Time (s) [39, 70, 71] Score [4] Pass/fail (%)	N; O; R	1-4	1262	50-80 (50.8-69.0)	215 F, 181 M

	eyes closed, thereby reducing visual and proprioceptive inputs and increasing their reliance on vestibular inputs; each for 30 s [72]						
Sharpened Romberg (2 studies)	Bipedal position with eyes open and eyes closed [73, 74]	Time (s)	R	2	76	62.5-72.8	26 F, 50 M
Romberg with Jendrassik maneuver (1 study)	Standing with both feet together, eyes closed and performing abduction of the upper limbs for 30 s [30]	n (%) able to hold the position >20 s	R	1	266	65-74 (69.5±3.0)	142 F, 124 M
Equi Test (1 study)	Twelve 20-s trials in a side-by-side position; 6 conditions (each condition twice): (1) normal vision, fixed support, (2) eyes closed, fixed support, (3) vision sway-referenced, fixed support, (4) normal vision, support sway-referenced, (5) eyes closed, support surface sway-referenced, (6) vision and support surface both sway-referenced [75]	N/A	R	6	55	61-83 (69.3±5.5)	36 F, 19 M
SOT (1 study)	Six 20-s trials, standing on a force platform, with the platform and/or visual surround sway referenced, according to subject's anteroposterior sway (1-3 motionless platform, 4-6 sway-referenced platform) [76]	Body sway angles	R	6	23	60-78 (66.2-71.3)	0 F, 23 M
CTSIB (2 studies)	Standing with hands at the sides, feet together, completing 6 sensory scenarios with various visual and supporting conditions; Visual disturbance provided by a rotating RDT image without a central rod [77] Modified, i.e. four 30-s trials: quite standing on a firm surface with eyes open and eyes closed, quite standing on a compliant (foam) surface with eyes open and eyes closed [22]	CoP displacement (cm) [22] Postural sway acceleration (m/s^2) [77]	R	4	61	64.0-69.0	19 F, 5 M
8-level balance scale (1 study)	Side-by-side standing, narrow base Romberg (eyes open; eyes closed), semi tandem (eyes open), tandem (eyes open; eyes closed), one leg stand (eyes open; eyes closed; eyes closed with cognitive distractor). Participants have to complete successfully a balance task for 30 s before progressing to the next task. The highest level performed successfully was rated [78, 79]	Score (0-8)	O	9	102	55-70 (66.4-69.9)	78 F, 25 M

Dynamic steady-state balance							
Tandem walk (7 studies)	On a beam, two trials [80] 3 m [78, 79] 6 m [81]; placing one foot in front of the other making sure that, with each step, the heel of one foot is directly in front of the toes of the other foot; walking forwards as fast as possible without falling or making any mistakes [42] 10-foot line, as quickly as possible without making mistakes (i.e. stepping completely off the line or failing to follow a heel-to-toe pattern), three trials [24] Pre-marked 9.14m tape line on the floor [82] On level ground and on a slightly-raised balance beam, both with and without the use of the anchors (i.e. two flexible cables, whose ends participants hold in each hand, to which 125g weights are attached at the opposing ends, and which rest on the ground; as the participants walk, they pull on the cables, dragging the anchors; 20 randomized trials [83]	Time (s) [24, 78, 79, 81, 82] Number of missteps [24, 42] n (%) who failed [80] Step speed; single-, double-support duration; trunk acceleration [83]	O; R	1	260	55-85 (65.9-77.0)	169 F, 48 M
Step test (2 studies)	Stepping one foot on, then off, a 7.5-cm block as quickly as possible in 15 s [84]; with the worse leg [85]	Number of steps on/off the block	R	1	67	53-83 (65.7-66.9)	38 F, 29 M
Four Square step test (6 studies)	Stepping as fast as possible in forward, sideways, and backward directions over 4 canes resting flat on the floor in a cross formation with the tips of the canes facing together, moving first in a clockwise direction and then counterclockwise position, without touching the canes, both feet make contact with the floor in each square before moving to the next [8, 45, 86-89]	Time (s)	R	1	470	55-81 (62.0-71.5)	363 F, 95 M
Step width & length, eyes open and eyes closed (1 study)	Footprints recorded on a 0.9x6.1 m (3x20 ft) paper walkway, triangular (base=5 cm) shapes were cut from adhesive moleskin and attached to the soles of the shoes at the midline of the toes. A square moleskin shape (5x5 cm) was attached to the midline of the heel. A stamp pad	Distance (mm)	R	1	56	66.7-72.8	41 F, 15 M

	inker was used to apply black ink to the triangular and squared shaped moleskin [74]						
MSL test (2 studies)	Standing with the feet together and then stepping out as far as possible with the preferred leg adjacent to a yardstick taped on the floor, before returning to the starting position, two trials [8] Stepping maximally with one leg while keeping the other leg planted and then return to the initial position in one step; for each leg and direction (front, side, back); five trials [24]	Distance (inches)	R	1	59	60-81 (67.7-77.0)	50 F, 9 M
360° turn (1 study)	Making a 360° turn, allowed to use assistive devices [5]	Score	O	1	282	60-74	228 F, 54 M
180° turn (2 studies)	Standing with arms by side and feet comfortably apart and pointing to the tape, then turn 180° on the spot within a designated area marked on the floor; as fast as possible, usual footwear; started from the word “GO” and stopped when shoulders and feet facing in the opposite directions; three turns to each direction with 1-minute rest break between trials [90] Hands by the side, positioned on a 40cm x 60cm square drawn on the ground; then turning 180° at self-selected speed by taking steps within the square to face the opposite direction [91]	Time (s), steps (n)	R; O	1	99	55+ (61.8-68.5)	52 F, 47 M
6m backwards walk (3 studies)	6m backwards walk, placing one foot directly behind the heel of the other with the shoes touching [92-94]	Time (s)	R	1	77	65-84 (68.9-69.7)	44 F, 40 M
10-m walk under single- and dual-task condition (1 study)	10 m instrumented walkway, single and dual task (walking while counting backwards aloud) conditions in 1) normal gait pattern, 2) narrow gait, 3) overlapping gait, and 4) tandem gait [70]	Stride time, stride length, stride width, stride velocity	R	8	54	65-80	N/A
Floor Transfer Task (1 study)	Standing upright on a mat, transferring to a sitting position on the floor mat, then returning to standing in any preferred way [95]	Time (s)	R	1	39	61.2±7.5	27 F, 12 M
SEBT (2 studies)	Balancing on the stance leg and reaching with the opposite leg as far as possible, five reaches in the anterior, medial, and posterior directions, calculating the star composite reach distance,	Distance (cm)	R	6	212	65.4-68.9	107 F, 99 M

	i.e. sum of the normalized reach distances for the right and left leg for all reach directions [96] Standing at the center of a grid placed on the floor, with eight lines extending at 45° increments from the center of the grid, placing one leg in the center of the grid, with the opposite leg reaching as far as possible along the eight defined directions in order to touch the furthest point on the floor as lightly as possible so as to avoid using the reach leg for support, and then return to the center of the grid without losing balance, the distance from the center of the grid to the reached point is measured [97]						
Dynamic balance/agility (2 studies)	Rapidly standing from a chair, walking around cones, and returning to the chair [98, 99]	Time (s)	R	1	120	60-84 (66.1-69.8)	43 F, 79 M
Narrow corridor walk (1 study)	N/A [100]	time (s)	R	1	40	60+ (69.8±7.5)	N/A
Sideways walk test (1 study)	Standing with the inner sides of the feet touching the starting line, and then walking sideways along a 5 m walkway at self-selected speed, i.e. abduction of the leading leg, followed by adduction of the trailing leg with the inner sides of feet touching each other; 3 trials toward both sides, randomized manner [101]	time (s), steps (n)	O; R	2	32	61.8±4.6	22 F, 10 M
Proactive balance							
TUG (91 studies)	As fast as possible [48, 50, 52, 81, 102, 103] Best of three trials [104] Two trials [105-107]; mean of two trials [108]; natural and fast speed [109] With and without a cognitive task (saying animal names) [110]; counting backwards (substracting 3 from 100) [77] 3 m version; two trials [111]; as fast as possible [112] Getting up from a chair with armrests, walk 3m, return and sit down again [97, 98]; walk 3 m to a mark placed on the floor [113]	Time (s) [9, 10, 16, 21, 28, 30, 33-35, 37, 38, 43, 51, 52, 54, 61, 65, 70, 77-79, 81, 86, 87, 89, 90, 95-98, 100-129, 131-137, 139, 141-148, 150-153] Time (m/s) [49, 140] time (s; each phase); amplitude; range of the movement; variability; smoothness of the movement for sit-	R	1-3	61826	46-99 (61.4-77.0)	9229 F, 12033 M

	Getting up from a chair, walk 3 m, turn, walk back, and sit down [9, 54, 114-117]; as quickly as possible, without running; two trials [118]; three trials [119]; habitual gait [34, 49, 70, 120, 121]; chair without armrests [95] Rising from a chair (40 cm high), walking 3 m, turning around, and sitting down again as fast as possible; two trials [122, 123] Rising from a chair, walking to and from a point located 3 m ahead at preferred speed, and then sitting down again [124]; two trials [61]; mean of three trials [90]; mean of two trials [107] On a command, participants get up from an armless, backless chair (43 cm high), walk forward 3 m, turn around, walk back to the chair and sit down again [38] On cue, participants rise from the chair, walk 3 meters to a line on the floor and return to their initial seated position [125, 126]; normal armchair (44 cm high [16]); marker of 20 cm diameter [127] Sitting in a free-standing padded armchair, then stand up without use of arms, walk at a comfortable and safe pace to a line on the floor 3 m away, turn and walk back to the chair and sit down again [128] Sitting in a normal chair (45 cm high), with the back against the chair, standing up, walking 3 m as quickly and safely as possible past a line on the floor, turn around, walk back to the chair, and sit down once again with the back against the chair; two trials [30] Rising from a chair (45 cm high) without using the arms to assist, walk 3 m to a cone, turn around the cone and return to the seat [129] Standing up from a standard chair of 45 cm height, walking 3 m, turning around a cone, returning to the chair, and sitting down again in the shortest time possible without running [78, 79]; two trials [52]	to-stand, stand-to-sit, and for the two turns; step-related variables (e.g., stride time, number of steps); ML and AP acceleration signals (measures of stability and smoothness of gait) [138]

	Standing up from a seated position, walk a distance of eight feet at usual pace, return to the chair, and sit back down [130]	
	Getting up from a chair, walking 10 ft, turning, walking back, and sitting down; three trials [54, 131]	
	Sitting with the back against the chair (approximately 46 cm high), on a command participants rise from a standard arm chair, wearing their own shoes and/or using an ambulatory aid, walk a distance of 10 ft, and return to the seat with their back resting against the back of the chair [132]	
	Sitting with the back against the chair (46 cm high, with armrests), on a command participants rise from a standard arm chair stand from a seated position, walk 3 m at their usual pace, turn around, walk back to the chair and sit down; walking aids permitted [133]	
	Sitting in a chair, then stand up without using the hands, walk to the end of a 10-ft pathway, turn around, walk back and sit down as quickly and safely as possible [21]	
	Getting up from a sitting position in an armless chair, walk 2.5 meters, return and sit down again in the same chair. A flag indicated the distance of 2.5 m from the chair; mean of three trials [134]	
	Sitting in a chair and on a command, standing and moving as quickly as possible around a cone placed 2.5 m away from the chair and return to the chair and sitting down [96, 135]	
	Sitting on a chair (43 cm high), with back support, travel a distance of 2.43 m, turn around a cone positioned at the end of the route, return, and sit down again at the chair; two trials [136]	
	Rising from a chair on a command, walk 8 ft, and return to sit in a chair [37]	
	3-trials: 1) Get up from a chair, walk 3 m straight on, turn around a cone, walk back to the chair, and sit down; 2) with an additional cognitive task (counting backwards in step 3, starting with 97),	

	and 3) with an additional motor control task (transporting a cup of water without spilling any water during the TUG) [137] Standing up and sitting down in a chair, walking and turning while simultaneously completing a cognitive task of counting backwards from 100 in 3's [87, 89] Dynaport was fixed with an elastic belt at the level of lumbar segment L3 over the participant's clothes; Standing up from a chair without the use of the arms, walking 7 m, turning around a pion, walking 7 m back to the chair, and sitting down without the use of the arms; as fast as possible without running; two trials [138] N/A [10, 28, 33, 35, 43, 51, 65, 86, 100, 101, 139-153]						
Chair rise and walk (1 study)	Starting from a seated position, then stand up and walk as quickly as possible in a predetermined straight line to a pylon 9.14 m, go around the pylon, and return to the original seated position [82]	Time (s)	R	1	39	65-85	20 F, 19 M
8-ft Up and Go (27 studies)	Part of the SFT [29, 42, 47, 57, 154-166] Getting out of a chair, walk 8 ft, turn around a cone, return to the chair and sit down as quickly as possible [167-171]; two trials [172] Sitting in a chair, hands on thighs and feet flat on the floor, on a command, stand up, walk as quickly as possible around a cone placed 8 ft ahead of the chair, and return to a fully seated position on the chair [173]; two trials [174] N/A [175, 176]	Time (s)	R	1	4724	51-89 (62.1-70.1)	2581 F, 992 M
FRT (30 studies)	Two trials [111], mean [107] Three trials [177] Part of the SFT [42, 105, 160, 162-166] Reaching forward as far as possible without moving the feet [11, 178] Maximum distance a person can reach forward beyond arm's length while standing in a fixed position, three trials [76]	Distance (cm) [11, 16, 18, 30, 42, 61, 66, 67, 70, 91, 93, 94, 97, 105, 107, 111, 122, 123, 151, 160, 162-166, 179-181] Distance (inches) [88] % (normalized using height) [76, 178]	R	1	13679	50-99 (61.5-71.3)	8577 F, 4072 M

	Measuring participant's balance with a tape measure horizontally on the wall and the participant reaching forward as far as possible from the waist without losing balance [97]	
	Standing with the feet shoulder-with apart, making a fist, and raising the arm to be parallel with the floor. The assessor took an initial reading on the yardstick, using the knuckle of the third metacarpal as the landmark, then reaching forward along the yardstick without moving the feet [88]	
	A yard stick was positioned horizontally next to the right side of the participant at the height of the acromion; Standing naturally and raising the right arm forward to 90° (parallel to the yard stick), and at the 'go' signal, reach forward as much as possible at their own pace [91]	
	Standing and then raising both arms in front to shoulder level while the heels touch the ground [66, 67, 179]; two trials [122]	
	Reaching forward beyond arm's length while maintaining a fixed base of support in the standing position; right and left arm recorded [70]; five trials [18]	
	Participants place their feet behind a marked line and whilst maintaining a fixed base of support reach forward along a preplaced measure tape [93]	
	Extending the right or left arm forward, while standing with legs apart, two trials [61]	
	Raising the arm closest to the wall to shoulder height; the position of the third metacarpal is recorded. Subjects are instructed to keep the feet flat on the floor and lean forward as far as possible without losing balance, touching the wall, or taking a step; two trials [16]	
	Raising one arm at 90 degree with fingers extended. A yardstick was mounted on the wall at shoulder height. The distance that a participant could reach while extending forward from the initial upright posture to the maximal anterior leaning posture without moving or lifting the feet is visually measured in cm, according to where the	

	middle finger tip is positioned on the mounted yardstick; two trials [123] Standing close against a wall with a measurement tape fixed on the wall and keep the shoulder in 90° flexion parallel to the tape; reach forward maximally with arm outstretched equal to shoulder's height without moving the feet or touching the wall; mean of three trials [180] Standing with the feet a comfortable distance apart and behind a line perpendicular and adjacent to a wall, the arm closest to the wall is then raised to shoulder height, and the position of the tip of the middle finger is measured; feet flat on the floor and leaning forward as far as possible without losing balance, touching the wall, or taking a step. The position of the tip of the middle finger is then recorded at the point of furthest reach, and the difference between the two points is recorded as the maximal distance; three trials [30] N/A [94, 107, 151]						
LRT (1 study)	Standing with the back to (but not in contact with) a wall, feet placed in a standardized position with 0.1 m between the most medial aspects of the heels, with each foot angle at 30°, then reaching directly sideward as far as possible without overbalancing, taking a step or touching a wall; two trials [16]	Distance (cm)	R	1	28	57-73 (65.9-66.0)	3 F, 25 M
7m obstacle walk (1 study)	7 m walk with stepping over a 30 cm obstacle at the 4 m point, normal pace; two trials [135]	Time (s)	R	1	134	69.6-70.3	85 F, 49 M
Zigzag walking (1 study)	Walk along a 10-m walkway with four cones placed 2 m apart on the floor between the start and finish points as quickly as possible. The cones were set to alternate from side to side with a distance of 0.5 m from a line drawn through the start and finish points. Participants walk around the outside of each cone and walk through the finish point; two trials [182]	Time (s)	R	1	81	50-74 (59.0-61.0)	40 F, 41 M
Curved walking (1 study)	Walking three times (i.e., 1,080°) around a marked circle on the floor with a diameter	Time (s), n (checked boxes/subtractions)	R	1	1054	65.0±7.0	526 F, 528 M

	of 1.2 m. Walks in the clockwise and counterclockwise directions were alternated to avoid the effects of direction. Participants started with the single walking trial (walking three times counterclockwise, followed by three times clockwise); followed by adding a dual task; and another dual task (subtracting serial 7 s) [183]						
Reactive balance							
Reactive balance test (1 study)	Stand erect in bipedal step stance with hands placed on hips and gaze fixated on a cross on the nearby wall on a two-dimensional balance platform. Medio-lateral perturbation impulses are unexpectedly be applied in order to investigate reactive postural control (10 s intervals); three trials [70]	Summed oscillations of the platform in medio-lateral and anterior-posterior directions	R	2	54	65-80	N/A
Push and release test (2 studies)	Standing in a comfortable stance with eyes open and pushing backward against a palm of the examiners' hand. After the examiner suddenly releases his or her hands, participants are required to regain balance [70, 147]	Amount of steps to regain balance	O	1	102	65-80 (69.8-70.0)	26 F, 22 M
Adaptive gait test (1 study)	Walking barefoot at self-selected comfortable pace within a narrow, 6.1-m-long path with a cognitive task (reciting the days of the week in reverse order); four trials [184]	gait speed (m/s), step errors (n)	R	1	20	61-81	69.1±8.6
Step Execution Test (2 studies)	Standing barefoot and upright on a force platform viewing an 'X' displayed on a screen, 3 m in front, step as quick as possible (step length 50-60 cm), following a tap cue on their heel, nine trials (forward, backward, sideward) [185] Stand with the foot of the preferred leg on a foot-pad, and react to an auditory stimulus by stepping rapidly onto a second foot-pad 18 inches away; two trials [8]	Reaction time (ms)	R	1	72	60-88 (67.7-69.6)	9 F, 28 M
Backwards stepping test (1 study)	When signaled, lean as far backwards as possible, and then take a backward step with the unloaded leg; three trials (eyes open, eyes closed) [186]	Ground reaction force (N/kg)	R	1	36	65-75 (66.2-68.3)	31 F, 5 M

Crossover stepping test (1 study)	When signaled, lean as far laterally as possible, and then take a crossover step with the unloaded leg; three trials (eyes open, eyes closed) [186]	Ground reaction force (N/kg)	R	1	36	65-75 (66.2-68.3)	31 F, 5 M
Limits of stability test (1 study)	Maximum distance that each subject could shift their center of gravity (COG) without losing balance (forward, backwards, and both sides) [112]	Reaction time (RT), movement velocity (MVL), maximum Excursion (MXE)	R	4	30	64.2±7.3	19 F, 11 M
Performance batteries							
BBS (35 studies)	14 balance tasks (5 static, 9 dynamic) with varied difficulty (e.g. sit-to-stand, standing with eyes open and eyes closed, tandem stand, one-leg stand, transfers, reaching for an object, a 360° turn; each scored from 0 to 4 [15, 21, 33, 47, 50, 52, 54, 65, 75, 90, 95, 98, 101, 109, 112, 113, 116, 117, 126, 132, 140, 142, 148, 149, 151, 153, 185, 187-193]	Score (0-56)	O	14	2324	56-88 (61.4-74.0)	1255 F, 8728 M
SPPB (34 studies)	Three hierarchical standing balance tests (side-by-side, semi-tandem, tandem position for 10 s each), 4-m walk at usual speed (m/s), and five repeated chair stands as quickly as possible (s), each scored from 0 to 4 [11, 52, 96, 100, 121, 129, 165, 176, 180, 194-217]	Score (0-12) [11, 52, 96, 100, 121, 129, 165, 176, 180, 194-204, 207, 208, 210-213, 215-217] Score summarized in quartiles (lower body function: poor, fair, good excellent) [206] % score (7-9; 10-12 points) [205]	O	3	17687	60-89 (63.0-72.3)	10992 F, 4187 M
Tinetti Test / Performance Oriented Mobility Assessment (7 studies)	Tinetti's balance and gait evaluation [73, 103, 218] 13-item balance and 9-item gait assessment, each scored from 0 (unable) to 1 (able to perform) [219] Static sitting balance (rising from the sitting position without using), standing balance (the first five seconds after the subject's sternum was gently pushed by the examiner, and when stance was stabilized, staggering or excessive sway of the subject was examined with the subject standing and his eyes closed); 360° turn,	Score (0-28) [73, 103, 218] Score (0-22) [219] POMA balance score [220-222]	O	3-28	8166	55.0-97.6 (62.5-66.8)	4916 F, 2524 M

	observing steadiness and continuity of steps [220-222]						
PPT (2 studies)	Two versions, i.e. a 9-item scale, including writing a sentence, simulated eating, 360° turn, putting on and removing a jacket, lifting a book and putting it on a shelf, picking up a penny from the floor, a 50-foot walk test, and climbing stairs (scored as two items); and a 7-item scale, not including stairs; each scored from 0-4 [85] Nine items, including Romberg test, chair sit-to-stand, lifting a book from waist height to a shelf at shoulder level, putting on and taking off a coat, picking up a penny from the floor, 360° turn, 15 m walk, ascending one flight of stairs, climbing 4 flights of stairs; each scored from 0-4 [118]	7-item score (0-28) [85] 9-item score (0-36) [85, 118]	O	7-9	91	60-83 (67.4-68.0)	54 F, 37 M
FAB scale (6 studies)	Ten static and dynamic balance tasks (stand, reach, turn in a circle, step up and over, tandem walk, one-leg stand, stand on foam with eyes closed, two-footed jump, walk with head turns, maintain a reactive posture), each scored from 0-4 [78, 79, 158, 164, 190, 223] Task 1-4 [161]	Score (0-40) Score (0-8) [161]	O	10	308	52-89 (61.8-69.5)	187 F, 72 M
CS-PFP-10 (1 study)	10 household tasks, including carrying a pot of water from one counter to another; carrying groceries onto and off a 4-step platform; transferring laundry; donning and removing a jacket; sweeping kitty litter into a dustpan; climbing stairs; sitting down and getting up from the floor; picking up 4 scarves from the floor; 6 m walk; maximal reach [202]	Score (0-100)	O	10	26	60+ (68.6-72.3)	22 F, 4 M
PPB (4 studies)	Modified version of the SPPB, more challenging tasks, i.e. 10 repeated chair stands, single leg stance, narrow walk (walking between 2 parallel lines separated by 20 cm) [144, 215, 224, 225]	Score [144, 215, 224] Time (s; individual tasks) [144]	O; R	N/A	2149	64.0-69.9	556 F, 494 M
CBM (3 studies)	13 tasks: One-leg stance, tandem walking, 180° tandem pivot, lateral foot scooting, hopping forward, crouch and walk, lateral dodging, walking and looking, running with controlled stop, forward to backward walking, walk, look and carry,	Score (0-96)	O	13	132	55-70 (66.4-69.9)	39 F, 12 M

	descending stairs, and step-up x 1 step [78, 79, 226]						
FMM (1 study)	Seven tasks including deep squat, hurdle step, in-line lunge, shoulder mobility, active straight leg rise, trunk stability push-up, and rotary stability [227]	Score (0-21)	O	7	90	65.3±4.6	N/A

Test characteristics			Scale		Study population		
Strength test	**Detailed Description**	**Unit**	**Level**	**Items**	**Na**	**Age**	**Sexb**
One repetition maximum							
Handgrip strength (81 studies)	Both hands, best trial [228] Bi-handgrip strength, two trials [105] Standing and then grasping a grip device; best of three trials [31]; mean of two trials [147]; best of two trials [229] Sitting, elbow fully extended in front on shoulder height; mean of three trials [230] Sitting position, shoulders adducted, neutrally rotated, elbow flexed at 90°, forearm neutral - dominant hand; best of three trials [231] - both hands, three trials for each hand, best score [121] Electronic / hydraulic dynamometer - best trial (number not specified) [9, 40] - two trials [199] - three trials [82]; mean of three trials [196] - best of three trials [49, 57, 146, 184, 188, 232, 233] - both arms, mean of six trials [124] - standing; mean of three trials [234] - Sitting, shoulders adducted, neutrally rotated, elbow flexed at 90°, forearm neutral; mean of two trials [235]; mean of three trials [236] - dominant hand [176]; arm by side; best of three trials [168, 237] - dominant hand, sitting in an upright position, arm of the measured hand unsupported and parallel to the body; one trial [70]	Force (kg) [2, 9, 31, 40, 49, 57, 69, 70, 105, 111, 128, 139, 167, 180, 184, 188, 197, 199, 201, 203, 207, 228, 231, 235, 237-243, 245-251, 253] [59, 179, 203, 204, 233, 234] [7, 32, 82, 91, 120, 121, 124, 133, 147, 150, 152, 165, 176, 214, 216, 217, 229, 230, 240, 252, 254-261] Force (pounds) [21] Force (kg)/body-weight (kg) [168] % of people with force scores (kg, classified as weakness) [196] kg/cm^2 [71] Force (pound per square) [244] Force (kPa) [28] Force (Newton) [63] Quartiles (% lower quartile reported) [236]	R	1-2	130821	34-89 (60.4-70.5)	75538 F, 49439 M

- dominant hand, sitting, dominant shoulder in rest position, elbow flexed 90° without support, forearm and wrist at neutral position; best of three trials [201]
- dominant hand, sitting comfortably, dominant arm by side, elbow flexed 90°, hand held in mid-supination/pronation position; best of three trials [180]
- both hands, one trial in each hand; best score [63]
- both hands; three trials in each hand, best score for each hand [238]; best trial of the dominant hand [239]
- best score of both hands [203]; best score of each hand [240]; sum of best score of each hand [111]; mean score of each hand [241]
- both hands, wrist in neutral position, elbow flexed at 90°; three trials for each hand; mean of each hand [242]
- both hands, two trials for each hand, mean for each hand and larger mean from one of the hands [207]
- N/A [28, 91]

Bulb hand dynamometer

- both hands, holding at shoulder level, two trials in each hand; mean of both hands added [243]
- dominant hand, medium (women) or large (men) dynamometer, sitting; best of three trials [71]
- both hands, three trials in each hand; mean for each hand [244]
- both hands, two trials in each hand; mean for each hand [245]

Calibrated dynamometer

- both hands, elbow flexed to 90° [167]
- preferred hand, arm raised overhead then slowly lowered towards floor [21]; three trials [32]

Handheld dynamometer

- both hands, standing, best of two trials (each hand) [2, 59, 69, 128]; mean of three trials (each hand) [110]

	- both hands, two trials on the dominant and three trials on each hand [246] - two trials; mean score [197]; best score [120, 247] - dominant hand [179]; mean of two trials [248]; mean of three trials [249]; best of two trials [133] - sitting, elbow flexed two 90°, best of two trials [204]; best of three trials [250, 251] - standing with their forearms bent at 90 °, best of two trials [7] - elbow positioned at 90° of the side of the body; dominant hand; mean of three trials [252] N/A [48, 103, 139, 150, 152, 165, 214, 216, 217, 253-261]						
Shoulder flexor strength (1 study)	Right arm, 90° shoulder flexion, elbow in full extension; mean of three trials [47]	Force (kg)	R	1	85	65-84 (69.0±0.4)	37 F, 48 M
Hip muscle strength (2 studies)	Supine on a plinth, both legs 10° abducted, a strap (5 cm wide) around the plinth and over the pelvis; for the examiner-resisted test, participants pushed as hard as possible against the HHD as the examiner provided resistance, stabilizing and positioning the HHD; for the belt-resisted test, HHD is placed between the side of the test leg and a second strap (5cm width), participants spread their legs apart simultaneously as hard as possible [262] Right, and left hip abduction strength on isotonic external resistance machine [263]	Torques kg/body mass [263]	R	2	45	55-75 (63.7-68.4)	31 F, 14 M
Knee extensor strength (1 study)	Computer-based manual muscle testing, knee at 30° flexion; mean of three trials [47]	Force (kg)	R	1	85	65-84 (69.0±0.4)	37 F, 48 M
Leg strength (6 studies)	Sitting in a standard chair (45 cm high), connected to a WBB (57° angle from the ground) via custom seatbelt straps; pressing the feet on the WBB as hard and as fast as possible; three trials [264] Dynametry [254] Leg press machine [34, 145, 176, 263]	Force (kg) Kg/body mass [263]	R	1	272	55-75 (61.1-69.3)	140 F, 76 M

Toe grasping strength (2 studies)	Barefoot, one-leg stand, both hands on the wall in front, holding the dynamometer grasping bar with the toes [265] Sitting upright on a chair, without leaning on the backrest; both hips and knees flexed at approximately 90°; ankles placed in a neutral position and fixed with a strap [107]	Force (kg)	R	1	7227	52-78 (66.3-67.6)	534 F, 188 M
Maximal Isometric Strength (MIS)							
Elbow extensor strength (1 study)	Instrumented wooden pole that subjects pressed against the ground; Subjects sit on a bench, with the shoulder in a neutral position, elbow angle of 90°, forearm was parallel to the ground; best of three trials [146]	MVC (kg)	R	1	26	69.2-70.0	17 F, 9 M
Hip extensor strength (1 study)	HHD, mean score [178]	% (strength/body weight)	R	1	39	60-78 (68.5-69.7)	15 F, 24 M
Hip flexor strength (2 studies)	HHD [91], mean score [178] HHD, sitting position, hip flexed at 90° and knee flexed at 90°; sensor of the HHD was applied to a distal site on the anterior surface of the thigh [55]	% (strength/body weight) Force (kg) [55, 91]	R	1	818	60-78 (68.5-69.7)	775 F, 313 M
Hip abductor strength (2 studies)	Supine position with the hip and knee fully extended and hip positioned in neutral abduction; sensor of the HHD was applied 5 cm proximal from the lateral malleolus [55] HHD, three trials (each hold for 3-5 s) [101]	Force (kg)	R	1	744	61.8-68.7	482 F, 262 M
Knee extensor strength (11 studies)	HHD, sitting upright, raising lower legs up 90°, parallel to the ground, holding this position as strongly as possible against the maximum persistent (5 s) force applied by the examiner through the HHD placed on the front of the ankle proximal to the medial malleolus; two trials for each leg, best score [266] HHD [91, 107]; mean score [10, 178] Leaning back in a chair, extending both legs at the knee while pulling against a dynamometer; best of two trials [66]	Force (kg) [34, 66, 107, 232, 266] % (strength/body weight) [100, 178] N/kg [10] Peak torque (Nm) [145, 232] MVC (kg) [146]	R	1-2	1595	60-78 (61.1-71.6)	1038 F, 402 M

	Sitting on a high chair and pushing against a strap linked to a spring gauge [100] Seated position using an adjustable chair with a 90° angle of hip and knee joints, dominant leg, as fast as forcefully, strongest of five trials [34] Dynamometer, right side; sitting on a backwardly-inclined (5°) chair, range of motion was set from a knee joint angle of 90° to 160° (180° represents full extension); best of two trials [145] Sitting in an upright position with back support and with both the hip and knee flexed at 70°; distal leg affixed to a strain gauge force transducer; best of three trials [146] Dominant limb [232]						
Knee flexor strength (1 study)	HHD, mean score [178]	% (strength/body weight)	R	1	39	60-78 (68.5-69.7)	15 F, 24 M
Leg strength (6 studies)	Dynamometer, both legs simultaneously [267]; mean of two trials [245]; best of two trials [268] Dynamometer, both legs simultaneously, standing with back straight against a wall and knees 115° flexed; a bar connected by a chain to the dynamometer was held in front of the thighs and has to be lifted upwards with maximum force using only the legs, and keeping the neck and back straight; mean of two trials [243, 269] Fitted with the harness around their hips and seated in a standard chair (45 cm height) with the seatbelt straps connecting the harness to the FysioMeter-mount. The lengths of the seatbelt straps were adjusted using a tape measure and a goniometer angle between sessions to reach a knee angle of approximately 120°; foot placed in the middle of the Wii Balance Board [270]	Force (kg)	R	1	2544	50-79 (61.4-69.0)	1230 F, 1277 M
Ankle dorsiflexor strength (7 studies)	HHD [91, 149]; mean score [10, 178]; mean of three trials [90, 112]; three trials (each for 3-5 s) [101]	% (strength/body weight) [178] N/kg [10] Force (kg) [90, 91, 101, 112, 149]	R	1	357	60-78 (61.8-69.7)	222 F, 135 M

Ankle plantar flexor strength (5 studies)	HHD [91, 142, 149]; mean of three trials [90, 112]; three trials (each for 3-5 s) [101]	Force (kg)	R	1	832	50-80 (61.8-68.5)	450 F, 392 M
Functional muscle power							
Upper body functional muscle power							
30 second arm curl (20 studies)	Part of the SFT [42, 45, 154, 156, 160, 162-164, 166, 170, 271] Performing as many biceps curls as possible in 30 s, using a 2.27-kg dumbbell (full range of motion; study in women) [169, 272] Flexing and extending the elbow of the dominant hand, lifting a weight (8 lb [3629g] dumbbell for men; 5lb dumbbell [2268g] for women) through the complete range of motion as many times as possible in 30 s [173] Sitting on a chair, using the dominant hand to bring a weight (2.0 kg) up and down (flex and extend the biceps) as many times as possible in 30 s [179] Hand curling a hand weight (5 pounds for women and 8 pounds for men) for 30 s [131] Siting on the chair, holding the dumbbell (women 5 lbs, men 8 lbs) in the hand with palm facing towards the body (with the arm in a vertically down position beside the chair), bracing the upper arm against the body so that only the lower arm is moving, then curling the arm up through a full range of motion, gradually performing elbow flexion with supination; as the arm is lowered through the full range of motion, gradually return to the starting position [152] N/A [144, 175]	Repetitions	R	1	5768	51-89 (61.9.0-69.9)	2652 F, 1111 M
Abdominal Strength (2 studies)	Lying down on an abdominal pad, with knees flexed at 90°, hands set on the pad frame. Rise with the chest up to approximately 30° from the floor as many times as possible in 30 s [167] Lying on sit-up equipment and performing sit-ups with the feet attached to the equipment's	Repetitions	R	1	252	59-60+ (63.0-66.9)	230 F, 122 M

	foot holders as many times as possible in 30 s [235]						
Single forearm contractions (1 study)	Dynamic single contractions in both arms, HHD at 10%, 20%, and 40% of the subject's maximum voluntary contraction [273]	MVC (kg)	R	1	32	59-85 (66.0±2)	13 F, 19 M
Seated medicine ball throw (1 study)	Participants horizontally propel a 3 kg ball without trunk flexion [144]	Distance (m)	R	1	36	68.8-68.9	20 F, 16 M
Lower body functional muscle power							
Five times Sit-to-Stand (61 studies)	Part of the SPPB [27, 28] Rising from a chair and sitting back down five times [9, 34, 112, 168]; without arm support [147, 240, 274, 275]; time measured at the final sitting down, best of three trials [127]; mean of three trials [90] Five repetitive chair stands as quickly as possible with arms folded across the chest [1, 7, 41, 63, 70, 203, 246, 249]); mean of two trials [8, 276]; mean of three trials [95, 188] Sitting in a standard chair, arms folded across the chest, standing up and sitting down five times [124]; as fast as possible [247] Sitting in the middle of an armless folding chair, then standing up and sitting down with an arm folded in front of the chest; mean of two trials [36] Standing all the way up and sitting all the way down 5 times as quickly as possible while keeping the arms folded across the chests; straight back, flat, level, firm seat (45 cm high) [215] Getting up and sitting from a chair (43 cm high, flat seat), arms crossed over the chest, rising until full extension at trunk and lower limb joints, and returning with the back fully supported at the back of the chair; best of two trials [136]	Time (s) [1-3, 5, 6, 9, 23, 27, 34, 36, 41, 53, 60, 63, 70, 81, 85, 90, 91, 93-95, 109, 112, 118, 124, 125, 127, 136, 137, 141, 142, 145, 147, 149, 168, 188, 203, 215, 232, 240, 246, 247, 249, 260, 261, 266, 274, 275, 277-284] Stands/min (60*[5/time, s]) [7] Score (0-4) [28]	R	1	81289	40-90+ (58.7-71.0)	41301 F, 36656 M

	Standard chair (43.2 cm high), transferring to a standing position and returning to a sitting position, not allowed to use arms [277]	
	Standard padded chair (43.2 cm high) without armrests, both arms crossed against the chest, starting from a seated position and standing up (legs straight) and sitting down (full weight on the chair) [278]	
	Getting up from and sitting down on the chair (43.6 cm high) without arm rests [137]	
	Standing and sitting five times from an armless chair (46 cm high), not permitted to use arms [125]	
	Straight-back chair, placed against a wall, with a hard seat and standard height, sitting with the feet on the floor and arms folded across the chests on the chair, time measured at the final standing position [2]	
	Standing up and sitting down as quickly as possible five times in a row from an armless straight-back chair, arms across the chest, time measured at the final standing position [5]; time measured at the final sitting position [3]	
	Sitting in a hard-backed chair (43 cm high), arms folded across the chest, rising as fast as possible to a full standing position, then returning to a full-sitting position five times [93, 94]	
	Rising fully from a standard armless, backless chair five times as fast as possible, arms folded closely to the trunk, no moving of the feet during the test, time measured at the final sitting position [38]	
	Standing up from a straight-backed chair (43 cm high) five times at a self-selected pace, arms folded across the chest [53]	
	Sitting on a chair with the back touching the backrest, seat height adjusted to participant's lower leg length, knees flexed at 90°, time measured at the final sitting position [279]	
	Standing up and sitting down five times as quickly as possible from a straight-backed chair	

	[280]; time measured at the final standing position [23] Standard chair with arm rests, both arms crossed against the chest, starting from a seated position (upper back against seat), standing up to full extension and sitting down again (upper back against seat), best of two trials [118] N/A [6, 60, 81, 85, 91, 109, 141, 142, 145, 149, 232, 260, 261, 266, 281-283]						
One time sit-to-stand (7 studies)	Sitting in a straight-back chair, barefooted, on cue, standing up and sitting down as quickly as possible, upper extremity use not permitted [21] Sitting on a chair (43 cm high), on cue, rising to full stance; best of three trials [285] Adjusted seat height (5 cm increments from 45 to 60 cm) to achieve a 90/90 (hip/knee angle), sitting on the front half of an instrumented chair, using the arms as normally during the task, while standing as quickly as possible, three trials [126] Chair rise from different seat heights (43 cm, 38 cm, 30 cm) [286] Standing up as quickly as possible from a standard chair (43 cm high), arms crossed across the chest and feet shoulder-width apart placed flat on the floor [4] N/A [144, 287]	Time (s) [4, 21, 285-287] Force (N/s [kg]; W [kg]) [126, 144]	R	1	414	60-74 (61.6-69.9)	235 F, 79 M
Ten times sit-to-stand (6 studies)	Rising from a sitting to a standing position with straight back and legs and sitting down again as fast as possible [49] Straight-backed chair (45cm high), arms crossed against the chest, rising as quickly as possible without the use of the hands [215, 250, 251] Rising from a chair as quickly as possible with arms placed across the chest [229] N/A [86]	Time (s) [86, 229, 250, 251] Speed (stands per minute: [10/s]*60) [49]	R	1	3283	50-81 (62.6-69.0)	1182 F, 1012 M

15 second Sit-to-stand (1 study)	Straight-backed, non-padded, flat-seated, armless chair, Standing without using hands or arms, arms folded across the chest; mean of two trials [197]	Repetitions	R	1	5777	65-79 (69.8-70.1)	5777 F
30 second sit-to-stand (51 studies)	Part of the SFT [42, 45, 57, 154-156, 158-164, 166, 271]; two trials [157] Part of the Fullerton Functional Fitness Test [170] Standing in front of a stable chair, hands across the chest, then practicing sitting down and standing up for 30 s [31, 272] Sitting in a chair (43 cm high) with arms crossed at the wrists and holding against the chest, then standing up as many times as possible [171, 172, 288]) Sitting on a standard armless chair (45 cm high), looking straight forward with arms folded across the chest, then standing up and sitting down as many times as possible [167] Rising up and sitting down with arms folded in front of the chest as quickly as possible on a firm, armless chair placed against a wall [82, 289] Standing up and sitting down from a bench without armrests and back support as many times as possible, feet flat on the floor, initial foot placement and chair height individually adjusted [290] Stand up from a seated position as many times as possible [291] Stand fully upright and then return to the seated position as many times as possible [66, 67, 87, 173, 174] Different chair heights (43 cm; then adjusted to 80, 90, 100, 110 and 120% of the participants' lower leg length), last attempt at the end of 30 s is counted as a full stand if the participant is more than halfway up from sitting [284] Standard padded chair (43.2 cm high) without armrests, starting from the seated position and	Repetitions	R	1	7493	51-91 (61.1-71.6)	3730 F, 1697 M

	standing up (legs straight) and sitting down (full weight on the chair); mean of two trials [278] Sitting on a chair, back straight, feet shoulder-width apart and flat on the floor, arms crossed at the wrists and held against the chest, then rising to a full stand and returning to a fully seated position as many times as possible [179] Chair (44 cm high) without arms, sitting in the middle of the chair, feet shoulder width apart and placed on the floor at an angle slightly behind the knees, arms crossed at the wrists and held against the chest, then rising to a full stand and returning back to the initial seated position, as many full stands as possible; mean of two trials [16, 17] Sitting in the middle of the chair, arms across the chest, then rising to a full stand and returning to a fully seated position as many times as possible [131] Standard chair with arm rests, both arms crossed against the chest, starting from a seated position (upper back against seat), standing up to full extension and sitting down again (upper back against seat); best of two trials [118] Sitting in a standard-height chair with arms crossed over the chest, then stand fully and sit down again as many times as possible [97] N/A [10, 26, 43, 96, 144, 145, 152, 153, 175, 232, 234]						
1 minute sit-to-stand (2 studies)	Stand up from and sit down from a standard height chair without the use of the arms [292] Sitting on the edge of a standard-height chair, arms crossed over the chest, and repeatedly standing up from and returning to a seated position without assistance [4]	Repetitions	R	1	123	55-70 (62.2-70.7)	76 F, 47 M
One time kneel-to-stand (1 study)	Part of MOD scale [286]	Score (0-5)	R	1	259	60+ (67.6±7.0)	143 F, 116 M

Floor rise to standing (6 studies)	Lying in a supine position, with feet together and hands palm down and at the side, then rising to a standing position [93, 94] Sit and rise from the floor (flat, non-slippery surface), using the minimum support needed [112] Stand up from a supine position [282] Sitting and rising unassisted from the floor with partial scores assigned from the two required actions of sitting (5 points) and rising (5 points) and a final composite SRT score [153] N/A [141]	Time (s) Score [112, 153]	R	1	172	58-84 (67.0-69.3)	50 F,49 M
Five Step Test (1 study)	N/A [266]	Time (s)	R	1	621	50+ (66.8-69.4)	428 F, 193 M
Stair climbing (2 studies)	Walking up and down a standard flight of stairs, three times at self-selected pace, using the handrail for support only if needed [53, 216]	Time (s)	R	1	1143	55-79 (63.8-67.5)	634 F, 509 M
Stair climbing (8 steps) (2 studies)	Climbing eight steps (17 cm high, 31cm long) without using the handrail, requiring a step by step pattern; best of two trials [136] N/A [261]	Time (s)	R	1	111	65.6-67.8	50 F, 35 M
Stair climbing (10 steps) (3 studies)	Climbing a flight of stairs (10 steps) as quickly as possible without using the handrails or any other aid (14 cm high [87]; 7.8 cm high [89] Ascending and descending a flight of stairs (10 steps, 0.27 m high and 0.18 m deep) as quickly and safely as possible, while having the option of using a single handrail for support [293]	SCP (W) [87, 89] Time (s) [293]	R	1	212	50-75 (62.7-71.5)	152 F, 67 M
Stair climbing (11 steps) (3 studies)	Ascending a standard fight of stairs (11 stairs, 16 cm high), avoiding the use of the handrail [93, 94]; as rapidly as possible [92]	Time (s) [92-94] SCP (W) [93, 94]	R	1	77	65-84 (68.9-69.3)	37 F,40 M
Stair climbing (12 steps) (2 studies)	Ascending and descending 12 stairs, permitted to use the handrail, but not allowed to use it to push or pull oneself [277] Ascending and descending 12 stairs (15cm high, 20 cm tread), as fast as possible while being safe [102]	Time (s)	R	1	337	45-80 (58.7-64.8)	183 F, 164 M

Stair climbing (14 steps) (1 study)	Walk as fast as possible up 14 stairs without the use of railings [289]	Time (s)	R	1	30	68.5±5.1	15 F, 15 M
Stair climbing (15 steps) (1 study)	Ascending and descending a flight of 15 stairs (18 cm high, 27 cm tread) at normal pace, preferably without using the handrail [135]	Time (s)	R	1	134	69.6-70.3	85 F, 49 M
Stair ascent (23 steps) (1 study)	Walking up one flight of stairs consisting of 23 steps (16.5 cm high, 19.2 cm wide) as quickly as possible; after 14 steps, the participants make a left-hand wrap-around turn and then completed the remaining nine steps; not allowed to use the handrails; best of the two trials [294]	Time (s)	R	1	62	60-83 (66.6-71.0)	N/A
Stair ascent (16 steps) (1 study)	16 steps, height of 15 cm; not allowed to hold the handrails [143]	Time (s)	R	1	48	60-80 (68.6±6.1)	N/A
Stair ascent (10 steps) (4 studies)	Ascending a 10-stair prop (17 cm high, 30 cm deep) at fast pace [168] Walking up 10 steps in an expeditious and safe manner, placing one hand close to the handrail for balance if necessary, but not on the handrail [275] Ascending a 10-stair flight (16.5 cm stair high) as fast as possible, use of handrail allowed [70] Climbing 10 steps as fast as comfortably possible with one hand near, but not on, the handrail [200]	Time (s)	R	1	158	62-80 (66.0-70.0)	69 F, 35 M
Stair ascent (9 steps) (2 studies)	Walking quickly but safely up and down a nine step flight of stairs (step height: 17 cm); time started after the cue to go and stopped when the second foot reached the top step [147, 283]	Time (s)	R	1	71	62.7-70.0	46 F, 25 M
Stair ascent (4 steps) (1 study)	Walking up 4 stairs (15 cm high), arriving on a full stance on the fourth step without any support or help, three trials, best score [285]	Time (s)	R	1	33	60-74 (64.4-65.7)	21 F, 12 M
Stair ascent (one time) (1 study)	Part of MOD scale [286]	Score (0-5)	R	1	259	60+ (67.6±7.0)	143 F, 116 M

Stair descent (16 steps) (1 study)	16 steps, height of 15 cm; not allowed to hold the handrails [143]	Time (s)	R	1	48	60-80 (68.6±6.1)	N/A
Stair descent (14 steps) (1 study)	14 steps (height 17 cm, length 30 cm); time starts with participants' initiation of first step and stops when both feet are on the landing [282]	Time (s)	R	1	33	67±4.5	N/A
Stair descent (10 steps) (1 study)	Walking down 10 steps in an expeditious and safe manner, placing one hand close to the handrail for balance if necessary, but not on the handrail [275]	Time (s)	R	1	19	66.0±1.0	14 F, 5 M
Stair descent (9 steps) (1 study)	Walking quickly but safely up and down a nine step flight of stairs (step height: 17 cm); time started after the cue to go and stopped when the second foot reached the floor [147]	Time (s)	R	1	48	69.8-70.0	26 F, 22 M
Stair descent (one time) (1 study)	Part of MOD scale [286]	Score (0-5)	R	1	259	60+ (67.6±7.0)	143 F, 116 M
Functional leg extensor strength (1 study)	Taking a short step forward, first with the right leg, squat down until the knee of the tracking leg lightly touches the mat, and then rise up immediately and step back to the starting position, then repeating with the left leg [53]	Maximal weight relative to the subject's body weight	R	1	1133	55-79 (63.8-64.1)	632 F, 501 M
Lift and reach (one minute) (2 studies)	Sitting at a standard height desk, then repeatedly lifting a weight onto and off a shelf placed on the desk located at shoulder level immediately in front (10 pound for women, 20 pound for men) [292] Sitting in a standard chair at a standard height desk (75 cm), then lifting a weight repeatedly onto and off a shelf positioned at approximate shoulder height, 37 cm above the desktop (5 kg dumbbell for women, 8 kg dumbbell for men) [4]	Repetitions	R	1	123	55-70 (62.6-70.7)	76 F, 47M
Standing long jump (2 studies)	Jumping horizontally, using a 2-ft. takeoff and landing, three trials, measured at the heel of the foot [32] Jumping with no restriction to arm movements as high and as fast without bending legs in air and landing with both feet on the jump mat; average of three trials [263]	Distance (cm [32]; m [263]) Power (kg/body mass), Velocity (m/s) [263]	R	1	98	50-79 (63.7±1.1)	54 F, 44 M

Squat jump (1 study)	Static position, knee bent in a 90° angle, hands on the hip during the whole jump; three maximal trials separated by 1.5 min of rest [34]	Maximal Ground Reaction Force (A; N*kg-1), Rate of Force Development (N*kg-1), Force (N)	R	1	63	65-70 (67.5±0.4)	63 F
Single knee extension contractions (1 study)	Single knee extension contractions with a handgrip device at 20%, 40%, and 60% of the subject's max voluntary contraction [273]	Maximum work rate (WR_{max})	R	3	32	59-85 (66.0±2.0)	12 F, 19 M

Test characteristics			Scale		Study population		
Strength test	**Description / Variation**	**Unit**	**Level**	**Items**	**N[a]**	**Age**	**Sex[b]**
One repetition maximum							
Handgrip strength (81 studies)	Both hands, best trial [228] Bi-handgrip strength, two trials [105] Standing and then grasping a grip device; best of three trials [31]; mean of two trials [147]; best of two trials [229] Sitting, elbow fully extended in front on shoulder height; mean of three trials [230] Sitting position, shoulders adducted, neutrally rotated, elbow flexed at 90°, forearm neutral - dominant hand; best of three trials [231] - both hands, three trials for each hand, best score [121] Electronic / hydraulic dynamometer - best trial (number not specified) [9, 40] - two trials [199] - three trials [82]; mean of three trials [196] - best of three trials [49, 57, 146, 184, 188, 232, 233] - both arms, mean of six trials [124] - standing; mean of three trials [234] - Sitting, shoulders adducted, neutrally rotated, elbow flexed at 90°, forearm neutral;	Force (kg) [2, 9, 31, 40, 49, 57, 69, 70, 105, 111, 128, 139, 167, 180, 184, 188, 197, 199, 201, 203, 207, 228, 231, 235, 237-243, 245-251, 253] [59, 179, 203, 204, 233, 234] [7, 32, 82, 91, 120, 121, 124, 133, 147, 150, 152, 165, 176, 214, 216, 217, 229, 230, 240, 252, 254-261] Force (pounds) [21] Force (kg)/body-weight (kg) [168] % of people with force scores (kg, classified as weakness) [196] kg/cm^2 [71]	R	1-2	130821	34-89 (60.4-70.5)	75538 F, 49439 M

	mean of two trials [235]; mean of three trials [236] - dominant hand [176]; arm by side; best of three trials [168, 237] - dominant hand, sitting in an upright position, arm of the measured hand unsupported and parallel to the body; one trial [70] - dominant hand, sitting, dominant shoulder in rest position, elbow flexed 90° without support, forearm and wrist at neutral position; best of three trials [201] - dominant hand, sitting comfortably, dominant arm by side, elbow flexed 90°, hand held in mid-supination/pronation position; best of three trials [180] - both hands, one trial in each hand; best score [63] - both hands; three trials in each hand, best score for each hand [238]; best trial of the dominant hand [239] - best score of both hands [203]; best score of each hand [240]; sum of best score of each hand [111]; mean score of each hand [241] - both hands, wrist in neutral position, elbow flexed at 90°; three trials for each hand; mean of each hand [242] - both hands, two trials for each hand, mean for each hand and larger mean from one of the hands [207] - N/A [28, 91] Bulb hand dynamometer - both hands, holding at shoulder level, two trials in each hand; mean of both hands added [243] - dominant hand, medium (women) or large (men) dynamometer, sitting; best of three trials [71] - both hands, three trials in each hand; mean for each hand [244] - both hands, two trials in each hand; mean for each hand [245] Calibrated dynamometer	Force (pound per square) [244] Force (kPa) [28] Force (Newton) [63] Quartiles (% lower quartile reported) [236]

	- both hands, elbow flexed to 90° [167] - preferred hand, arm raised overhead then slowly lowered towards floor [21]; three trials [32] Handheld dynamometer - both hands, standing, best of two trials (each hand) [2, 59, 69, 128]; mean of three trials (each hand) [110] - both hands, two trials on the dominant and three trials on each hand [246] - two trials; mean score [197]; best score [120, 247] - dominant hand [179]; mean of two trials [248]; mean of three trials [249]; best of two trials [133] - sitting, elbow flexed two 90°, best of two trials [204]; best of three trials [250, 251] - standing with their forearms bent at 90 °, best of two trials [7] - elbow positioned at 90° of the side of the body; dominant hand; mean of three trials [252] N/A [48, 103, 139, 150, 152, 165, 214, 216, 217, 253-261]						
Shoulder flexor strength (1 study)	Right arm, 90° shoulder flexion, elbow in full extension; mean of three trials [47]	Force (kg)	R	1	85	65-84 (69.0±0.4)	37 F, 48 M
Hip muscle strength (2 studies)	Supine on a plinth, both legs 10° abducted, a strap (5 cm wide) around the plinth and over the pelvis; for the examiner-resisted test, participants pushed as hard as possible against the HHD as the examiner provided resistance, stabilizing and positioning the HHD; for the belt-resisted test, HHD is placed between the side of the test leg and a second strap (5cm width), participants spread their legs apart simultaneously as hard as possible [262] Right, and left hip abduction strength on isotonic external resistance machine [263]	Torques kg/body mass [263]	R	2	45	55-75 (63.7-68.4)	31 F, 14 M
Knee extensor strength (1 study)	Computer-based manual muscle testing, knee at 30° flexion; mean of three trials [47]	Force (kg)	R	1	85	65-84 (69.0±0.4)	37 F, 48 M

Leg strength (6 studies)	Sitting in a standard chair (45 cm high), connected to a WBB (57° angle from the ground) via custom seatbelt straps; pressing the feet on the WBB as hard and as fast as possible; three trials [264] Dynametry [254] Leg press machine [34, 145, 176, 263]	Force (kg) Kg/body mass [263]	R	1	272	55-75 (61.1-69.3)	140 F, 76 M
Toe grasping strength (2 studies)	Barefoot, one-leg stand, both hands on the wall in front, holding the dynamometer grasping bar with the toes [265] Sitting upright on a chair, without leaning on the backrest; both hips and knees flexed at approximately 90°; ankles placed in a neutral position and fixed with a strap [107]	Force (kg)	R	1	7227	52-78 (66.3-67.6)	534 F, 188 M
Maximal Isometric Strength (MIS)							
Elbow extensor strength (1 study)	Instrumented wooden pole that subjects pressed against the ground; Subjects sit on a bench, with the shoulder in a neutral position, elbow angle of 90°, forearm was parallel to the ground; best of three trials [146]	MVC (kg)	R	1	26	69.2-70.0	17 F, 9 M
Hip extensor strength (1 study)	HHD, mean score [178]	% (strength/body weight)	R	1	39	60-78 (68.5-69.7)	15 F, 24 M
Hip flexor strength (2 studies)	HHD [91], mean score [178] HHD, sitting position, hip flexed at 90° and knee flexed at 90°; sensor of the HHD was applied to a distal site on the anterior surface of the thigh [55]	% (strength/body weight) Force (kg) [55, 91]	R	1	818	60-78 (68.5-69.7)	775 F, 313 M
Hip abductor strength (2 studies)	Supine position with the hip and knee fully extended and hip positioned in neutral abduction; sensor of the HHD was applied 5 cm proximal from the lateral malleolus [55] HHD, three trials (each hold for 3-5 s) [101]	Force (kg)	R	1	744	61.8-68.7	482 F, 262 M
Knee extensor strength (11 studies)	HHD, sitting upright, raising lower legs up 90°, parallel to the ground, holding this position as strongly as possible against the maximum persistent (5 s) force applied by the examiner	Force (kg) [34, 66, 107, 232, 266] % (strength/body weight) [100, 178]	R	1-2	1595	60-78 (61.1-71.6)	1038 F, 402 M

	through the HHD placed on the front of the ankle proximal to the medial malleolus; two trials for each leg, best score [266] HHD [91, 107]; mean score [10, 178] Leaning back in a chair, extending both legs at the knee while pulling against a dynamometer; best of two trials [66] Sitting on a high chair and pushing against a strap linked to a spring gauge [100] Seated position using an adjustable chair with a 90° angle of hip and knee joints, dominant leg, as fast as forcefully, strongest of five trials [34] Dynamometer, right side; sitting on a backwardly-inclined (5°) chair, range of motion was set from a knee joint angle of 90° to 160° (180° represents full extension); best of two trials [145] Sitting in an upright position with back support and with both the hip and knee flexed at 70°; distal leg affixed to a strain gauge force transducer; best of three trials [146] Dominant limb [232]	N/kg [10] Peak torque (Nm) [145, 232] MVC (kg) [146]					
Knee flexor strength (1 study)	HHD, mean score [178]	% (strength/body weight)	R	1	39	60-78 (68.5-69.7)	15 F, 24 M
Leg strength (6 studies)	Dynamometer, both legs simultaneously [267]; mean of two trials [245]; best of two trials [268] Dynamometer, both legs simultaneously, standing with back straight against a wall and knees 115° flexed; a bar connected by a chain to the dynamometer was held in front of the thighs and has to be lifted upwards with maximum force using only the legs, and keeping the neck and back straight; mean of two trials [243, 269] Fitted with the harness around their hips and seated in a standard chair (45 cm height) with the seatbelt straps connecting the harness to the FysioMeter-mount. The lengths of the seatbelt straps were adjusted using a tape measure and a goniometer angle between sessions to reach a knee angle of approximately 120°; foot	Force (kg)	R	1	2544	50-79 (61.4-69.0)	1230 F, 1277 M

	placed in the middle of the Wii Balance Board [270]						
Ankle dorsiflexor strength (7 studies)	HHD [91, 149]; mean score [10, 178]; mean of three trials [90, 112]; three trials (each for 3-5 s) [101]	% (strength/body weight) [178] N/kg [10] Force (kg) [90, 91, 101, 112, 149]	R	1	357	60-78 (61.8-69.7)	222 F, 135 M
Ankle plantar flexor strength (5 studies)	HHD [91, 142, 149]; mean of three trials [90, 112]; three trials (each for 3-5 s) [101]	Force (kg)	R	1	832	50-80 (61.8-68.5)	450 F, 392 M
Functional muscle power							
Upper body functional muscle power							
30 second arm curl (20 studies)	Part of the SFT [42, 45, 154, 156, 160, 162-164, 166, 170, 271] Performing as many biceps curls as possible in 30 s, using a 2.27-kg dumbbell (full range of motion; study in women) [169, 272] Flexing and extending the elbow of the dominant hand, lifting a weight (8 lb [3629g] dumbbell for men; 5lb dumbbell [2268g] for women) through the complete range of motion as many times as possible in 30 s [173] Sitting on a chair, using the dominant hand to bring a weight (2.0 kg) up and down (flex and extend the biceps) as many times as possible in 30 s [179] Hand curling a hand weight (5 pounds for women and 8 pounds for men) for 30 s [131] Siting on the chair, holding the dumbbell (women 5 lbs, men 8 lbs) in the hand with palm facing towards the body (with the arm in a vertically down position beside the chair), bracing the upper arm against the body so that only the lower arm is moving, then curling the arm up through a full range of motion, gradually performing elbow flexion with supination; as the	Repetitions	R	1	5768	51-89 (61.9.0-69.9)	2652 F, 1111 M

	arm is lowered through the full range of motion, gradually return to the starting position [152] N/A [144, 175]						
Abdominal Strength (2 studies)	Lying down on an abdominal pad, with knees flexed at 90°, hands set on the pad frame. Rise with the chest up to approximately 30° from the floor as many times as possible in 30 s [167] Lying on sit-up equipment and performing sit-ups with the feet attached to the equipment's foot holders as many times as possible in 30 s [235]	Repetitions	R	1	252	59-60+ (63.0-66.9)	230 F, 122 M
Single forearm contractions (1 study)	Dynamic single contractions in both arms, HHD at 10%, 20%, and 40% of the subject's maximum voluntary contraction [273]	MVC (kg)	R	1	32	59-85 (66.0±2)	13 F, 19 M
Seated medicine ball throw (1 study)	Participants horizontally propel a 3 kg ball without trunk flexion [144]	Distance (m)	R	1	36	68.8-68.9	20 F, 16 M
Lower body functional muscle power							
Five times Sit-to-Stand (61 studies)	Part of the SPPB [27, 28] Rising from a chair and sitting back down five times [9, 34, 112, 168]; without arm support [147, 240, 274, 275]; time measured at the final sitting down, best of three trials [127]; mean of three trials [90] Five repetitive chair stands as quickly as possible with arms folded across the chest [1, 7, 41, 63, 70, 203, 246, 249]); mean of two trials [8, 276]; mean of three trials [95, 188] Sitting in a standard chair, arms folded across the chest, standing up and sitting down five times [124]; as fast as possible [247] Sitting in the middle of an armless folding chair, then standing up and sitting down with an arm folded in front of the chest; mean of two trials [36]	Time (s) [1-3, 5, 6, 9, 23, 27, 34, 36, 41, 53, 60, 63, 70, 81, 85, 90, 91, 93-95, 109, 112, 118, 124, 125, 127, 136, 137, 141, 142, 145, 147, 149, 168, 188, 203, 215, 232, 240, 246, 247, 249, 260, 261, 266, 274, 275, 277-284] Stands/min (60*[5/time, s]) [7] Score (0-4) [28]	R	1	81289	40-90+ (58.7-71.0)	41301 F, 36656 M

	Standing all the way up and sitting all the way down 5 times as quickly as possible while keeping the arms folded across the chests; straight back, flat, level, firm seat (45 cm high) [215]	
	Getting up and sitting from a chair (43 cm high, flat seat), arms crossed over the chest, rising until full extension at trunk and lower limb joints, and returning with the back fully supported at the back of the chair; best of two trials [136]	
	Standard chair (43.2 cm high), transferring to a standing position and returning to a sitting position, not allowed to use arms [277]	
	Standard padded chair (43.2 cm high) without armrests, both arms crossed against the chest, starting from a seated position and standing up (legs straight) and sitting down (full weight on the chair) [278]	
	Getting up from and sitting down on the chair (43.6 cm high) without arm rests [137]	
	Standing and sitting five times from an armless chair (46 cm high), not permitted to use arms [125]	
	Straight-back chair, placed against a wall, with a hard seat and standard height, sitting with the feet on the floor and arms folded across the chests on the chair, time measured at the final standing position [2]	
	Standing up and sitting down as quickly as possible five times in a row from an armless straight-back chair, arms across the chest, time measured at the final standing position [5]; time measured at the final sitting position [3]	
	Sitting in a hard-backed chair (43 cm high), arms folded across the chest, rising as fast as possible to a full standing position, then returning to a full-sitting position five times [93, 94]	
	Rising fully from a standard armless, backless chair five times as fast as possible, arms folded closely to the trunk, no moving of the feet during the test, time measured at the final sitting position [38]	

	Standing up from a straight-backed chair (43 cm high) five times at a self-selected pace, arms folded across the chest [53] Sitting on a chair with the back touching the backrest, seat height adjusted to participant's lower leg length, knees flexed at 90°, time measured at the final sitting position [279] Standing up and sitting down five times as quickly as possible from a straight-backed chair [280]; time measured at the final standing position [23] Standard chair with arm rests, both arms crossed against the chest, starting from a seated position (upper back against seat), standing up to full extension and sitting down again (upper back against seat), best of two trials [118] N/A [6, 60, 81, 85, 91, 109, 141, 142, 145, 149, 232, 260, 261, 266, 281-283]					
One time sit-to-stand (7 studies)	Sitting in a straight-back chair, barefooted, on cue, standing up and sitting down as quickly as possible, upper extremity use not permitted [21] Sitting on a chair (43 cm high), on cue, rising to full stance; best of three trials [285] Adjusted seat height (5 cm increments from 45 to 60 cm) to achieve a 90/90 (hip/knee angle), sitting on the front half of an instrumented chair, using the arms as normally during the task, while standing as quickly as possible, three trials [126] Chair rise from different seat heights (43 cm, 38 cm, 30 cm) [286] Standing up as quickly as possible from a standard chair (43 cm high), arms crossed across the chest and feet shoulder-width apart placed flat on the floor [4] N/A [144, 287]	Time (s) [4, 21, 285-287] Force (N/s [kg]; W [kg]) [126, 144]	R	1	414	60-74 (61.6-69.9)	235 F, 79 M
Ten times sit-to-stand (6 studies)	Rising from a sitting to a standing position with straight back and legs and sitting down again as fast as possible [49]	Time (s) [86, 229, 250, 251]	R	1	3283	50-81 (62.6-69.0)	1182 F, 1012 M

	Straight-backed chair (45cm high), arms crossed against the chest, rising as quickly as possible without the use of the hands [215, 250, 251] Rising from a chair as quickly as possible with arms placed across the chest [229] N/A [86]	Speed (stands per minute: [10/s]*60) [49]					
15 second Sit-to-stand (1 study)	Straight-backed, non-padded, flat-seated, arm-less chair, Standing without using hands or arms, arms folded across the chest; mean of two trials [197]	Repetitions	R	1	5777	65-79 (69.8-70.1)	5777 F
30 second sit-to-stand (51 studies)	Part of the SFT [42, 45, 57, 154-156, 158-164, 166, 271]; two trials [157] Part of the Fullerton Functional Fitness Test [170] Standing in front of a stable chair, hands across the chest, then practicing sitting down and standing up for 30 s [31, 272] Sitting in a chair (43 cm high) with arms crossed at the wrists and holding against the chest, then standing up as many times as possible [171, 172, 288]) Sitting on a standard armless chair (45 cm high), looking straight forward with arms folded across the chest, then standing up and sitting down as many times as possible [167] Rising up and sitting down with arms folded in front of the chest as quickly as possible on a firm, armless chair placed against a wall [82, 289] Standing up and sitting down from a bench with-out armrests and back support as many times as possible, feet flat on the floor, initial foot placement and chair height individually adjusted [290] Stand up from a seated position as many times as possible [291] Stand fully upright and then return to the seated position as many times as possible [66, 67, 87, 173, 174]	Repetitions	R	1	7493	51-91 (61.1-71.6)	3730 F, 1697 M

	Different chair heights (43 cm; then adjusted to 80, 90, 100, 110 and 120% of the participants' lower leg length), last attempt at the end of 30 s is counted as a full stand if the participant is more than halfway up from sitting [284] Standard padded chair (43.2 cm high) without armrests, starting from the seated position and standing up (legs straight) and sitting down (full weight on the chair); mean of two trials [278] Sitting on a chair, back straight, feet shoulder-width apart and flat on the floor, arms crossed at the wrists and held against the chest, then rising to a full stand and returning to a fully seated position as many times as possible [179] Chair (44 cm high) without arms, sitting in the middle of the chair, feet shoulder width apart and placed on the floor at an angle slightly behind the knees, arms crossed at the wrists and held against the chest, then rising to a full stand and returning back to the initial seated position, as many full stands as possible; mean of two trials [16, 17] Sitting in the middle of the chair, arms across the chest, then rising to a full stand and returning to a fully seated position as many times as possible [131] Standard chair with arm rests, both arms crossed against the chest, starting from a seated position (upper back against seat), standing up to full extension and sitting down again (upper back against seat); best of two trials [118] Sitting in a standard-height chair with arms crossed over the chest, then stand fully and sit down again as many times as possible [97] N/A [10, 26, 43, 96, 144, 145, 152, 153, 175, 232, 234]						
1 minute sit-to-stand (2 studies)	Stand up from and sit down from a standard height chair without the use of the arms [292] Sitting on the edge of a standard-height chair, arms crossed over the chest, and repeatedly	Repetitions	R	1	123	55-70 (62.2-70.7)	76 F, 47 M

	standing up from and returning to a seated position without assistance [4]						
One time kneel-to-stand (1 study)	Part of MOD scale [286]	Score (0-5)	R	1	259	60+ (67.6±7.0)	143 F, 116 M
Floor rise to standing (6 studies)	Lying in a supine position, with feet together and hands palm down and at the side, then rising to a standing position [93, 94] Sit and rise from the floor (flat, non-slippery surface), using the minimum support needed [112] Stand up from a supine position [282] Sitting and rising unassisted from the floor with partial scores assigned from the two required actions of sitting (5 points) and rising (5 points) and a final composite SRT score [153] N/A [141]	Time (s) Score [112, 153]	R	1	172	58-84 (67.0-69.3)	50 F,49 M
Five Step Test (1 study)	N/A [266]	Time (s)	R	1	621	50+ (66.8-69.4)	428 F, 193 M
Stair climbing (2 studies)	Walking up and down a standard flight of stairs, three times at self-selected pace, using the handrail for support only if needed [53, 216]	Time (s)	R	1	1143	55-79 (63.8-67.5)	634 F, 509 M
Stair climbing (8 steps) (2 studies)	Climbing eight steps (17 cm high, 31cm long) without using the handrail, requiring a step by step pattern; best of two trials [136] N/A [261]	Time (s)	R	1	111	65.6-67.8	50 F, 35 M
Stair climbing (10 steps) (3 studies)	Climbing a flight of stairs (10 steps) as quickly as possible without using the handrails or any other aid (14 cm high [87]; 7.8 cm high [89] Ascending and descending a flight of stairs (10 steps, 0.27 m high and 0.18 m deep) as quickly and safely as possible, while having the option of using a single handrail for support [293]	SCP (W) [87, 89] Time (s) [293]	R	1	212	50-75 (62.7-71.5)	152 F, 67 M
Stair climbing (11 steps)	Ascending a standard fight of stairs (11 stairs, 16 cm high), avoiding the use of the handrail [93, 94]; as rapidly as possible [92]	Time (s) [92-94] SCP (W) [93, 94]	R	1	77	65-84 (68.9-69.3)	37 F,40 M

(3 studies)							
Stair climbing (12 steps) (2 studies)	Ascending and descending 12 stairs, permitted to use the handrail, but not allowed to use it to push or pull oneself [277] Ascending and descending 12 stairs (15cm high, 20 cm tread), as fast as possible while being safe [102]	Time (s)	R	1	337	45-80 (58.7-64.8)	183 F, 164 M
Stair climbing (14 steps) (1 study)	Walk as fast as possible up 14 stairs without the use of railings [289]	Time (s)	R	1	30	68.5±5.1	15 F, 15 M
Stair climbing (15 steps) (1 study)	Ascending and descending a flight of 15 stairs (18 cm high, 27 cm tread) at normal pace, preferably without using the handrail [135]	Time (s)	R	1	134	69.6-70.3	85 F, 49 M
Stair ascent (23 steps) (1 study)	Walking up one flight of stairs consisting of 23 steps (16.5 cm high, 19.2 cm wide) as quickly as possible; after 14 steps, the participants make a left-hand wrap-around turn and then completed the remaining nine steps; not allowed to use the handrails; best of the two trials [294]	Time (s)	R	1	62	60-83 (66.6-71.0)	N/A
Stair ascent (16 steps) (1 study)	16 steps, height of 15 cm; not allowed to hold the handrails [143]	Time (s)	R	1	48	60-80 (68.6±6.1)	N/A
Stair ascent (10 steps) (4 studies)	Ascending a 10-stair prop (17 cm high, 30 cm deep) at fast pace [168] Walking up 10 steps in an expeditious and safe manner, placing one hand close to the handrail for balance if necessary, but not on the handrail [275] Ascending a 10-stair flight (16.5 cm stair high) as fast as possible, use of handrail allowed [70] Climbing 10 steps as fast as comfortably possible with one hand near, but not on, the handrail [200]	Time (s)	R	1	158	62-80 (66.0-70.0)	69 F, 35 M

Stair ascent (9 steps) (2 studies)	Walking quickly but safely up and down a nine step flight of stairs (step height: 17 cm); time started after the cue to go and stopped when the second foot reached the top step [147, 283]	Time (s)	R	1	71	62.7-70.0	46 F, 25 M
Stair ascent (4 steps) (1 study)	Walking up 4 stairs (15 cm high), arriving on a full stance on the fourth step without any support or help, three trials, best score [285]	Time (s)	R	1	33	60-74 (64.4-65.7)	21 F, 12 M
Stair ascent (one time) (1 study)	Part of MOD scale [286]	Score (0-5)	R	1	259	60+ (67.6±7.0)	143 F, 116 M
Stair descent (16 steps) (1 study)	16 steps, height of 15 cm; not allowed to hold the handrails [143]	Time (s)	R	1	48	60-80 (68.6±6.1)	N/A
Stair descent (14 steps) (1 study)	14 steps (height 17 cm, length 30 cm); time starts with participants' initiation of first step and stops when both feet are on the landing [282]	Time (s)	R	1	33	67±4.5	N/A
Stair descent (10 steps) (1 study)	Walking down 10 steps in an expeditious and safe manner, placing one hand close to the handrail for balance if necessary, but not on the handrail [275]	Time (s)	R	1	19	66.0±1.0	14 F, 5 M
Stair descent (9 steps) (1 study)	Walking quickly but safely up and down a nine step flight of stairs (step height: 17 cm); time started after the cue to go and stopped when the second foot reached the floor [147]	Time (s)	R	1	48	69.8-70.0	26 F, 22 M
Stair descent (one time) (1 study)	Part of MOD scale [286]	Score (0-5)	R	1	259	60+ (67.6±7.0)	143 F, 116 M
Functional leg extensor strength (1 study)	Taking a short step forward, first with the right leg, squat down until the knee of the tracking leg lightly touches the mat, and then rise up immediately and step back to the starting position, then repeating with the left leg [53]	Maximal weight relative to the subject's body weight	R	1	1133	55-79 (63.8-64.1)	632 F, 501 M
Lift and reach (one minute) (2 studies)	Sitting at a standard height desk, then repeatedly lifting a weight onto and off a shelf placed on the desk located at shoulder level immediately in front (10 pound for women, 20 pound for men) [292]	Repetitions	R	1	123	55-70 (62.6-70.7)	76 F, 47M

	Sitting in a standard chair at a standard height desk (75 cm), then lifting a weight repeatedly onto and off a shelf positioned at approximate shoulder height, 37 cm above the desktop (5 kg dumbbell for women, 8 kg dumbbell for men) [4]						
Standing long jump (2 studies)	Jumping horizontally, using a 2-ft. takeoff and landing, three trials, measured at the heel of the foot [32] Jumping with no restriction to arm movements as high and as fast without bending legs in air and landing with both feet on the jump mat; average of three trials [263]	Distance (cm [32]; m [263]) Power (kg/body mass), Velocity (m/s) [263]	R	1	98	50-79 (63.7±1.1)	54 F, 44 M
Squat jump (1 study)	Static position, knee bent in a 90° angle, hands on the hip during the whole jump; three maximal trials separated by 1.5 min of rest [34]	Maximal Ground Reaction Force (A; N*kg-1), Rate of Force Development (N*kg-1), Force (N)	R	1	63	65-70 (67.5±0.4)	63 F
Single knee extension contractions (1 study)	Single knee extension contractions with a handgrip device at 20%, 40%, and 60% of the subject's max voluntary contraction [273]	Maximum work rate (WR_{max})	R	3	32	59-85 (66.0±2.0)	12 F, 19 M

[a] [a]The total number included was the total number of participants in all studies per balance test; [b]M: Male; F: Female; (sex was not reported in all papers, and the number will differ from total included in [a]); OLS: One-leg standing balance; SEBT: Star Excursion Balance Test; TUG: Timed Up and Go; FRT: Functional Reach Test; LRT: Lateral Reach Test; SOT: Sensory Organization Test; BBS: Berg Balance Scale; SPPB: Short Physical Performance Battery; PPT: Physical Performance Test; mPPT: modified Physical Performance Test; FAB: Fullerton Advanced Balance; CS-PFP-10: Continuous Scale-Physical Functional Performance-10 item test; PPB: Physical Performance Battery; CBM: Community Balance & Mobility scale; FFM : Functional Movement Measurement; N: Nominal; O: Ordinal; R: Ratio; RMS: Root Mean Square; MPF: Mean Power Frequency; CoP: Center of Pressure; CoG: Center of Gravity; MSL: Maximum Step Length; ft: feet; max: maximum; s: seconds; rep: repetitions; %: percentage; cm: centimeter; HHD: Hand Held Dynamometer; WBB: Wii Balance Board; SFT: Senior Fitness Test; MVC: Maximum Voluntary Contraction; MOD scale: Modification scale; SCP: Stair Climbing Power; n: number of participants; N/A: not applicable.

References

1. Al Snih S, Kaushik V, Eschbach K, Markides K: **Ethnic differences in physical performance in older Americans: data from the Third National Health and Nutrition Examination Survey (1988-1994)**. *Aging-Clinical & Experimental Research* 2008, **20**(2):139-144.
2. Keevil VL, Hayat S, Dalzell N, Moore S, Bhaniani A, Luben R, Wareham NJ, Khaw KT: **The physical capability of community-based men and women from a British cohort: the European Prospective Investigation into Cancer (EPIC)-Norfolk study**. *BMC Geriatrics* 2013, **13**:93.
3. Nickel KJ, Acree LS, Montgomery PS, Gardner AW: **Association between lower-extremity function and arterial compliance in older adults**. *Angiology* 2008, **59**(2):203-208.
4. Ritchie C, Trost SG, Brown W, Armit C: **Reliability and validity of physical fitness field tests for adults aged 55 to 70 years**. *Journal of Science & Medicine in Sport* 2005, **8**(1):61-70.
5. Sharkey JR, Ory MG, Branch LG: **Severe elder obesity and 1-year diminished lower extremity physical performance in homebound older adults**. *Journal of the American Geriatrics Society* 2006, **54**(9):1407-1413.
6. Tiedemann A, O'Rourke S, Sesto R, Sherrington C: **A 12-week Iyengar yoga program improved balance and mobility in older community-dwelling people: a pilot randomized controlled trial**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2013, **68**(9):1068-1075.
7. Keevil VL, Luben R, Hayat S, Sayer AA, Wareham NJ, Khaw K-T: **Physical capability predicts mortality in late mid-life as well as in old age: Findings from a large British cohort study**. *Archives of Gerontology and Geriatrics* 2018, **74**:77-82.
8. Goldberg A, Talley SA: **Performance on a test of rapid stepping in community-dwelling older adults: validity, relative and absolute reliability and minimum detectable change**. *Physiotherapy Theory & Practice* 2015, **31**(7):483-488.
9. Alfred T, Ben-Shlomo Y, Cooper R, Hardy R, Cooper C, Deary IJ, Gaunt TR, Gunnell D, Harris SE, Kumari M *et al*: **A multi-cohort study of polymorphisms in the GH/IGF axis and physical capability: the HALCyon programme**. *PLoS ONE [Electronic Resource]* 2012, **7**(1):e29883.
10. Scaglioni-Solano P, Aragon-Vargas LF: **Gait characteristics and sensory abilities of older adults are modulated by gender**. *Gait & Posture* 2015, **42**(1):54-59.
11. Won H, Singh DK, Din NC, Badrasawi M, Manaf ZA, Tan ST, Tai CC, Shahar S: **Relationship between physical performance and cognitive performance measures among community-dwelling older adults**. *Clinical Epidemiology* 2014, **6**:343-350.
12. Hsieh KL, Roach KL, Wajda DA, Sosnoff JJ: **Smartphone technology can measure postural stability and discriminate fall risk in older adults**. *Gait & Posture* 2019, **67**:160-165.
13. McBean AL, Najjar RP, Schuchard RA, Hall CD, Wang C-A, Ku B, Zeitzer JM: **Standing balance and spatiotemporal aspects of gait are impaired upon nocturnal awakening in healthy late middle-aged and older adults**. *Journal of Clinical Sleep Medicine* 2016, **12**(11):1477-1486.
14. Quek J, Treleaven J, Clark RA, Brauer SG: **An exploratory study examining factors underpinning postural instability in older adults with idiopathic neck pain**. *Gait & Posture* 2018, **60**:93-98.
15. Fleury A, Mourcou Q, Franco C, Diot B, Demongeot J, Vuillerme N: **Evaluation of a Smartphone-based audio-biofeedback system for improving balance in older adults--a pilot study**. *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine & Biology Society* 2013, **2013**:1198-1201.

16. Nicholson VP, McKean MR, Burkett BJ: **Twelve weeks of BodyBalance training improved balance and functional task performance in middle-aged and older adults**. *Clinical Interventions In Aging* 2014, **9**:1895-1904.
17. Nicholson VP, McKean MR, Burkett BJ: **Low-load high-repetition resistance training improves strength and gait speed in middle-aged and older adults**. *Journal of Science & Medicine in Sport* 2015, **18**(5):596-600.
18. Hageman PA, Leibowitz JM, Blanke D: **Age and gender effects on postural control measures**. *Archives of Physical Medicine & Rehabilitation* 1995, **76**(10):961-965.
19. Shaw JA, Stefanyk LE, Frank JS, Jog MS, Adkin AL: **Effects of age and pathology on stance modifications in response to increased postural threat**. *Gait & Posture* 2012, **35**(4):658-661.
20. Quek JM, Pua YH, Bryant AL, Clark RA: **The influence of cervical spine flexion-rotation range-of-motion asymmetry on postural stability in older adults**. *Spine* 2013, **38**(19):1648-1655.
21. Brotherton SS, Williams HG, Gossard JL, Hussey JR, McClenaghan BA, Eleazer P: **Are measures employed in the assessment of balance useful for detecting differences among groups that vary by age and disease state?** *Journal of Geriatric Physical Therapy* 2005, **28**(1):14-19.
22. Scaglioni-Solano P, Aragon-Vargas LF: **Validity and reliability of the Nintendo Wii Balance Board to assess standing balance and sensory integration in highly functional older adults**. *International Journal of Rehabilitation Research* 2014, **37**(2):138-143.
23. Stijntjes M, de Craen AJ, van der Grond J, Meskers CG, Slagboom PE, Maier AB: **Cerebral Microbleeds and Lacunar Infarcts Are Associated with Walking Speed Independent of Cognitive Performance in Middle-Aged to Older Adults**. *Gerontology* 2016, **15**:15.
24. Medell JL, Alexander NB: **A clinical measure of maximal and rapid stepping in older women**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2000, **55**(8):M429-433.
25. Abe T, Ogawa M, Loenneke JP, Thiebaud RS, Loftin M, Mitsukawa N: **Association between site-specific muscle loss of lower body and one-leg standing balance in active women: the HIREGASAKI study**. *Geriatrics & gerontology international* 2014, **14**(2):381-387.
26. Balachandran A, C NV, Potiaumpai M, Ni M, Signorile JF: **Validity and reliability of a video questionnaire to assess physical function in older adults**. *Experimental Gerontology* 2016, **16**:16.
27. Martin H, Aihie Sayer A, Jameson K, Syddall H, Dennison EM, Cooper C, Robinson S: **Does diet influence physical performance in community-dwelling older people? Findings from the Hertfordshire Cohort Study**. *Age & Ageing* 2011, **40**(2):181-186.
28. St-Arnaud-McKenzie D, Payette H, Gray-Donald K: **Low physical function predicts either 2-year weight loss or weight gain in healthy community-dwelling older adults. the NuAge Longitudinal Study**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2010, **65**(12):1362-1368.
29. Vaughan S, Wallis M, Polit D, Steele M, Shum D, Morris N: **The effects of multimodal exercise on cognitive and physical functioning and brain-derived neurotrophic factor in older women: a randomised controlled trial**. *Age & Ageing* 2014, **43**(5):623-629.
30. Jalali MM, Gerami H, Heidarzadeh A, Soleimani R: **Balance performance in older adults and its relationship with falling**. *Aging-Clinical & Experimental Research* 2015, **27**(3):287-296.
31. Chen HT, Lin CH, Yu LH: **Normative physical fitness scores for community-dwelling older adults**. *Journal of Nursing Research* 2009, **17**(1):30-41.
32. Rudisill ME, Toole T: **Gender differences in motor performance of 50- to 79-year-old adults**. *Perceptual & Motor Skills* 1993, **77**(3 Pt 1):939-947.

33. Liu XY, Gao J, Yin BX, Yang XY, Bai DX: **Efficacy of Ba Duan Jin in improving balance: A study in Chinese community-dwelling older adults**. *Journal of Gerontological Nursing* 2016, **42**(5):38-46.
34. Edholm P, Strandberg E, Kadi F: **Lower limb explosive strength capacity in elderly women: effects of resistance training and healthy diet**. *Journal of Applied Physiology* 2017, **123**(1):190-196.
35. Wayne PM, Gow BJ, Costa MD, Peng CK, Lipsitz LA, Hausdorff JM, Davis RB, Walsh JN, Lough M, Novak V *et al*: **Complexity-based measures inform effects of Tai Chi training on standing postural control: Cross-sectional and randomized trial Studies**. *PLoS One* 2014, **9**(12):e114731.
36. Kokai Y, Mikami N, Tada M, Tomonobu K, Ochiai R, Osaki N, Katsuragi Y, Sohma H, Ito YM: **Effects of dietary supplementation with milk fat globule membrane on the physical performance of community-dwelling Japanese adults: a randomised, double-blind, placebo-controlled trial**. *Journal of Nutritional Science* 2018, **7**.
37. Schilling BK, Falvo MJ, Karlage RE, Weiss LW, Lohnes CA, Chiu LZ: **Effects of unstable surface training on measures of balance in older adults**. *Journal of Strength & Conditioning Research* 2009, **23**(4):1211-1216.
38. Kuo YL, Huang KY, Chiang PT, Lee PY, Tsai YJ: **Steadiness of spinal regions during single-leg standing in older adults with and without chronic low back pain**. *PLoS One* 2015, **10**(5):e0128318.
39. Fujisawa T, Suzuki S, Tanaka K, Kamekura N, Fukushima K, Kemmotsu O: **Recovery of postural stability following conscious sedation with midazolam in the elderly**. *Journal of Anesthesia* 2002, **16**(3):198-202.
40. Hardy R, Cooper R, Aihie Sayer A, Ben-Shlomo Y, Cooper C, Deary IJ, Demakakos P, Gallacher J, Martin RM, McNeill G *et al*: **Body mass index, muscle strength and physical performance in older adults from eight cohort studies: the HALCyon programme**. *PLoS ONE [Electronic Resource]* 2013, **8**(2):e56483.
41. Demnitz N, Zsoldos E, Mahmood A, Mackay CE, Kivimäki M, Singh-Manoux A, Dawes H, Johansen-Berg H, Ebmeier KP, Sexton CE: **Associations between mobility, cognition, and brain structure in healthy older adults**. *Frontiers in Aging Neuroscience* 2017, **9**:155.
42. Chen TC-C, Tseng W-C, Huang G-L, Chen H-L, Tseng K-W, Nosaka K: **Superior effects of eccentric to concentric knee extensor resistance training on physical fitness, insulin sensitivity and lipid profiles of elderly men**. *Frontiers in Physiology* 2017, **8**:209.
43. Paes T, Belo LF, da Silva DR, Morita AA, Donária L, Furlanetto KC, Sant'Anna T, Pitta F, Hernandes NA: **Londrina activities of daily living protocol: reproducibility, validity, and reference values in physically independent adults Age 50 years and older**. *Respiratory Care* 2017:respcare. 05059.
44. Porto JM, Júnior RCF, Bocarde L, Fernandes JA, Marques NR, Rodrigues NC, de Abreu DCC: **Contribution of hip abductor–adductor muscles on static and dynamic balance of community-dwelling older adults**. *Aging Clinical and Experimental Eesearch* 2018:1-7.
45. Gothe NP, McAuley E: **Yoga Is as good as stretching-strengthening exercises in improving functional fitness outcomes: results from a randomized controlled trial**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2016, **71**(3):406-411.
46. Santoni G, Angleman SB, Ek S, Heiland EG, Lagergren M, Fratiglioni L, Welmer A-K: **Temporal trends in impairments of physical function among older adults during 2001–16 in Sweden: towards a healthier ageing**. *Age and Ageing* 2018.
47. Hulya TD, Sevi YS, Serap A, Ayse OE: **Factors affecting the benefits of a six-month supervised exercise program on community-dwelling older adults: interactions among age, gender, and participation**. *Journal of Physical Therapy Science* 2015, **27**(5):1421-1427.

48. Coelho-Junior HJ, Rodrigues B, de Oliveira Gonçalves I, Asano RY, Uchida MC, Marzetti E: **The physical capabilities underlying timed "Up and Go" test are time-dependent in community-dwelling older women**. *Experimental Gerontology* 2018, **104**:138-146.
49. Cooper AJ, Simmons RK, Kuh D, Brage S, Cooper R, scientific N, data collection t: **Physical activity, sedentary time and physical capability in early old age: British birth cohort study**. *PLoS ONE [Electronic Resource]* 2015, **10**(5):e0126465.
50. Yıldırım P, Ofluoglu D, Aydogan S, Akyuz G: **Tai chi vs. combined exercise prescription: a comparison of their effects on factors related to falls**. *Journal of Back and Musculoskeletal Rehabilitation* 2016, **29**(3):493-501.
51. Chun SH, Cho B, Yang H-K, Ahn E, Han MK, Oh B, Shin DW, Son KY: **Performance on physical function tests and the risk of fractures and admissions: findings from a national health screening of 557,648 community-dwelling older adults**. *Archives of Gerontology and Geriatrics* 2017, **68**:174-180.
52. Benavent-Caballer V, Rosado-Calatayud P, Segura-Orti E, Amer-Cuenca JJ, Lison JF: **The effectiveness of a video-supported group-based Otago exercise programme on physical performance in community-dwelling older adults: a preliminary study**. *Physiotherapy* 2015, **1**:1.
53. Malmberg JJ, Miilunpalo SI, Vuori IM, Pasanen ME, Oja P, Haapanen-Niemi NA: **A health-related fitness and functional performance test battery for middle-aged and older adults: feasibility and health-related content validity**. *Archives of Physical Medicine & Rehabilitation* 2002, **83**(5):666-677.
54. Rogers CE, Nseir S, Keller C: **Sign Chi Do and physical function: a pilot study**. *Geriatric Nursing* 2013, **34**(1):12-18.
55. Inoue W, Ikezoe T, Tsuboyama T, Sato I, Malinowska KB, Kawaguchi T, Tabara Y, Nakayama T, Matsuda F, Ichihashi N: **Are there different factors affecting walking speed and gait cycle variability between men and women in community-dwelling older adults?** *Aging Clinical and Experimental Research* 2017, **29**(2):215-221.
56. Brown M, Holloszy JO: **Effects of walking, jogging and cycling on strength, flexibility, speed and balance in 60- to 72-year olds**. *Aging-Clinical & Experimental Research* 1993, **5**(6):427-434.
57. Chung PK, Zhao Y, Liu JD, Quach B: **Functional fitness norms for community-dwelling older adults in Hong Kong**. *Archives of Gerontology & Geriatrics* 2016, **65**:54-62.
58. Mickle KJ, Caputi P, Potter JM, Steele JR: **Efficacy of a progressive resistance exercise program to increase toe flexor strength in older people**. *Clinical Biomechanics* 2016, **40**:14-19.
59. Nakamura M, Hashizume H, Oka H, Okada M, Takakura R, Hisari A, Yoshida M, Utsunomiya H: **Physical performance measures associated with locomotive syndrome in middle-aged and older Japanese women**. *Journal of Geriatric Physical Therapy* 2015, **38**(4):202-207.
60. Demnitz N, Hogan DB, Dawes H, Johansen-Berg H, Ebmeier KP, Poulin MJ, Sexton CE: **Cognition and mobility show a global association in middle-and late-adulthood: Analyses from the Canadian longitudinal study on aging**. *Gait & Posture* 2018.
61. Urushihata T, Kinugasa T, Soma Y, Miyoshi H: **Aging effects on the structure underlying balance abilities tests**. *Journal of the Japanese Physical Therapy Association* 2010, **13**(1):1-8.
62. Heiland EG, Welmer AK, Wang R, Santoni G, Angleman S, Fratiglioni L, Qiu C: **Association of mobility limitations with incident disability among older adults: a population-based study**. *Age & Ageing* 2016, **28**:28.
63. Welmer AK, Kareholt I, Angleman S, Rydwik E, Fratiglioni L: **Can chronic multimorbidity explain the age-related differences in strength, speed and balance in older adults?** *Aging-Clinical & Experimental Research* 2012, **24**(5):480-489.

64. Heiland EG, Welmer A-K, Wang R, Santoni G, Angleman S, Fratiglioni L, Qiu C: **Association of mobility limitations with incident disability among older adults: a population-based study**. *Age and Ageing* 2016, **45**(6):812-819.
65. Hafström A: **Perceived and functional balance control is negatively affected by diminished touch and vibration sensitivity in relatively healthy older adults and elderly**. *Gerontology and Geriatric Medicine* 2018, **4**:2333721418775551.
66. Sasai H, Matsuo T, Numao S, Sakai T, Mochizuki M, Kuroda K, Okamoto M, Tanaka K: **Aotake: a modified stepping exercise as a useful means of improving lower-extremity functional fitness in older adults**. *Geriatrics & gerontology international* 2010, **10**(3):244-250.
67. Shigematsu R, Okura T, Sakai T, Rantanen T: **Square-stepping exercise versus strength and balance training for fall risk factors**. *Aging-Clinical & Experimental Research* 2008, **20**(1):19-24.
68. Hong Y, Li JX, Robinson PD: **Balance control, flexibility, and cardiorespiratory fitness among older Tai Chi practitioners**. *British Journal of Sports Medicine* 2000, **34**(1):29-34.
69. Moriya S, Notani K, Murata A, Inoue N, Miura H: **Analysis of moment structures for assessing relationships among perceived chewing ability, dentition status, muscle strength, and balance in community-dwelling older adults**. *Gerodontology* 2014, **31**(4):281-287.
70. Gschwind YJ, Kressig RW, Lacroix A, Muehlbauer T, Pfenninger B, Granacher U: **A best practice fall prevention exercise program to improve balance, strength / power, and psychosocial health in older adults: study protocol for a randomized controlled trial**. *BMC Geriatrics* 2013, **13**:105.
71. Daly RM, Ahlborg HG, Ringsberg K, Gardsell P, Sernbo I, Karlsson MK: **Association between changes in habitual physical activity and changes in bone density, muscle strength, and functional performance in elderly men and women**. *Journal of the American Geriatrics Society* 2008, **56**(12):2252-2260.
72. Xie YJ, Liu EY, Anson ER, Agrawal Y: **Age-related imbalance is associated with slower walking speed: an analysis from the National Health and nutrition examination survey**. *Journal of Geriatric Physical Therapy* 2017, **40**(4):183-189.
73. Bloem BR, Valkenburg VV, Slabbekoorn M, van Dijk JG: **The multiple tasks test. Strategies in Parkinson's disease**. *Experimental Brain Research* 2001, **137**(3-4):478-486.
74. Bulbulian R, Hargan ML: **The effect of activity history and current activity on static and dynamic postural balance in older adults**. *Physiology & Behavior* 2000, **70**(3-4):319-325.
75. Misic MM, Rosengren KS, Woods JA, Evans EM: **Muscle quality, aerobic fitness and fat mass predict lower-extremity physical function in community-dwelling older adults**. *Gerontology* 2007, **53**(5):260-266.
76. Gao KL, Hui-Chan CW, Tsang WW: **Golfers have better balance control and confidence than healthy controls**. *European Journal of Applied Physiology* 2011, **111**(11):2805-2812.
77. Lee S-C: **Relationship of visual dependence to age, balance, attention, and vertigo**. *Journal of Physical Therapy Science* 2017, **29**(8):1318-1322.
78. Gordt K, Mikolaizak AS, Nerz C, Barz C, Gerhardy T, Weber M, Becker C, Schwenk M: **German version of the Community Balance and Mobility Scale**. *Zeitschrift für Gerontologie und Geriatrie* 2018:1-9.
79. Weber M, Van Ancum J, Bergquist R, Taraldsen K, Gordt K, Mikolaizak AS, Nerz C, Pijnappels M, Jonkman NH, Maier AB: **Concurrent validity and reliability of the Community Balance and Mobility scale in young-older adults**. *BMC Geriatrics* 2018, **18**(1):156.

80. Frey DJ, Ortega JD, Wiseman C, Farley CT, Wright KP, Jr.: **Influence of zolpidem and sleep inertia on balance and cognition during nighttime awakening: a randomized placebo-controlled trial**. *Journal of the American Geriatrics Society* 2011, **59**(1):73-81.
81. Goudarzian M, Ghavi S, Shariat A, Shirvani H, Rahimi M: **Effects of whole body vibration training and mental training on mobility, neuromuscular performance, and muscle strength in older men**. *Journal of Exercise Rehabilitation* 2017, **13**(5):573.
82. Tarnopolsky M, Zimmer A, Paikin J, Safdar A, Aboud A, Pearce E, Roy B, Doherty T: **Creatine monohydrate and conjugated linoleic acid improve strength and body composition following resistance exercise in older adults**. *PLoS ONE [Electronic Resource]* 2007, **2**(10):e991.
83. da Silva Costa AA, dos Santos LO, Mauerberg-deCastro E, Moraes R: **Task difficulty has no effect on haptic anchoring during tandem walking in young and older adults**. *Neuroscience Letters* 2018, **666**:133-138.
84. Bird M, Hill KD, Ball M, Hetherington S, Williams AD: **The long-term benefits of a multi-component exercise intervention to balance and mobility in healthy older adults**. *Archives of Gerontology & Geriatrics* 2011, **52**(2):211-216.
85. Vrantsidis F, Hill K, Haralambous B, Renehan E, Ledgerwood K, Pinikahana J, Harper S, Penberthy M: **Living Longer Living StrongerTM: a community-delivered strength training program improving function and quality of life**. *Australasian Journal on Ageing* 2014, **33**(1):22-25.
86. Bird ML, Hill KD, Fell JW: **A randomized controlled study investigating static and dynamic balance in older adults after training with Pilates**. *Archives of Physical Medicine & Rehabilitation* 2012, **93**(1):43-49.
87. Gianoudis J, Bailey CA, Ebeling PR, Nowson CA, Sanders KM, Hill K, Daly RM: **Effects of a targeted multimodal exercise program incorporating high-speed power training on falls and fracture risk factors in older adults: a community-based randomized controlled trial**. *Journal of Bone & Mineral Research* 2014, **29**(1):182-191.
88. Justine M, Ruzali D, Hazidin E, Said A, Bukry SA, Manaf H: **Range of motion, muscle length, and balance performance in older adults with normal, pronated, and supinated feet**. *Journal of Physical Therapy Science* 2016, **28**(3):916-922.
89. Pirotta S, Kidgell DJ, Daly RM: **Effects of vitamin D supplementation on neuroplasticity in older adults: a double-blinded, placebo-controlled randomised trial**. *Osteoporosis International* 2015, **26**(1):131-140.
90. Robinson RL, Ng SS: **The timed 180° turn test for assessing people with hemiplegia from chronic stroke**. *BioMed Research International* 2018, **2018**.
91. Tsai Y-J, Yang Y-C, Lu F-H, Lee P-Y, Lee I-T, Lin S-I: **Functional balance and its determinants in older people with diabetes**. *PloS One* 2016, **11**(7):e0159339.
92. Galvao DA, Taaffe DR: **Resistance exercise dosage in older adults: single- versus multiset effects on physical performance and body composition**. *Journal of the American Geriatrics Society* 2005, **53**(12):2090-2097.
93. Henwood TR, Taaffe DR: **Short-term resistance training and the older adult: the effect of varied programmes for the enhancement of muscle strength and functional performance**. *Clinical Physiology & Functional Imaging* 2006, **26**(5):305-313.
94. Henwood TR, Taaffe DR: **Detraining and retraining in older adults following long-term muscle power or muscle strength specific training**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2008, **63**(7):751-758.
95. Ng SS, Fong SS, Chan CW, Fung FM, Pang PK, Tsang NN, Yeung CL, Kwong PW, Tse MM, Tam EW: **Floor transfer test for assessing people with chronic stroke**. *Journal of Rehabilitation Medicine* 2015, **47**(6):489-494.
96. Ward CL, Valentine RJ, Evans EM: **Greater effect of adiposity than physical activity or lean mass on physical function in community-dwelling older adults**. *Journal of Aging & Physical Activity* 2014, **22**(2):284-293.

97. Zhuang J, Huang L, Wu Y, Zhang Y: **The effectiveness of a combined exercise intervention on physical fitness factors related to falls in community-dwelling older adults**. *Clinical Interventions In Aging* 2014, **9**:131-140.
98. Teixeira CV, Gobbi S, Pereira JR, Ueno DT, Shigematsu R, Gobbi LT: **Effect of square-stepping exercise and basic exercises on functional fitness of older adults**. *Geriatrics & gerontology international* 2013, **13**(4):842-848.
99. Wood RH, Reyes R, Welsch MA, Favaloro-Sabatier J, Sabatier M, Matthew Lee C, Johnson LG, Hooper PF: **Concurrent cardiovascular and resistance training in healthy older adults**. *Medicine & Science in Sports & Exercise* 2001, **33**(10):1751-1758.
100. Kwok BC, Pua YH: **Effects of WiiActive exercises on fear of falling and functional outcomes in community-dwelling older adults: a randomised control trial**. *Age and Ageing* 2016, **45**(5):621-627.
101. Ng SS, Lau BK, Law GT, Wom CW, Liu T-W, Tam EW, Tse MM, Fong SS: **Sideways walk test: reliability and association with lower limb motor function after stroke**. *Journal of Rehabilitation Medicine* 2016, **48**(8):657-665.
102. Pozzi F, White DK, Snyder-Mackler L, Zeni JA: **Restoring physical function after knee replacement: a cross sectional comparison of progressive strengthening vs standard physical therapy**. *Physiotherapy Theory and Practice* 2018:1-12.
103. Fife E, Kostka J, Kroc Ł, Guligowska A, Pigłowska M, Sołtysik B, Kaufman-Szymczyk A, Fabianowska-Majewska K, Kostka T: **Relationship of muscle function to circulating myostatin, follistatin and GDF11 in older women and men**. *BMC Geriatrics* 2018, **18**(1):200.
104. Coetsee C, Terblanche E: **The effect of three different exercise training modalities on cognitive and physical function in a healthy older population**. *European Review of Aging and Physical Activity* 2017, **14**(1):13.
105. Gusi N, Hernandez-Mocholi MA, Olivares PR: **Changes in HRQoL after 12 months of exercise linked to primary care are associated with fitness effects in older adults**. *European Journal of Public Health* 2015, **25**(5):873-879.
106. Weiss A, Herman T, Plotnik M, Brozgol M, Maidan I, Giladi N, Gurevich T, Hausdorff JM: **Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson's disease?** *Medical Engineering & Physics* 2010, **32**(2):119-125.
107. Uritani D, Fukumoto T, Matsumoto D, Shima M: **The relationship between toe grip strength and dynamic balance or functional mobility among community-dwelling Japanese older adults: a cross-sectional study**. *Journal of Aging and Physical Activity* 2016, **24**(3):459-464.
108. Hoang OTT, Jullamate P, Piphatvanitcha N, Rosenberg E: **Factors related to fear of falling among community-dwelling older adults**. *Journal of Clinical Nursing* 2017, **26**(1-2):68-76.
109. Duclos C, Mieville C, Gagnon D, Leclerc C: **Dynamic stability requirements during gait and standing exergames on the wii fit system in the elderly**. *Journal of Neuroengineering & Rehabilitation* 2012, **9**:28.
110. Alonso AC, Ribeiro SM, Luna NMS, Peterson MD, Bocalini DS, Serra MM, Brech GC, Greve JMDA, Garcez-Leme LE: **Association between handgrip strength, balance, and knee flexion/extension strength in older adults**. *PloS One* 2018, **13**(6):e0198185.
111. Olivares PR, Gusi N, Prieto J, Hernandez-Mocholi MA: **Fitness and health-related quality of life dimensions in community-dwelling middle aged and older adults**. *Health & Quality of Life Outcomes* 2011, **9**:117.
112. Ng SS, Fong SS, Chan WL, Hung BK, Chung RK, Chim TH, Kwong PW, Liu T-W, Tse MM, Chung RC: **The sitting and rising test for assessing people with chronic stroke**. *Journal of Physical Therapy Science* 2016, **28**(6):1701-1708.
113. Campos de Oliveira L, Goncalves de Oliveira R, Pires-Oliveira DA: **Effects of Pilates on muscle strength, postural balance and quality of life of older adults: a randomized, controlled, clinical trial**. *Journal of Physical Therapy Science* 2015, **27**(3):871-876.

114. Beauchet O, Launay CP, Sejdic E, Allali G, Annweiler C: **Motor imagery of gait: a new way to detect mild cognitive impairment?** *Journal of Neuroengineering & Rehabilitation* 2014, **11**:66.
115. Coetsee C, Terblanche E: **The time course of changes induced by resistance training and detraining on muscular and physical function in older adults**. *European Reviews of Aging & Physical Activity* 2015, **12**:7.
116. Simpson LA, Miller WC, Eng JJ: **Effect of stroke on fall rate, location and predictors: a prospective comparison of older adults with and without stroke**. *PLoS ONE [Electronic Resource]* 2011, **6**(4):e19431.
117. Yu W, An C, Kang H: **Effects of resistance exercise using thera-band on balance of elderly adults: A randomized controlled trial**. *Journal of Physical Therapy Science* 2013, **25**(11):1471-1473.
118. Van Roie E, Delecluse C, Coudyzer W, Boonen S, Bautmans I: **Strength training at high versus low external resistance in older adults: effects on muscle volume, muscle strength, and force-velocity characteristics**. *Experimental Gerontology* 2013, **48**(11):1351-1361.
119. Lamarche L, Gammage KL, Klentrou P, Adkin AL: **What will they think? The relationship between self-presentational concerns and balance and mobility outcomes in older women**. *Experimental Aging Research* 2014, **40**(4):426-435.
120. Wang X, Ma Y, Wang J, Han P, Dong R, Kang L, Zhang W, Shen S, Wang J, Li D *et al*: **Mobility and muscle strength together are more strongly correlated with falls in suburb-dwelling older Chinese**. *Scientific Reports* 2016, **6**:25420.
121. Wu IC, Chang HY, Hsu CC, Chiu YF, Yu SH, Tsai YF, Shen SC, Kuo KN, Chen CY, Liu K *et al*: **Association between dietary fiber intake and physical performance in older adults: a nationwide study in Taiwan**. *PLoS ONE [Electronic Resource]* 2013, **8**(11):e80209.
122. Tsuji T, Kitano N, Tsunoda K, Himori E, Okura T, Tanaka K: **Short-term effects of whole-body vibration on functional mobility and flexibility in healthy, older adults: a randomized crossover study**. *Journal of Geriatric Physical Therapy* 2014, **37**(2):58-64.
123. Uemura SI, Kanbayashi T, Wakasa M, Satake M, Ito W, Shimizu K, Shioya T, Shimizu T, Nishino S: **Residual effects of zolpidem, triazolam, rilmazafone and placebo in healthy elderly subjects: a randomized double-blind study**. *Sleep Medicine* 2015, **16**(11):1395-1402.
124. Siervo M, Oggioni C, Jakovljevic DG, Trenell M, Mathers JC, Houghton D, Celis-Morales C, Ashor AW, Ruddock A, Ranchordas M: **Dietary nitrate does not affect physical activity or outcomes in healthy older adults in a randomized, cross-over trial**. *Nutrition Research* 2016, **36**(12):1361-1369.
125. Furness TP, Maschette WE, Lorenzen C, Naughton GA, Williams MD: **Efficacy of a whole-body vibration intervention on functional performance of community-dwelling older adults**. *Journal of Alternative & Complementary Medicine* 2010, **16**(7):795-797.
126. Houck J, Kneiss J, Bukata SV, Puzas JE: **Analysis of vertical ground reaction force variables during a Sit to Stand task in participants recovering from a hip fracture**. *Clinical Biomechanics* 2011, **26**(5):470-476.
127. Gault ML, Clements RE, Willems ME: **Functional mobility of older adults after concentric and eccentric endurance exercise**. *European Journal of Applied Physiology* 2012, **112**(11):3699-3707.
128. Meng Y, Wu H, Yang Y, Du H, Xia Y, Guo X, Liu X, Li C, Niu K: **Relationship of anabolic and catabolic biomarkers with muscle strength and physical performance in older adults: a population-based cross-sectional study**. *BMC Musculoskeletal Disorders* 2015, **16**:202.
129. Kwok BC, Clark RA, Pua YH: **Novel use of the Wii Balance Board to prospectively predict falls in community-dwelling older adults**. *Clinical Biomechanics* 2015, **30**(5):481-484.
130. Nagamatsu LS, Weinstein AM, Erickson KI, Fanning J, Awick EA, Kramer AF, McAuley E: **Exercise mode moderates the relationship between mobility and basal ganglia volume in healthy older adults**. *Journal of the American Geriatrics Society* 2016, **64**(1):102-108.

131. Rogers HL, Cromwell RL, Grady JL: **Adaptive changes in gait of older and younger adults as responses to challenges to dynamic balance**. *Journal of Aging & Physical Activity* 2008, **16**(1):85-96.
132. Mulford D, Taggart HM, Nivens A, Payrie C: **Arch support use for improving balance and reducing pain in older adults**. *Applied Nursing Research* 2008, **21**(3):153-158.
133. Canney M, Sexton DJ, O'connell MD, Kenny RA, Little MA, O'seaghdha CM: **Kidney function estimated from cystatin C, but not creatinine, is related to objective tests of physical performance in community-dwelling older adults**. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 2017, **72**(11):1554-1560.
134. Araujo TB, Silva NA, Costa JN, Pereira MM, Safons MP: **Effect of equine-assisted therapy on the postural balance of the elderly**. *Revista Brasileira de Fisioterapia* 2011, **15**(5):414-419.
135. Valentine RJ, Misic MM, Rosengren KS, Woods JA, Evans EM: **Sex impacts the relation between body composition and physical function in older adults**. *Menopause* 2009, **16**(3):518-523.
136. Dias CP, Toscan R, de Camargo M, Pereira EP, Griebler N, Baroni BM, Tiggemann CL: **Effects of eccentric-focused and conventional resistance training on strength and functional capacity of older adults**. *Age* 2015, **37**(5):99.
137. Morat T, Mechling H: **The functional movement circle for older adults: feasibility and effects on physical performance**. *Aging-Clinical & Experimental Research* 2014, **26**(5):529-537.
138. Vervoort D, Vuillerme N, Kosse N, Hortobágyi T, Lamoth CJ: **Multivariate analyses and classification of inertial sensor data to identify aging effects on the Timed-Up-and-Go test**. *PloS One* 2016, **11**(6):e0155984.
139. Merchant RA, Banerji S, Singh G, Chew E, Poh CL, Tapawan SC, Guo YR, Pang YW, Sharma M, Kambadur R *et al*: **Is trunk posture in walking a better marker than gait speed in predicting decline in function and subsequent frailty?** *Journal of the American Medical Directors Association* 2016, **17**(1):65-70.
140. Voos MC, Custodio EB, Malaquias J, Jr.: **Relationship of executive function and educational status with functional balance in older adults**. *Journal of Geriatric Physical Therapy* 2011, **34**(1):11-18.
141. Buskard A, Zalma B, Cherup N, Armitage C, Dent C, Signorile JF: **Effects of linear periodization versus daily undulating periodization on neuromuscular performance and activities of daily living in an elderly population**. *Experimental Gerontology* 2018, **113**:199-208.
142. Foong YC, Chherawala N, Aitken D, Scott D, Winzenberg T, Jones G: **Accelerometer-determined physical activity, muscle mass, and leg strength in community-dwelling older adults**. *Journal of Cachexia, Sarcopenia and Muscle* 2016, **7**(3):275-283.
143. Raj IS, Bird SR, Westfold BA, Shield AJ: **Determining criteria to predict repeatability of performance in older adults: Using coefficients of variation for strength and functional measures**. *Journal of Aging and Physical Activity* 2017, **25**(1):94-98.
144. Balachandran AT, Gandia K, Jacobs KA, Streiner DL, Eltoukhy M, Signorile JF: **Power training using pneumatic machines vs. plate-loaded machines to improve muscle power in older adults**. *Experimental Gerontology* 2017, **98**:134-142.
145. Van Roie E, Walker S, Van Driessche S, Baggen R, Coudyzer W, Bautmans I, Delecluse C: **Training load does not affect detraining's effect on muscle volume, muscle strength and functional capacity among older adults**. *Experimental Gerontology* 2017, **98**:30-37.
146. Bellumori M, Uygur M, Knight CA: **High-speed cycling intervention improves rate-dependent mobility in older adults**. *Medicine and Science in Sports and Exercise* 2017, **49**(1):106.

147. Eckardt N: **Lower-extremity resistance training on unstable surfaces improves proxies of muscle strength, power and balance in healthy older adults: a randomised control trial**. *BMC Geriatrics* 2016, **16**(1):191.
148. Noopud P, Suputtitada A, Khongprasert S, Kanungsukkasem V: **Effects of thai traditional dance on balance performance in daily life among older women**. *Aging Clinical and Experimental Research* 2018:1-7.
149. Ng SS, Tse MM, Tam EW, Lai CY: **The psychometric properties of the toe tap test in people with stroke**. *Disability and Rehabilitation* 2018:1-9.
150. Merchant RA, van Dam R, Tan L, Lim M, Low J, Morley J: **Vitamin D binding protein and vitamin D levels in multi-ethnic population**. *The Journal of Nutrition, Health & Aging* 2018:1-6.
151. Afridi A, Malik AN, Ali S, Amjad I: **Effect of balance training in older adults using Wii fit plus**. *JPMA The Journal of the Pakistan Medical Association* 2018, **68**(3):480-483.
152. Lou L, Zou L, Fang Q, Wang H, Liu Y, Tian Z, Han Y: **Effect of Taichi softball on function-related outcomes in older adults: A randomized control trial**. *Evidence-Based Complementary and Alternative Medicine* 2017, **2017**.
153. Shahtahmassebi B, Hebert JJ, Hecimovich MD, Fairchild TJ: **Associations between trunk muscle morphology, strength and function in older adults**. *Scientific Reports* 2017, **7**(1):10907.
154. Bates A, Donaldson A, Lloyd B, Castell S, Krolik P, Coleman R: **Staying active, staying strong: pilot evaluation of a once-weekly, community-based strength training program for older adults**. *Health Promotion Journal of Australia* 2009, **20**(1):42-47.
155. Lee M, Son J, Kim J, Yoon B: **Individualized feedback-based virtual reality exercise improves older women's self-perceived health: a randomized controlled trial**. *Archives of Gerontology & Geriatrics* 2015, **61**(2):154-160.
156. Reichert T, Kanitz AC, Delevatti RS, Bagatini NC, Barroso BM, Kruel LF: **Continuous and interval training programs using deep water running improves functional fitness and blood pressure in the older adults**. *Age* 2016, **38**(1):20.
157. Sousa N, Mendes R, Abrantes C, Sampaio J, Oliveira J: **Effectiveness of combined exercise training to improve functional fitness in older adults: A randomized controlled trial**. *Geriatrics & gerontology international* 2014, **14**(4):892-898.
158. Zettel-Watson L, Suen M, Wehbe L, Rutledge DN, Cherry BJ: **Aging well: Processing speed inhibition and working memory related to balance and aerobic endurance**. *Geriatrics & gerontology international* 2015, **23**:23.
159. Zhao Q, Wang J, Feng W, Jia W, Burke LE, Zgibor JC, Sun M: **Assessing physical performance in free-living older adults with a wearable computer**. *Proceedings of the IEEE Annual Northeast Bioengineering Conference* 2015.
160. Mason RC, Horvat M, Nocera J: **The effects of exercise on the physical fitness of high and moderate-low functioning older adult women**. *Journal of Aging Research* 2016, **2016**.
161. Ramalho F, Santos-Rocha R, Branco M, Moniz-Pereira V, André H-I, Veloso AP, Carnide F: **Effect of 6-month community-based exercise interventions on gait and functional fitness of an older population: a quasi-experimental study**. *Clinical Interventions in Aging* 2018, **13**:595.
162. Lee M, Lim T, Lee J, Kim K, Yoon B: **Optimal retraining time for regaining functional fitness using multicomponent training after long-term detraining in older adults**. *Archives of Gerontology and Geriatrics* 2017, **73**:227-233.
163. Noradechanunt C, Worsley A, Groeller H: **Thai Yoga improves physical function and well-being in older adults: A randomised controlled trial**. *Journal of Science and Medicine in Sport* 2017, **20**(5):494-501.

164. Zettel-Watson L, Suen M, Wehbe L, Rutledge DN, Cherry BJ: **Aging well: Processing speed inhibition and working memory related to balance and aerobic endurance**. *Geriatrics & Gerontology International* 2017, **17**(1):108-115.
165. Richardson DL, Duncan MJ, Jimenez A, Juris PM, Clarke ND: **Effects of movement velocity and training frequency of resistance exercise on functional performance in older adults: a randomised controlled trial**. *European Journal of Sport Science* 2018:1-13.
166. Prusik K, Kortas J, Prusik K, Mieszkowski J, Jaworska J, Skrobot W, Lipinski M, Ziemann E, Antosiewicz J: **Nordic walking training causes a decrease in blood cholesterol in elderly women supplemented with vitamin D**. *Frontiers in Endocrinology* 2018, **9**:42.
167. Bergamin M, Gobbo S, Bullo V, Zanotto T, Vendramin B, Duregon F, Cugusi L, Camozzi V, Zaccaria M, Neunhaeuserer D *et al*: **Effects of a Pilates exercise program on muscle strength, postural control and body composition: results from a pilot study in a group of post-menopausal women**. *Age* 2015, **37**(6):118.
168. Canuto Wanderley FA, Oliveira NL, Marques E, Moreira P, Oliveira J, Carvalho J: **Aerobic versus resistance training effects on health-related quality of life, body composition, and function of older adults**. *Journal of Applied Gerontology* 2015, **34**(3):NP143-165.
169. Carvalho MJ, Marques E, Mota J: **Training and detraining effects on functional fitness after a multicomponent training in older women**. *Gerontology* 2009, **55**(1):41-48.
170. Krause MP, Januario RS, Hallage T, Haile L, Miculis CP, Gama MP, Goss FL, da Silva SG: **A comparison of functional fitness of older Brazilian and American women**. *Journal of Aging & Physical Activity* 2009, **17**(4):387-397.
171. Toraman F, Sahin G: **Age responses to multicomponent training programme in older adults**. *Disability & Rehabilitation* 2004, **26**(8):448-454.
172. Marques EA, Mota J, Viana JL, Tuna D, Figueiredo P, Guimaraes JT, Carvalho J: **Response of bone mineral density, inflammatory cytokines, and biochemical bone markers to a 32-week combined loading exercise programme in older men and women**. *Archives of Gerontology & Geriatrics* 2013, **57**(2):226-233.
173. Milanovic Z, Pantelic S, Trajkovic N, Sporis G, Kostic R, James N: **Age-related decrease in physical activity and functional fitness among elderly men and women.[Erratum appears in Clin Interv Aging. clin interv aging. 2014;9:979]**. *Clinical Interventions In Aging* 2013, **8**:549-556.
174. Yamauchi T, Islam MM, Koizumi D, Rogers ME, Rogers NL, Takeshima N: **Effect of home-based well-rounded exercise in community-dwelling older adults**. *Journal of Sports Science & Medicine* 2005, **4**(4):563-571.
175. Palmer RC, Batra A, Anderson C, Page T, Vieira E, Seff L: **Implementation of an evidence-based exercise program for older adults in South Florida**. *Journal of Aging Research* 2016, **2016**.
176. Timmons JF, Minnock D, Hone M, Cogan KE, Murphy JC, Egan B: **Comparison of time-matched aerobic, resistance, or concurrent exercise training in older adults**. *Scandinavian Journal of Medicine & Science in Sports* 2018, **28**(11):2272-2283.
177. Yagci N, Cavlak U, Aslan UB, Akdag B: **Relationship between balance performance and musculoskeletal pain in lower body comparison healthy middle aged and older adults**. *Archives of Gerontology & Geriatrics* 2007, **45**(1):109-119.
178. Lin SI, Tsai TT: **Muscle weakness and imbalance in older dizzy patients**. *Aging-Clinical & Experimental Research* 2005, **17**(3):168-173.
179. Nakamura Y, Tanaka K, Yabushita N, Sakai T, Shigematsu R: **Effects of exercise frequency on functional fitness in older adult women**. *Archives of Gerontology & Geriatrics* 2007, **44**(2):163-173.

180. Singh DK, Pillai SG, Tan ST, Tai CC, Shahar S: **Association between physiological falls risk and physical performance tests among community-dwelling older adults**. *Clinical Interventions In Aging* 2015, **10**:1319-1326.
181. !!! INVALID CITATION !!! [13, 70].
182. Abe T, Patterson KM, Stover CD, Geddam DA, Tribby AC, Lajza DG, Young KC: **Site-specific thigh muscle loss as an independent phenomenon for age-related muscle loss in middle-aged and older men and women**. *Age* 2014, **36**(3):9634.
183. Salkovic D, Hobert MA, Bellut C, Funer F, Renno S, Haertner L, Hasmann SE, Staebler J, Geritz J, Suenkel U: **Evidence for a selectively regulated prioritization shift depending on walking situations in older adults**. *Frontiers in Aging Neuroscience* 2017, **9**:75.
184. Schrager MA, Hilton J, Gould R, Kelly VE: **Effects of blueberry supplementation on measures of functional mobility in older adults**. *Applied Physiology, Nutrition, & Metabolism = Physiologie Appliquee, Nutrition et Metabolisme* 2015, **40**(6):543-549.
185. Elbar O, Tzedek I, Vered E, Shvarth G, Friger M, Melzer I: **A water-based training program that includes perturbation exercises improves speed of voluntary stepping in older adults: a randomized controlled cross-over trial**. *Archives of Gerontology & Geriatrics* 2013, **56**(1):134-140.
186. Kim J, Son J, Ko N, Yoon B: **Unsupervised virtual reality-based exercise program improves hip muscle strength and balance control in older adults: a pilot study**. *Archives of Physical Medicine & Rehabilitation* 2013, **94**(5):937-943.
187. Garland SJ, Stevenson TJ, Ivanova T: **Postural responses to unilateral arm perturbation in young, elderly, and hemiplegic subjects**. *Archives of Physical Medicine & Rehabilitation* 1997, **78**(10):1072-1077.
188. Lip RW, Fong SS, Ng SS, Liu KP, Guo X: **Effects of Ving Tsun Chinese martial art training on musculoskeletal health, balance performance, and self-efficacy in community-dwelling older adults**. *Journal of Physical Therapy Science* 2015, **27**(3):667-672.
189. Misic MM, Valentine RJ, Rosengren KS, Woods JA, Evans EM: **Impact of training modality on strength and physical function in older adults**. *Gerontology* 2009, **55**(4):411-416.
190. Nematollahi A, Kamali F, Ghanbari A, Etminan Z, Sobhani S: **Improving balance in older people: A double-blind randomized clinical trial of three modes of balance training**. *Journal of Aging & Physical Activity* 2015, **23**:23.
191. Stanley J, Hollands M: **A novel video-based paradigm to study the mechanisms underlying age- and falls risk-related differences in gaze behaviour during walking**. *Ophthalmic & Physiological Optics* 2014, **34**(4):459-469.
192. Stellefson M, Yannessa JF, Martel GF: **Using canonical commonality analysis to examine the predictive quality of aging and falls efficacy on balance functioning in older adults**. *Evaluation & the Health Professions* 2012, **35**(2):239-255.
193. Li Z, Wang X-X, Liang Y-Y, Chen S-Y, Sheng J, Ma S-J: **Effects of the visual-feedback-based force platform training with functional electric stimulation on the balance and prevention of falls in older adults: a randomized controlled trial**. *PeerJ* 2018, **6**:e4244.
194. Barone Gibbs B, Brach JS, Byard T, Creasy S, Davis KK, McCoy S, Peluso A, Rogers RJ, Rupp K, Jakicic JM: **Reducing sedentary behavior versus increasing moderate-to-vigorous intensity physical activity in older adults: A 12-week randomized, clinical trial**. *Journal of Aging & Health* 2016, **3**:3.
195. Beavers KM, Gordon MM, Easter L, Beavers DP, Hairston KG, Nicklas BJ, Vitolins MZ: **Effect of protein source during weight loss on body composition, cardiometabolic risk and physical performance in abdominally obese, older adults: a pilot feeding study**. *Journal of Nutrition, Health & Aging* 2015, **19**(1):87-95.

196. da Camara SM, Alvarado BE, Guralnik JM, Guerra RO, Maciel AC: **Using the Short Physical Performance Battery to screen for frailty in young-old adults with distinct socioeconomic conditions**. *Geriatrics & gerontology international* 2013, **13**(2):421-428.
197. Gray SL, Aragaki AK, LaMonte MJ, Cochrane BB, Kooperberg C, Robinson JG, Woods NF, LaCroix AZ: **Statins, angiotensin-converting enzyme inhibitors, and physical performance in older women**. *Journal of the American Geriatrics Society* 2012, **60**(12):2206-2214.
198. Hannam K, Deere K, Worrall S, Hartley A, Tobias JH: **Characterization of vertical accelerations experienced by older people attending an aerobics class designed to produce high impacts**. *Journal of Aging & Physical Activity* 2016, **24**(2):268-274.
199. Hau C, Reid KF, Wong KF, Chin RJ, Botto TJ, Eliasziw M, Bermudez OI, Fielding RA: **Collaborative evaluation of the healthy habits program: an effective community intervention to improve mobility and cognition of chinese older adults living in the U.S**. *Journal of Nutrition, Health & Aging* 2016, **20**(4):391-397.
200. Kortebein P, Symons TB, Ferrando A, Paddon-Jones D, Ronsen O, Protas E, Conger S, Lombeida J, Wolfe R, Evans WJ: **Functional impact of 10 days of bed rest in healthy older adults**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2008, **63**(10):1076-1081.
201. Martins WR, Safons MP, Bottaro M, Blasczyk JC, Diniz LR, Fonseca RM, Bonini-Rocha AC, de Oliveira RJ: **Effects of short term elastic resistance training on muscle mass and strength in untrained older adults: a randomized clinical trial**. *BMC Geriatrics* 2015, **15**:99.
202. Moore-Harrison TL, Speer EM, Johnson FT, Cress ME: **The effects of aerobic training and nutrition education on functional performance in low socioeconomic older adults**. *Journal of Geriatric Physical Therapy* 2008, **31**(1):18-23.
203. Nicklas BJ, Chmelo E, Delbono O, Carr JJ, Lyles MF, Marsh AP: **Effects of resistance training with and without caloric restriction on physical function and mobility in overweight and obese older adults: a randomized controlled trial**. *American Journal of Clinical Nutrition* 2015, **101**(5):991-999.
204. Reeves GR, Whellan DJ, Patel MJ, O'Connor CM, Duncan P, Eggebeen JD, Morgan TM, Hewston LA, Pastva AM, Kitzman DW: **Comparison of frequency of frailty and severely impaired physical function in patients >60 years hospitalized with acute decompensated heart failure versus chronic stable heart failure with reduced and preserved left ventricular ejection fraction**. *American Journal of Cardiology* 2016, **6**:6.
205. Roma MF, Busse AL, Betoni RA, Melo AC, Kong J, Santarem JM, Jacob Filho W: **Effects of resistance training and aerobic exercise in elderly people concerning physical fitness and ability: a prospective clinical trial**. *Einstein* 2013, **11**(2):153-157.
206. Satariano WA, Kealey M, Hubbard A, Kurtovich E, Ivey SL, Bayles CM, Hunter RH, Prohaska TR: **Mobility disability in older adults: At the intersection of people and places**. *Gerontologist* 2016, **56**(3):525-534.
207. Stehle JR, Jr., Leng X, Kitzman DW, Nicklas BJ, Kritchevsky SB, High KP: **Lipopolysaccharide-binding protein, a surrogate marker of microbial translocation, is associated with physical function in healthy older adults**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2012, **67**(11):1212-1218.
208. Gomez JF, Curcio CL, Alvarado B, Zunzunegui MV, Guralnik J: **Validity and reliability of the Short Physical Performance Battery (SPPB): a pilot study on mobility in the Colombian Andes**. *Colombia Medica* 2013, **44**(3):165-171.
209. Lana A, Struijk E, Guallar-Castillón P, Martín-Moreno JM, Rodríguez Artalejo F, Lopez-Garcia E: **Leptin concentration and risk of impaired physical function in older adults: the Seniors-ENRICA cohort**. *Age and Ageing* 2016, **45**(6):819-826.
210. Wilson C, Perkin OJ, McGuigan MP, Stokes KA: **The effect of age on technique variability and outcome variability during a leg press**. *PloS One* 2016, **11**(10):e0163764.

211. Ahmed T, Vafaei A, Auais M, Guralnik J, Zunzunegui MV: **Gender roles and physical function in older adults: Cross-sectional analysis of the International Mobility in Aging Study (IMIAS)**. *PloS One* 2016, **11**(6):e0156828.
212. Bollaert RE, Marsh AP, Cutter GR, Motl RW: **The virtual Short Physical Performance Battery: psychometric properties and validation in older adults with multiple sclerosis**. *Journal of Applied Gerontology* 2017:0733464817709532.
213. Holwerda AM, Overkamp M, Paulussen KJ, Smeets JS, van Kranenburg J, Backx EM, Gijsen AP, Goessens JP, Verdijk LB, van Loon LJ: **Protein supplementation after exercise and before sleep does not further augment muscle mass and strength gains during resistance exercise training in active older men**. *The Journal of Nutrition* 2018, **148**(11):1723-1732.
214. Auais M, French S, Alvarado B, Pirkle C, Belanger E, Guralnik J: **Fear of falling predicts incidence of functional disability 2 years later: A perspective from an international cohort study**. *The Journals of Gerontology: Series A* 2017, **73**(9):1212-1215.
215. Osawa Y, Studenski SA, Ferrucci L: **Knee extension rate of torque development and peak torque: associations with lower extremity function**. *Journal of Cachexia, Sarcopenia and Muscle* 2018.
216. Distefano G, Standley RA, Zhang X, Carnero EA, Yi F, Cornnell HH, Coen PM: **Physical activity unveils the relationship between mitochondrial energetics, muscle quality, and physical function in older adults**. *Journal of Cachexia, Sarcopenia and Muscle* 2018, **9**(2):279-294.
217. Cegielski J, Brook MS, Quinlan JI, Wilkinson DJ, Smith K, Atherton PJ, Phillips BE: **A 4-week, lifestyle-integrated, home-based exercise training programme elicits improvements in physical function and lean mass in older men and women: a pilot study**. *F1000Research* 2017, **6**.
218. Ćwirlej-Sozańska A, Wiśniowska-Szurlej A, Wilmowska-Pietruszyńska A, Sozański B, Wołoszyn N: **Assessment of psychophysical capacities for professional work in late middle age and at the beginning of old age**. *Medycyna Pracy* 2018, **69**(4):375-381.
219. Manckoundia P, Thomas F, Buatois S, Guize L, Jego B, Aquino JP, Benetos A: **Impact of clinical, psychological, and social factors on decreased Tinetti test score in community-living elderly subjects: a prospective study with two-year follow-up**. *Medical Science Monitor* 2008, **14**(6):CR316-322.
220. Aung KC, Feng L, Yap KB, Sitoh YY, Leong IY, Ng TP: **Serum albumin and hemoglobin are associated with physical function in community-living older persons in Singapore**. *Journal of Nutrition, Health & Aging* 2011, **15**(10):877-882.
221. Ng TP, Aung KC, Feng L, Feng L, Nyunt MS, Yap KB: **Tea consumption and physical function in older adults: a cross-sectional study**. *Journal of Nutrition, Health & Aging* 2014, **18**(2):161-166.
222. Ng TP, Aung KC, Feng L, Scherer SC, Yap KB: **Homocysteine, folate, vitamin B-12, and physical function in older adults: cross-sectional findings from the Singapore Longitudinal Ageing Study**. *American Journal of Clinical Nutrition* 2012, **96**(6):1362-1368.
223. Sobhani S, Sinaei E, Motealleh A, Hooshyar F, Kashkooli NS, Yoosefinejad AK: **Combined effects of whole body vibration and unstable shoes on balance measures in older adults: A randomized clinical trial**. *Archives of Gerontology and Geriatrics* 2018, **78**:30-37.
224. Schrack JA, Kuo P-L, Wanigatunga AA, Di J, Simonsick EM, Spira AP, Ferrucci L, Zipunnikov V: **Active-to-sedentary behavior transitions, fatigability, and physical functioning in older adults**. *The Journals of Gerontology: Series A* 2018.
225. Martinez-Amezcua P, Matsushita K, Simonsick EM, Ferrucci L, Schrack JA: **Fatigability and functional performance among older adults with low-normal ankle-brachial index: Cross-sectional findings from the Baltimore Longitudinal Study of Aging**. *Atherosclerosis* 2018, **272**:200-206.

226. Zhang W, Schwenk M, Mellone S, Paraschiv-Ionescu A, Vereijken B, Pijnappels M, Mikolaizak A, Boulton E, Jonkman N, Maier A: **Complexity of daily physical activity is more sensitive than conventional metrics to assess functional change in younger older adults**. *Sensors* 2018, **18**(7):2032.
227. Wang H, Wei A, Lu Y, Yu B, Chen W, Lu Y, Liu Y, Yu D, Zou L: **Simplified tai chi program training versus traditional tai chi on the functional movement screening in older adults**. *Evidence-based Complementary and Alternative Medicine* 2016, **2016**.
228. Coelho VA, Probst VS, Nogari BM, Teixeira DC, Felcar JM, Santos DC, Gomes MV, Andraus RA, Fernandes KB: **Angiotensin-II blockage, muscle strength, and exercise capacity in physically independent older adults**. *Journal of Physical Therapy Science* 2016, **28**(2):547-552.
229. Yaginuma Y, Abe T, Thiebaud RS, Kitamura T, Kawanishi M, Fukunaga T: **Can handgrip strength improve following body mass-based lower body exercise?** *BioResearch Open Access* 2017, **6**(1):19-27.
230. Scott D, Shore-Lorenti C, McMillan L, Mesinovic J, Clark RA, Hayes A, Sanders KM, Duque G, Ebeling PR: **Associations of components of sarcopenic obesity with bone health and balance in older adults**. *Archives of Gerontology and Geriatrics* 2018, **75**:125-131.
231. de Souza Barbosa JF, Zepeda MU, Beland F, Guralnik JM, Zunzunegui MV, Guerra RO: **Clinically relevant weakness in diverse populations of older adults participating in the International Mobility in Aging Study**. *Age* 2016, **38**(1):25.
232. Francis P, Mc Cormack W, Toomey C, Norton C, Saunders J, Kerin E, Lyons M, Jakeman P: **Twelve weeks' progressive resistance training combined with protein supplementation beyond habitual intakes increases upper leg lean tissue mass, muscle strength and extended gait speed in healthy older women**. *Biogerontology* 2017, **18**(6):881-891.
233. Garcia IFF, Tiuganji CT, Simões MdSMP, Santoro IL, Lunardi AC: **Systemic effects of chronic obstructive pulmonary disease in young-old adults' life-space mobility**. *International Journal of Chronic Obstructive Pulmonary Disease* 2017, **12**:2777.
234. Muller AM, Khoo S, Morris T: **Text messaging for exercise promotion in older adults from an upper-middle-income country: randomized controlled trial**. *Journal of Medical Internet Research* 2016, **18**(1):e5.
235. Hwang HJ, Kim SH: **The association among three aspects of physical fitness and metabolic syndrome in a Korean elderly population**. *Diabetology & metabolic syndrome* 2015, **7**:112.
236. Dip RM, Cabrera MA, Prato SF: **Association between body composition and stair negotiation ability among individuals> 55 years of age: a cross-sectional study**. *Clinical Interventions in Aging* 2017, **12**:1289.
237. Wanderley FA, Silva G, Marques E, Oliveira J, Mota J, Carvalho J: **Associations between objectively assessed physical activity levels and fitness and self-reported health-related quality of life in community-dwelling older adults**. *Quality of Life Research* 2011, **20**(9):1371-1378.
238. Gonzales JU, Defferari E, Fisher A, Shephard J, Proctor DN: **Calf exercise-induced vasodilation is blunted in healthy older adults with increased walking performance fatigue**. *Experimental Gerontology* 2014, **57**:1-5.
239. Aribisala BS, Gow AJ, Bastin ME, del Carmen Valdes Hernandez M, Murray C, Royle NA, Munoz Maniega S, Starr JM, Deary IJ, Wardlaw JM: **Associations between level and change in physical function and brain volumes**. *PLoS ONE [Electronic Resource]* 2013, **8**(11):e80386.
240. Zeng P, Han Y, Pang J, Wu S, Gong H, Zhu J, Li J, Zhang T: **Sarcopenia-related features and factors associated with lower muscle strength and physical performance in older Chinese: a cross sectional study**. *BMC Geriatrics* 2016, **16**:45.
241. Woo J, Leung J, Kwok T: **BMI, body composition, and physical functioning in older adults**. *Obesity* 2007, **15**(7):1886-1894.
242. Ismail N, Hairi F, Choo WY, Hairi NN, Peramalah D, Bulgiba A: **The Physical Activity Scale for the Elderly (PASE): Validity and reliability among community-dwelling older adults in Malaysia**. *Asia-Pacific Journal of Public Health* 2015, **27**(8 Suppl):62S-72S.

243. Antony B, Jones G, Stannus O, Blizzard L, Ding C: **Body fat predicts an increase and limb muscle strength predicts a decrease in leptin in older adults over 2.6 years**. *Clinical Endocrinology* 2013, **79**(5):652-660.
244. Hashemi R, Shafiee G, Motlagh AD, Pasalar P, Esmailzadeh A, Siassi F, Larijani B, Heshmat R: **Sarcopenia and its associated factors in Iranian older individuals: Results of SARIR study**. *Archives of Gerontology & Geriatrics* 2016, **66**:18-22.
245. Scott D, Hayes A, Sanders KM, Aitken D, Ebeling PR, Jones G: **Operational definitions of sarcopenia and their associations with 5-year changes in falls risk in community-dwelling middle-aged and older adults**. *Osteoporosis International* 2014, **25**(1):187-193.
246. Goldman N, Glei DA, Rosero-Bixby L, Chiou ST, Weinstein M: **Performance-based measures of physical function as mortality predictors: Incremental value beyond self-reports**. *Demographic Research [electronic resource]* 2014, **30**(7):227-252.
247. Leite JC, Forte R, de Vito G, Boreham CA, Gibney MJ, Brennan L, Gibney ER: **Comparison of the effect of multicomponent and resistance training programs on metabolic health parameters in the elderly**. *Archives of Gerontology & Geriatrics* 2015, **60**(3):412-417.
248. Hong GR, Cho SH, Tak Y: **Falls among Koreans 45 years of age and older: incidence and risk factors**. *Journal of Advanced Nursing* 2010, **66**(9):2014-2024.
249. Hamer M, Molloy GJ: **Association of C-reactive protein and muscle strength in the English Longitudinal Study of Ageing**. *Age* 2009, **31**(3):171-177.
250. Wong CH, Wong SF, Pang WS, Azizah MY, Dass MJ: **Habitual walking and its correlation to better physical function: implications for prevention of physical disability in older persons**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2003, **58**(6):555-560.
251. Wong CH, Wong SF, Yusoff AM, Karunananthan S, Bergman H: **The effect of later-life health promotion on functional performance and body composition**. *Aging-Clinical & Experimental Research* 2008, **20**(5):454-460.
252. Shah KN, Lin FV, Yu F, McMahon JM: **Activity engagement and physical function in old age sample**. *Archives of Gerontology and Geriatrics* 2017, **69**:55-60.
253. Guedes RC, Dias RC, Pereira LS, Silva SL, Lustosa LP, Dias JM: **Influence of dual task and frailty on gait parameters of older community-dwelling individuals**. *Brazilian Journal of Physical Therapy* 2014, **0**:0.
254. Evans M, Guthrie N, Pezzullo J, Sanli T, Fielding RA, Bellamine A: **Efficacy of a novel formulation of L-Carnitine, creatine, and leucine on lean body mass and functional muscle strength in healthy older adults: a randomized, double-blind placebo-controlled study**. *Nutrition & Metabolism* 2017, **14**(1):7.
255. Patel A, Edwards M, Jameson K, Ward K, Fuggle N, Cooper C, Dennison E: **Longitudinal change in peripheral quantitative computed tomography assessment in older adults: the hertfordshire cohort study**. *Calcified tissue international* 2018, **103**(5):476-482.
256. Arias-Fernández L, Machado-Fragua MD, Graciani A, Guallar-Castillón P, Banegas JR, Rodríguez-Artalejo F, Lana A, Lopez-Garcia E: **Prospective association between nut consumption and physical function in older men and women**. *The Journals of Gerontology: Series A* 2018.
257. Hurst C, Weston KL, Weston M: **The effect of 12 weeks of combined upper-and lower-body high-intensity interval training on muscular and cardiorespiratory fitness in older adults**. *Aging Clinical and Experimental Research* 2018:1-11.
258. McGrath RP, Vincent BM, Lee I-M, Kraemer WJ, Peterson MD: **Handgrip strength, function, and mortality in older adults: a time-varying approach**. *Medicine and science in sports and exercise* 2018, **50**(11):2259-2266.

259. Ryan A, Murphy C, Boland F, Galvin R, Smith SM: **What Is the impact of physical activity and physical function on the development of multimorbidity in older adults over time? A population-based cohort study**. *Journals of Gerontology: Series A Medical Sciences* 2018, **73**(11):1538-1544.
260. Yiengprugsawan V, Steptoe A: **Impacts of persistent general and site-specific pain on activities of daily living and physical performance: A prospective analysis of the English Longitudinal Study of Ageing**. *Geriatrics & Gerontology International* 2018.
261. Trudelle-Jackson E, Jackson AW: **Do older adults who meet 2008 Physical Activity Guidelines have better physical performance than those who do not meet?** *Journal of Geriatric Physical Therapy* 2018, **41**(3):180-185.
262. Kramer JF, Vaz MD, Vandervoort AA: **Reliability of isometric hip abductor torques during examiner- and belt-resisted tests**. *Journal of Gerontology* 1991, **46**(2):M47-51.
263. Singh H, Kim D, Bemben M, Bemben D: **Relationship between muscle performance and DXA-derived bone parameters in community-dwelling older adults**. *Journal of Musculoskeletal & Neuronal Interactions* 2017, **17**(2):50.
264. Gronbech Jorgensen M, Andersen S, Ryg J, Masud T: **Novel use of the Nintendo Wii Board for measuring isometric lower limb strength: a reproducible and valid method in older adults**. *PLoS ONE* 2015, **10**(10):e0138660.
265. Abe T, Thiebaud RS, Loenneke JP, Mitsukawa N: **Association between toe grasping strength and accelerometer-determined physical activity in middle-aged and older women**. *Journal of Physical Therapy Science* 2015, **27**(6):1893-1897.
266. Assantachai P, Muangpaisan W, Intalapaporn S, Sitthichai K, Udompunturak S: **Cut-off points of quadriceps strength, declines and relationships of sarcopenia-related variables among Thai community-dwelling older adults**. *Geriatrics & gerontology international* 2014, **14 Suppl 1**:61-68.
267. Ahedi H, Aitken D, Scott D, Blizzard L, Cicuttini F, Jones G: **The association between hip muscle cross-sectional area, muscle strength, and bone mineral density**. *Calcified Tissue International* 2014, **95**(1):64-72.
268. Johnson LG, Butson ML, Polman RC, Raj IS, Borkoles E, Scott D, Aitken D, Jones G: **Light physical activity is positively associated with cognitive performance in older community dwelling adults**. *Journal of Science and Medicine in Sport* 2016, **19**(11):877-882.
269. Scott D, Blizzard L, Fell J, Ding C, Winzenberg T, Jones G: **A prospective study of the associations between 25-hydroxy-vitamin D, sarcopenia progression and physical activity in older adults**. *Clinical endocrinology* 2010, **73**(5):581-587.
270. Blomkvist AW, Andersen S, de Bruin E, Jorgensen MG: **Unilateral lower limb strength assessed using the Nintendo Wii Balance Board: a simple and reliable method**. *Aging Clinical and Experimental Research* 2017, **29**(5):1013-1020.
271. Gouveia ER, Maia JA, Beunen GP, Blimkie CJ, Fena EM, Freitas DL: **Functional fitness and physical activity of Portuguese community-residing older adults**. *Journal of Aging & Physical Activity* 2013, **21**(1):1-19.
272. Trapé AA, Lizzi EAdS, Gonçalves TCP, Rodrigues JAL, Tavares SS, Lacchini R, Pinheiro LC, Ferreira GC, Tanus-Santos JE, Ovídio PP: **Effect of multicomponent training on blood pressure, nitric oxide, redox status, and physical fitness in older adult women: influence of endothelial nitric oxide synthase (NOS3) haplotypes**. *Oxidative Medicine and Cellular Longevity* 2017, **2017**.
273. Hughes WE, Ueda K, Casey DP: **Chronic endurance exercise training offsets the age-related attenuation in contraction-induced rapid vasodilation**. *Journal of Applied Physiology* 2016:jap.00057.02016.
274. Borschmann K, Moore K, Russell M, Ledgerwood K, Renehan E, Lin X, Brown C, Sison J: **Overcoming barriers to physical activity among culturally and linguistically diverse older adults: a randomised controlled trial**. *Australasian Journal on Ageing* 2010, **29**(2):77-80.

275. Coker RH, Hays NP, Williams RH, Wolfe RR, Evans WJ: **Bed rest promotes reductions in walking speed, functional parameters, and aerobic fitness in older, healthy adults**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2015, **70**(1):91-96.
276. Goldberg A, Chavis M, Watkins J, Wilson T: **The five-times-sit-to-stand test: validity, reliability and detectable change in older females**. *Aging-Clinical & Experimental Research* 2012, **24**(4):339-344.
277. Judd DL, Thomas AC, Dayton MR, Stevens-Lapsley JE: **Strength and functional deficits in individuals with hip osteoarthritis compared to healthy, older adults**. *Disability & Rehabilitation* 2014, **36**(4):307-312.
278. McCarthy EK, Horvat MA, Holtsberg PA, Wisenbaker JM: **Repeated chair stands as a measure of lower limb strength in sexagenarian women**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2004, **59**(11):1207-1212.
279. Ng SS, Kwong PW, Chau MS, Luk IC, Wan SS, Fong SS: **Effect of arm position and foot placement on the five times sit-to-stand test completion times of female adults older than 50 years of age**. *Journal of Physical Therapy Science* 2015, **27**(6):1755-1759.
280. Stijntjes M, Meskers CG, de Craen AJ, van Lummel RC, Rispens SM, Slagboom PE, Maier AB: **Effect of calendar age on physical performance: A comparison of standard clinical measures with instrumented measures in middle-aged to older adults**. *Gait & Posture* 2016, **45**:12-18.
281. Mikhael M, Orr R, Amsen F, Greene D, Singh MA: **Effect of standing posture during whole body vibration training on muscle morphology and function in older adults: a randomised controlled trial**. *BMC Geriatrics* 2010, **10**:74.
282. Gluchowski A, Dulson D, Merien F, Plank L, Harris N: **Comparing the effects of two distinct eccentric modalities to traditional resistance training in resistance trained, higher functioning older adults**. *Experimental Gerontology* 2017, **98**:224-229.
283. Dobbs TJ, Simonson SR, Conger SA: **Improving power output in older adults using plyometrics in a body mass–supported treadmill**. *The Journal of Strength & Conditioning Research* 2018, **32**(9):2458-2465.
284. Kuo YL: **The influence of chair seat height on the performance of community-dwelling older adults' 30-second chair stand test**. *Aging-Clinical & Experimental Research* 2013, **25**(3):305-309.
285. Kalapotharakos VI, Michalopoulos M, Tokmakidis SP, Godolias G, Gourgoulis V: **Effects of a heavy and a moderate resistance training on functional performance in older adults**. *Journal of Strength & Conditioning Research* 2005, **19**(3):652-657.
286. Naugle KM, Higgins TJ, Manini TM: **Obesity and use of compensatory strategies to perform common daily activities in pre-clinically disabled older adults**. *Archives of Gerontology & Geriatrics* 2012, **54**(2):e134-138.
287. Baer GD, Ashburn AM: **Trunk movements in older subjects during sit-to-stand**. *Archives of Physical Medicine & Rehabilitation* 1995, **76**(9):844-849.
288. Marques E, Carvalho J, Pizarro A, Wanderlay F, Mota J: **The influence of physical activity, body composition, and lower extremity strength on walking ability**. *Motor Control* 2011, **15**(4):494-506.
289. Brose A, Parise G, Tarnopolsky MA: **Creatine supplementation enhances isometric strength and body composition improvements following strength exercise training in older adults**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2003, **58**(1):11-19.
290. Crockett K, Ardell K, Hermanson M, Penner A, Lanovaz J, Farthing J, Arnold C: **The relationship of knee-extensor strength and rate of torque development to sit-to-stand performance in older adults**. *Physiotherapy Canada* 2013, **65**(3):229-235.
291. Yasuda T, Fukumura K, Fukuda T, Uchida Y, Iida H, Meguro M, Sato Y, Yamasoba T, Nakajima T: **Muscle size and arterial stiffness after blood flow-restricted low-intensity resistance training in older adults**. *Scandinavian Journal of Medicine & Science in Sports* 2014, **24**(5):799-806.

292. King AC, Pruitt LA, Phillips W, Oka R, Rodenburg A, Haskell WL: **Comparative effects of two physical activity programs on measured and perceived physical functioning and other health-related quality of life outcomes in older adults**. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences* 2000, **55**(2):M74-83.
293. Smith JW, Marcus RL, Tracy BL, Foreman KB, Christensen JC, LaStayo PC: **Stance time variability during stair stepping before and after total knee arthroplasty: A pilot study**. *Human Movement Science* 2016, **45**:53-62.
294. Vincent KR, Braith RW, Feldman RA, Magyari PM, Cutler RB, Persin SA, Lennon SL, Gabr AH, Lowenthal DT: **Resistance exercise and physical performance in adults aged 60 to 83**. *Journal of the American Geriatrics Society* 2002, **50**(6):1100-1107.

Additional file 3

Additional file 3. The quality of studies assessing validity and/or reliability of included balance and strength tools and the rating of the reported results

			Validity				Reliability							
Assessment Tool			Construct		Criterion		Internal consistency		Inter-rater		Intra-rater		Test-retest	
	ref	n	Quality of study	Reported result	Quality of study	Reported result	Quality of study	Reported result	Quality of study	Reported result	Quality of study	Reported result	Quality of study	Reported result
Tandem, 10s	[1]	37			F	β. 92 R^2 .98								
	[1]	20											F	$ICC_{2,1}$ .82
TUG	[2]	12											P	PCC .90-.97
	[3]	10							P	$ICC_{3,1}$ .97				
	[4]	60	G	RR 3.2										
SPPB	[5]	150	E	N/A									G	ICC .87
CBM	[6]	51			F	SCC .32-.85	P	α .998	G	$ICC_{2,k}$ 0.97	G	$ICC_{3,k}$ 1.00		
	[7]	25	F	PCC .21-.63										
CBM (German)	[8]	51			F	SCC .32-.85	G	α .998	G	$ICC_{2,k}$ .996	G	$ICC_{3,k}$ .998		
5x STS	[9]	12									F	ICC .89		

Notes: Timed Up-and-Go (TUG), Short Physical Performance Battery (SPPB), Community Balance & Mobility scale (CBM), Five times sit-to-stand (5x STS); Quality of study (COSMIN): excellent (E), good (G), fair (F), poor (P); Cronbach's alpha (α), beta Correlation coefficient (β), Coefficient of determination (R^2), Risk Ratio (RR), Pearson Correlation Coefficient (PCC), Spearman Correlation Coefficient (SCC); Intraclass Correlation Coefficient (ICC), number (n), reference (ref), Not Applicable (N/A),

References

1. Scaglioni-Solano P, Aragón-Vargas LF: **Validity and reliability of the Nintendo Wii Balance Board to assess standing balance and sensory integration in highly functional older adults**. *International Journal of Rehabilitation Research* 2014, **37**(2):138-143.
2. Morris S, Morris ME, Iansek R: **Reliability of measurements obtained with the Timed "Up & Go" test in people with Parkinson disease**. *Physical Therapy* 2001, **81**(2):810-818.
3. Ng SS, Hui-Chan CW: **The timed up & go test: its reliability and association with lower-limb impairments and locomotor capacities in people with chronic stroke**. *Archives of physical medicine and rehabilitation* 2005, **86**(8):1641-1647.
4. Alexandre TS, Meira DM, Rico NC, Mizuta SK: **Accuracy of Timed Up and Go Test for screening risk of falls among community-dwelling elderly**. *Brazilian Journal of Physical Therapy* 2012, **16**(5):381-388.
5. Gomez JF, Curcio CL, Alvarado B, Zunzunegui MV, Guralnik J: **Validity and reliability of the Short Physical Performance Battery (SPPB): a pilot study on mobility in the Colombian Andes**. *Colombia Medica* 2013, **44**(3):165-171.
6. Weber M, Van Ancum J, Bergquist R, Taraldsen K, Gordt K, Mikolaizak AS, Nerz C, Pijnappels M, Jonkman NH, Maier AB: **Concurrent validity and reliability of the Community Balance and Mobility scale in young-older adults**. *BMC Geriatrics* 2018, **18**(1):156.
7. Takacs J, Garland SJ, Carpenter MG, Hunt MA: **Validity and reliability of the Community Balance and Mobility scale in individuals with knee osteoarthritis**. *Physical Therapy* 2014, **94**(6):866-874.
8. Gordt K, Mikolaizak AS, Nerz C, Barz C, Gerhardy T, Weber M, Becker C, Schwenk M: **German version of the Community Balance and Mobility Scale**. *Zeitschrift für Gerontologie und Geriatrie* 2018:1-9.
9. Cani KC, Silva IJCS, Karloh M, Gulart AA, Matte DL, Mayer AF: **Reliability of the five-repetition sit-to-stand test in patients with chronic obstructive pulmonary disease on domiciliary oxygen therapy**. *Physiotherapy Theory and Practice* 2018:1-7.

Paper II

Paper III

Original Paper

App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study

Ronny Bergquist[1], MSc; Beatrix Vereijken[1], PhD; Sabato Mellone[2], PhD; Mattia Corzani[2], MSc; Jorunn L Helbostad[1], PhD; Kristin Taraldsen[1], PhD

[1]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway
[2]Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Bologna, Italy

Corresponding Author:
Ronny Bergquist, MSc
Department of Neuromedicine and Movement Science
Norwegian University of Science and Technology
Edvard Griegs gate 8
Trondheim, 7030
Norway
Phone: 47 47634462
Email: ronny.bergquist@ntnu.no

Abstract

Background: Objective measures of physical function in older adults are widely used to predict health outcomes such as disability, institutionalization, and mortality. App-based clinical tests allow users to assess their own physical function and have objective tracking of changes over time by use of their smartphones. Such tests can potentially guide interventions remotely and provide more detailed prognostic information about the participant's physical performance for the users, therapists, and other health care personnel. We developed 3 smartphone apps with instrumented versions of the Timed Up and Go (Self-TUG), tandem stance (Self-Tandem), and Five Times Sit-to-Stand (Self-STS) tests.

Objective: This study aimed to test the usability of 3 smartphone app–based self-tests of physical function using an iterative design.

Methods: The apps were tested in 3 iterations: the first (n=189) and second (n=134) in a lab setting and the third (n=20) in a separate home-based study. Participants were healthy adults between 60 and 80 years of age. Assessors observed while participants self-administered the tests without any guidance. Errors were recorded, and usability problems were defined. Problems were addressed in each subsequent iteration. Perceived usability in the home-based setting was assessed by use of the System Usability Scale, the User Experience Questionnaire, and semi-structured interviews.

Results: In the first iteration, 7 usability problems were identified; 42 (42/189, 22.0%) and 127 (127/189, 67.2%) participants were able to correctly perform the Self-TUG and Self-Tandem, respectively. In the second iteration, errors caused by the problems identified in the first iteration were drastically reduced, and 108 (108/134, 83.1%) and 106 (106/134, 79.1%) of the participants correctly performed the Self-TUG and Self-Tandem, respectively. The first version of the Self-STS was also tested in this iteration, and 40 (40/134, 30.1%) of the participants performed it correctly. For the third usability test, the 7 usability problems initially identified were further improved. Testing the apps in a home setting gave rise to some new usability problems, and for Self-TUG and Self-STS, the rates of correctly performed trials were slightly reduced from the second version, while for Self-Tandem, the rate increased. The mean System Usability Scale score was 77.63 points (SD 16.1 points), and 80-95% of the participants reported the highest or second highest positive rating on all items in the User Experience Questionnaire.

Conclusions: The study results suggest that the apps have the potential to be used to self-test physical function in seniors in a nonsupervised home-based setting. The participants reported a high degree of ease of use. Evaluating the usability in a home setting allowed us to identify new usability problems that could affect the validity of the tests. These usability problems are not easily found in the lab setting, indicating that, if possible, app usability should be evaluated in both settings. Before being made available to end users, the apps require further improvements and validation.

(JMIR Mhealth Uhealth 2020;8(4):e16507) doi: 10.2196/16507

KEYWORDS

physical function; mHealth app; usability; older people; seniors

Introduction

At the time of retirement, at the age of 60-70 years, many people experience a significant decline in physical activity levels [1], and balance, gait, and mobility typically start to decline at a higher rate than before [2,3]. Thus, detection of changes in physical function at an early stage could be crucial to improve or prevent future declines in physical function and to sustain physical function over time. Objective assessment of physical function in health care settings is resource-demanding and therefore limited to people with a pressing need to have their function assessed, such as individuals who have experienced falls or who have been diagnosed with a condition known to affect physical functioning. Because functional decline typically occurs slowly, it might not pose an issue for the individual until their ability to perform activities of daily life is affected. Consequently, it might not be obvious why younger or well-functioning seniors should have their physical function assessed until it has come to this stage.

Innovations in mobile health (mHealth) technology have paved the way for new possibilities in assessing physical function. Most smartphones are equipped with sensors such as accelerometers, gyroscopes, and magnetometers and have high computational power; therefore, smartphones can be considered an inertial measurement unit enabling an objective and reliable assessment of physical function [4]. Considering that seniors are the fastest growing group of smartphone users [5] and that, in 2017, 42% of adults aged 65 or older in the United States owned a smartphone [6], there is great potential for using smartphones as a tool for self-assessing physical function [7]. Furthermore, well-designed and evidence-based apps represent new opportunities in preventive strategies for the senior population as a valuable tool in helping to make changes in their lives that can prevent functional decline.

Three such smartphone apps for self-assessment of physical function were developed as part of the PreventIT (early risk detection and prevention in aging people by self-administered ICT-supported assessment and a behavioral change intervention, delivered by use of smartphones and smartwatches) project. PreventIT was a European Union Horizon 2020 Personalising Health and Care project. The apps allow users to self-administer instrumented versions of the Timed Up and Go (Self-TUG), tandem stance (Self-Tandem), and Five Times Sit-to-Stand (Self-STS) tests in order to measure mobility and dynamic balance, static balance, and leg strength, respectively.

When developing an mHealth app for self-assessment of physical function, the usability of the app must be carefully considered, as it has been shown to be a fundamental determinant for technology adoption among older adults [8]. Usability is defined in the official International Organization for Standardization (ISO) guidelines as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [9]. Furthermore, when measuring aspects of one’s health, the accuracy of the results relies on correct administration of the test. Thus, any usability problem associated with using an app-based test should be identified and addressed before it is made available to end users. This is usually done through several iterations of testing with target user groups, ideally until no major usability problems exist with regards to using the apps and administering the test. Usability studies are most often carried out in a lab setting, which is convenient and offers a high degree of control, as opposed to field-based usability testing. However, field-based testing, which, in this context, would be a home setting, provides insight into how the system is used under more realistic situations. Depending on the system being tested and the phase of development, usability should ideally be tested in both lab and home settings.

The overall aim of this study was to evaluate whether people in our target group of seniors between the ages of 60 years and 80 years were able to reliably self-administer the tests on their own using the smartphone, apps, and instructions we provided without any interaction with the assessors. In this paper, we describe the 3 iterations of usability testing with target user groups that were needed to identify all major usability problems. Each iteration consisted of a development phase and subsequent testing phase. In the first 2 iterations, we performed the usability tests in a controlled lab setting, where the assessors had prepared the test setup and necessary materials for the participants beforehand. For the third testing phase, participants were in their own homes, where they needed to prepare the test setup themselves by following instructions presented within the apps. This study does not address the topic of algorithms for signals and data processing nor how to present specific information and feedback to the users based on the test results.

Methods

Design Overview

We developed 3 app-based self-tests of physical function within the European Union Horizon 2020 project PreventIT [10]. Technology development in PreventIT followed the ISO standard 9241-210 [9] on user-centered development of products, and an iterative design approach was used to develop and test the usability of the apps. Because our target group is community dwellers and not clinical patients, we did not follow the ISO norm for medical devices. The target group of the apps was community-dwelling people aged 60 years and older, able to walk independently, and without any cognitive, functional hearing, or visual impairments. The overall aim of the mobile-based, self-administrable functional tests is early identification of risk for age-related functional decline by extracting relevant digital biomarkers from the smartphone-embedded inertial sensors. The intended context of use for the apps is to guide preventive intervention strategies for the general population.

An initial version (version 1) of the Self-TUG and Self-Tandem apps was included for the first iteration. The apps were upgraded based on the results of this testing, and the Self-STS was added

as a third self-test. All 3 apps were tested under similar conditions during the second iteration (apps version 2). After further upgrades, version 3 of the apps was tested in a summative usability evaluation with a new group of volunteers in a home setting.

Participants

We included participants from two studies. First, we included participants from a multicenter, 3-armed, feasibility randomized controlled trial conducted within the PreventIT project. For the first and second iterations, we included participants from the main study if they had performed the self-administration of the apps during baseline (iteration 1) and follow-up (iteration 2). The inclusion criteria for the participants are described in detail in the protocol paper for the PreventIT trial [10]. In short, for iterations 1 and 2, we included 189 and 134 community-dwelling adults, respectively, aged between 61 and 70 years from Trondheim, Norway; Stuttgart, Germany; and Amsterdam, the Netherlands.

For iteration 3, we included 20 community-dwelling adults ranging in age from 60 years to 80 years (mean 68.7 years, SD 5.2 years) in Trondheim, Norway. Inclusion criteria were community-dwelling status, age between 60 years and 80 years, ability to walk 500 meters independently, Norwegian-speaking, ability to hear sound from a smartphone, and current user of a smartphone. Participants were excluded if they reported any severe cardiovascular, pulmonary, neurological, or mental diseases.

Description and Development of the Apps – From Version 1 to Version 3

We developed the apps using Android Studio 3.1.2 (Google, Mountain View, CA). Versions 1 and 2 of the apps were installed on a Samsung Galaxy S3 (Samsung, Seoul, Korea), while version 3 was installed on a Samsung Galaxy S8 (Samsung, Seoul, Korea).

Self-Timed Up and Go, Self-Sit to Stand, and Self-Tandem Apps

We created separate apps for each of the clinical tests (TUG, Five Times STS, and tandem stance). The apps were developed to be used as standalone tests, so one or more tests could be skipped if participants felt unsafe or did not want to perform a test. The TUG is a measure of mobility, in which the participant is timed while rising up from a chair, walking 3 meters, turning around, walking back, and sitting down again. In the Five Times STS, the participant is supposed to stand up from a chair and sit down again repeatedly 5 times as fast as possible, while being timed. In the tandem stance, the participant is supposed to place one foot in front of the other, heel-to-toe, in a straight line for 15 seconds, if possible. The Self-TUG uses an algorithm to detect the different phases of the TUG and the transitions between them (ie, sit-to-stand, walking, turning, turn-to-sit) from the sensor signals. Further, it calculates features from these phases, such as duration, velocity, jerkiness, and signal range, as well as gait features including number of steps, step duration, and gait speed. For the Self-STS, the algorithms analyze the sensor signals and calculate several features from the whole task, transitions, and separate sit-to-stand and stand-to-sit phases of each repetition. Finally, for Self-Tandem, the algorithms analyze the sensor signals and calculate features such as signal frequency, ellipse area, velocity, sway path, jerkiness, signal range, and spectral entropy.

Version 1

A multidisciplinary team designed the apps with emphasis on ease of use for the target group, corresponding to the term "perceived satisfaction" in the ISO terminology [9]. This included displaying buttons and icons in relatively large sizes and using contrasting colors on a white background. In addition, to ease the demands on working memory, the app screens were designed with as few elements and text as possible.

All apps are based on the same structure (Figure 1). For example, when opening the self-TUG app, a green "play" button appears. Pressing the button prompts a dialog box with a 5-second countdown and a red stop button. The countdown gives the user time to attach the smartphone to the lower back by means of a waist belt case (see Procedures). After the countdown and as soon as no movement is detected by the inertial sensors, an audio signal initiates the start of the test. At the end of the test, when the user is again sitting still, an audio signal indicates that the test is completed. The Self-STS has the same structure (ie, audio signals for both the start and end of the test when the participant is sitting still). One important difference for the Self-Tandem is that the start and end signals are activated by time and not by movement. Thus, the audio start signal is initiated immediately after a 5-second countdown, followed by the end signal after 15 seconds.

Version 2

In version 2 of the Self-TUG and Self-STS, the smartphone was worn in the front trouser pocket instead of the waist belt case. We also integrated instruction videos into the apps. By pressing "play," a dialog box appears with a question asking whether the participants want to see the instruction video (with a yes/no choice; Figure 2). Pressing "yes" starts the instruction video for how to perform the self-test. Pressing "no" results in the question "Do you want to start the test?" with the options "Yes" or "No, play the instruction video." A reminder of what to do (insert phone in pocket for Self-TUG and Self-STS, hold against chest for Self-Tandem) was added to the countdown dialog box. The apps were otherwise similar in structure as in version 1.

Version 3

The upgrades made for the third version of the apps included new instructional videos, with updated voiceover and footage, and new graphical elements in the video to emphasize important details of how to perform the tests (Multimedia Appendix 1) as well as a new menu structure where the user could choose to view instructions or start the test. Instructions consisted of a submenu with two options: watch the instructions for how to prepare the test setup or how to perform the test.

In addition, new features (Figure 3) included a warning message that popped up if a user tried to perform a test without having watched both instructions; voiceover that instructed the user on what to do once the test sequence had been initiated (ie, "Put the phone gently in your right pocket. Sit down and wait for my

instructions"); and real-time verbal feedback based on the inertial signals from the smartphone (eg, "sit down," "get up from the chair," "proceed with the test," and count of repetitions for the Self-STS). The instruction videos were made for Norwegian study participants; thus, voiceover and text elements were in Norwegian. The text on the menu and dialog box was automatically adapted to the system language of the phone.

Figure 1. Screenshots of the first version of the Self-Timed Up and Go test.

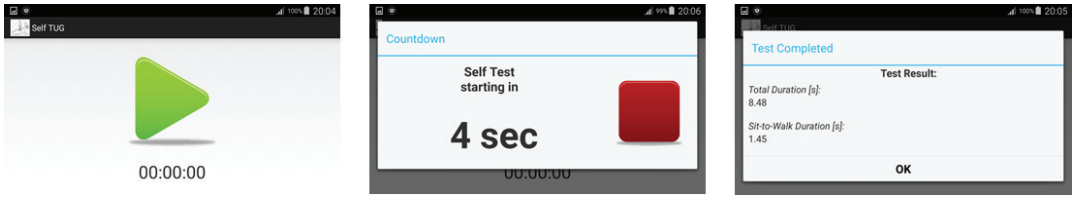

Figure 2. Screenshots of the second version of the Self-Timed Up and Go test.

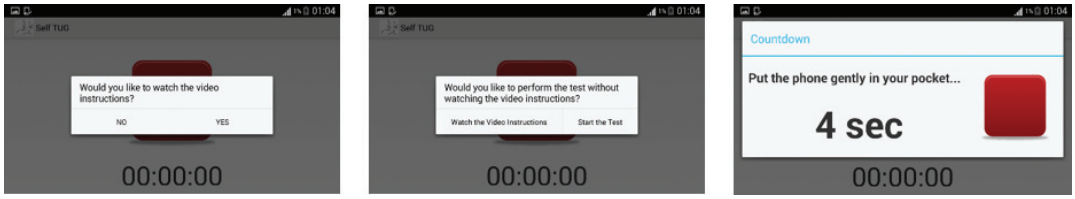

Figure 3. Screenshots of the third version of the Self-Timed Up and Go, including the start menu, instructions menu, test setup, instruction video screenshots, warning prompt when trying to start the test without having opened the instructions first, and instructions after starting the test.

Procedures

Testing of Version 1

The testing of version 1 was carried out in a lab setting by trained assessors. Before testing, the assessors prepared the setup, which included a chair placed against the wall with a 3-meter walkway in front of it and a beanbag at the 3-meter mark. Following a standardized introduction of the general purpose and procedures of a usability test, the participants were asked to self-administer the tests using the app on the

smartphone. The participants did not receive any guidance from the assessors during the tests, and the materials they needed to perform the test were placed on a chair in front of them. This included the smartphone, written instructions, and a belt case for wearing the phone during the Self-TUG.

The assessors observed while the participants attempted to self-administer the tests, recording issues and errors on a sheet with predefined errors and issues that we expected in addition to a free-text box to record other errors and issues.

Testing of Version 2

For the second step, the Self-STS was added to the self-test battery. Testing happened under the same conditions as during testing of the first version, with one exception for the Self-TUG and Self-STS, namely that the smartphone placement was changed from the belt case to the front trousers pocket.

Testing of Version 3

An instructor visited the participants in their homes to get a realistic impression of how the system would be used in a real-world home setting. After a standardized introduction, the participants were asked to prepare and self-administer each of the 3 self-tests 3 times, without guidance from the instructor. One GoPro camera (GoPro, San Mateo, CA) was attached with a harness to the participant's chest and worn during the test sequence to record the participant's interaction with the smartphone. A second GoPro camera was placed in the room in a position where all movements could be recorded. The participants were encouraged to think aloud when using the system. After performing the self-tests, we asked participants to complete two questionnaires: User Experience Questionnaire (UEQ) and Systems Usability Scale (SUS) [11]. This was followed by an audiotaped semistructured interview that was developed specifically for this study, where we aimed to collect end users' views on topics relevant to the apps, such as user experience, feedback/results, suggestions for improvements, and general usefulness of the apps.

Data Processing and Analyses

We defined errors as deviations from the test instructions and counted the number of errors from the clinical record forms in the first and second iterations and from video recordings in the third iteration (Multimedia Appendix 2). SUS scores, based on 5-point Likert-scale items, were averaged for each participant and converted into a usability score with a range from 0 to 100 (with a higher score indicating better usability). UEQ Likert-scale items were scored from 0 (highly agree) to 4 (highly disagree), and frequencies of responses within each category across all items were calculated.

The interview transcripts were analyzed using thematic analysis [12] to identify relevant themes. Quotes were extracted for each subtheme and translated from Norwegian to English for analysis. The questions presented to the participants were "What did you think about using these apps to test your physical function?" and "Do you have any ideas for how the apps can be improved?"

Ethics

The data collection was performed in accordance with the Helsinki ethical guidelines. The first and second usability testing phases were approved by the ethical committees in Norway (REK midt, 2016/1891), Stuttgart (registration number 770/2016BO1), and Amsterdam (METc VUmc registration number 2016.539 [NL59977.029.16]). The Norwegian Center for Research Data approved that the data protection for the third usability testing was in accordance with current regulations (ref. no. 391684). All participants included in this study gave their informed consent.

Results

Participants' characteristics are presented in Table 1.

Table 1. Participant characteristics.

Cohort	App version	n	Age (years), mean (SD)	Male gender, n (%)	Has smartphone experience, n (%)	Years of education, mean (SD)
PreventIT study	1	189	66.3 (2.4)	90 (47.4)	157 (83.1)	15.6 (4.6)
PreventIT study	2	134	66.3 (2.5)	64 (47.8)	108 (80.0)	15.9 (4.8)
Summative usability evaluation	3	20	68.7 (5.2)	11 (55.0)	20 (100.0)	__[a]

[a]Data were not collected.

The usability problems identified, numbers of participants who experienced these problems, and what was done to eliminate or reduce these problems are presented in Tables 2-4.

Table 2. Usability problems in version 1 of the Self-Timed Up and Go and Self-Tandem apps, rate of errors, and solutions (n=189).

Problem ID	Usability problem	Rate of errors/trials	Improvements made
1	Incorrect performance	120/378 (32%)	Added instruction video
2	Performed test without starting app	22/378 (6%)	Implemented instruction video that clearly demonstrates that the play button needs to be pressed before performing the test
3	Did not sit still and wait for start signal after test was started	23/189 (12%)	Added instruction video (demonstrating sitting still and waiting for the start signal before starting the test) and shortened the delay in the algorithm to limit any confusion
4	Incorrect placement of phone	32/378 (8%)	Changed placement to front pocket for Self-TUG[a] and Self-STS[b] and added a reminder in the countdown screen on what to do first (eg, "put the phone gently in your pocket")
5	Did not hear/perceive instructions	18/378 (5%)	Changed placement to front pocket for Self-TUG and Self-STS
6	Accidentally cancelled the test	15/378 (4%)	Not possible to override the home button function in the android OS, change of placement the preferred solution to reduce this problem

[a]TUG: Timed Up and Go.

[b]STS: Sit to Stand.

Table 3. Usability problems in version 2 of all 3 self-tests, rate of errors, and solutions (n=134).

Problem ID	Usability problem	Rate of errors/trials	Improvements made
1	Incorrect performance	66/402 (16%)	Added new, improved instructions to the videos (new voiceover and added graphical elements to draw attention to the details of the test procedures); added a warning message that appears if trying to start the test without watching instructions; and added real-time TTS[a] voice feedback on the number of repetitions in the Self-STS[b]
2	Started performing test (during instruction video) without starting the test in the app	28/402 (7%)	Changed structure of the app: main window now has two separate buttons, one for "start test" and one for "instructions"
3	Did not sit still and wait for start signal after test was started in the app	39/268 (15%)	Added real-time verbal step-by-step instructions that are initiated after the test is started in the app
4	Incorrect placement of phone	11/402 (3%)	Added real-time verbal instruction explaining where to place the phone and when to do this
5	Did not hear/perceive instructions	4/402 (1%)	Changed settings in the app so that the volume is always on maximum levels during testing, to prevent participants from accidentally pressing the "volume down" button
6	Accidentally cancelled the test	8/402 (2%)	Reduced the size of the "stop" button

[a]TTS: text-to-speech.

[b]STS: Sit to Stand.

Table 4. Usability problems identified in version 3 of all 3 self-tests and the rate of errors (n=20).

Problem ID	Usability problem	Rate of errors/trials
1	Incorrect performance	19/60 (32%)
2	Performed test (during instruction video) without starting the test in the app	5/60 (8%)
3	Did not sit still and wait for start signal after test was started in the app	0/40 (0%)
4	Incorrect placement of phone	0/60 (0%)
5	Did not hear/perceive instructions	2/60 (3%)
6	Accidentally cancelled the test	1/60 (2%)

Iteration 1

In total, at least 1 error was made in 120 of 378 (32%) trials during the first usability testing with the Self-TUG and Self-Tandem. Forgetting or misunderstanding the written instructions were the leading causes of errors. In order to reduce the errors caused by this usability problem, we created video instructions to replace the written instructions.

Iteration 2

In the second usability test, errors due to usability problem 1 (incorrect performance of test) were made in 66 of 402 trials (16%). Percentage of errors due to usability problems 2 (performs test without starting app) and 3 (did not sit still and wait for start signal after test was started) increased from the first usability test, while the frequency of problems 4-6 (incorrect placement of phone, did not hear/perceive instructions, accidentally cancelled the test, respectively) decreased.

Iteration 3

In the third summative usability evaluation, errors due to usability problem 1 (incorrect performance of test) were made in 19 of 60 (32%) trials. Usability problems 3 (did not sit still and wait for start signal after test was started) and 4 (incorrect placement of phone) were eliminated, while the frequencies of usability problems 2 (performs test without starting app), 5 (did not hear/perceive instructions), and 6 (accidentally cancels the test) remained similar.

Table 5 presents an overview of the proportions of correctly performed (first) trials of self-tests for all tests in all iterations.

Table 5. Number of correctly performed self-tests (first trial) with versions 1, 2, and 3.

	Self-TUG[a], n (%)	Self-STS[b], n (%)	Self-Tandem, n (%)
Testing of version 1	42 (22.0)	N/A[c]	127 (67.2)
Testing of version 2	108 (83.1)	40 (30.1)	106 (79.1)
Testing of version 3	14 (70.0)	5 (25.0)	18 (90.0)

[a]TUG: Timed Up and Go.

[b]STS: Sit to Stand.

[c]Not yet developed.

Perceived Ease of Use

UEQ scores for iteration 3 are presented in Table 6, indicating a positive or very positive user experience on all 6 items. Seven sub-themes of perceived ease of use were identified in the analysis of interview transcripts and are presented in Multimedia Appendix 3 with accompanying sample quotes, mapped to proposed solutions.

Table 6. Frequency of scores across the 6 items in the User Experience Questionnaire administered in the summative user evaluation (n=19).

Likert scale items	Strongly agree	Agree	Neither agree/disagree	Disagree	Strongly disagree
The set-up instructions were clear and easy to follow	14	4	1	0	0
The text in the app was easy to read	15	4	0	0	0
The buttons in the app were easy to discern from other elements	12	5	0	2	0
The signals were easy to hear	15	3	0	1	0
It was easy to navigate around in the apps	10	6	0	1	1
The instruction videos were clear and easy to follow	14	5	0	0	0

The mean score on the SUS for version 3 was 77.63 points (SD 16.1 points, range 42.5-97.4 points). Of the 20 participants, 14 participants scored the SUS above 66.5 points, which is the average SUS score for cell phones [13].

Discussion

Principal Findings

This paper describes the development and usability testing of the Self-TUG, Self-STS, and Self-Tandem through 2 iterations in the lab and 1 in a home setting. Our aim was to develop app-based, self-administrable tests of physical function that participants could use with a high degree of effectiveness and perceived ease of use. The first phase of testing revealed usability problems that affected the validity of the test results, illustrating a clear need for improvements. We addressed all usability problems by making changes to the app design, test algorithms, and test setup, which led to a large decrease in the number of trial errors in the second usability testing. The work on the third version of the apps then started, which included updating and adding new instructions for a version fully adapted for use in a home-based setting.

The results from the SUS, UEQ, and thematic analysis from the usability testing in the home setting indicated that the participants experienced high levels of perceived ease of use when using the apps. Still, errors were made that may affect the validity of the tests, most of which were caused by misunderstanding the instructions. As an example, the most

common error for Self-STS was not performing it as fast as possible, which was the main reason why only 25.0% (5/20) performed it correctly on their first attempt. This is lower than in the second version tested in the lab (40/134, 30.1%). This misunderstanding was caused by a delay in the real-time counting of repetitions that was implemented in the third version of Self-STS. The verbal announcement of repetitions, which is also done by the assessor when the original Five Times STS is administered in the clinic, was implemented to motivate the participant to perform it faster and as a way for the participant to keep track. However, as can be seen in the sample quotes from the participant interviews (Multimedia Appendix 3), there was a delay in the real-time feedback, making many of the participants stop and wait in a standing position for the TTS to announce the repetition before sitting down. This slowed down the performance and thus impaired instead of improving the validity of the test.

During the Self-TUG, a common error was to measure an incorrect distance for the walkway during the set-up. Although the instructions state that the walkway should be 3 meters, the participant responses indicate that they did not consider it crucial to measure exactly 3 meters. However, it has to be exactly 3 meters if the total test duration, walking duration, or gait speed is to be used as an outcome measure, as these features rely on a standardized distance walked. A clarification in the instruction, where it is specified that the walkway needs to follow a straight line of exactly 3 meters, could be one way to increase the reliability of the test. However, as the distance walked by the participant cannot be accurately measured by the app, another approach could be to only exploit the distance-independent signal features, such as those calculated from the sit-to-stand, turning, and turn-to-sit phases. This will improve the system reliability in assessing motor performance, but it will not ensure full compatibility with the standard clinical measure of the total test duration.

Another common error with the Self-TUG was to press “Start test” without watching the instructions first. Although we had implemented a pop-up warning if this happened, a bug prevented this from happening in 5 of the 7 times this occurred. For the 2 participants who received the pop-up warning, 1 ended up watching the instruction video and performing the test correctly, while the other ignored the warning and proceeded to perform the test without watching the instruction video, thus performing it incorrectly. Because of the bug, we do not have sufficient information to make a safe claim regarding the effectiveness of the warning message. However, we assume that this problem will be resolved by fixing the bug and specifying in the warning message that a correct trial depends on having watched the instruction video first.

A common usability problem observed with the Self-Tandem, and also mentioned by many of the participants in the interviews, was the discrepancy between the instructions and actual duration required to stand in the tandem position. The instructions state that the participant is supposed to place the feet in tandem, hold the phone against the chest, and, after hearing the start signal, keep as still as possible for 15 seconds. What often happened in the current version, however, was that the app tried to detect and verify the position of the smartphone after the participant had been instructed to place the feet in a tandem position. The TTS then instructed the participant to hold the tandem position and keep as still as possible for 15 seconds, until the end signal. Depending on how fast they placed the phone against the chest, the participant could thus stand up to 25 seconds in total. However, if the instructions had said that the participant should assume the tandem position after hearing the start signal, different people would likely need a different amount of time to get into the correct position, thereby risking that we would get less than 15 seconds of actual tandem balancing. The outcome measure in Self-Tandem is mediolateral sway, which was found to be a strong predictor of age-related decline in a study in which an eyes-open condition was used [14]. We therefore designed the test in a way that would ensure, or at least increase the chance, that we would have at least 15 seconds of the participant standing in tandem.

A limitation with the Self-Tandem test is that we cannot infer whether the participant was keeping the correct tandem position for the entire 15 seconds from the inertial sensor signals. This is not true for the Self-STS and Self-TUG tests, where the correct performance of all phases of the test can be identified reliably from the signal. A potential solution could be to implement a pop-up question where the user self-reports whether they actually held the position for the entire duration. Such a solution has been implemented in the mHealth app “Steady” [14]. Steady is a falls risk app that consists of a health history questionnaire, 4 balance tasks (eyes open, eyes closed, tandem, and single leg), and a 30-second sit-to-stand test. The binary outcome measure of whether a user is able to complete a static steady-state balance task in various conditions and durations, such as those used in Steady, has been used extensively in studies assessing healthy young seniors [15].Therefore, adding this feature could potentially increase the Self-Tandem’s diagnostic/prognostic abilities.

Implications and Future Work

The 3 iterations of usability testing described in this paper were sufficient to identify all major usability problems with the self-tests. The only problem remaining after the third cycle is the real-time counting in the Self-STS, described earlier in the discussion, which can easily be fixed.

We have demonstrated what challenges can be expected when developing app-based tests of physical function for seniors and how solutions to specific usability problems identified in one iteration affected the same problems in the next iterations (Tables 2-5). In addition, we described the perspectives of the seniors regarding their experience of using the apps to self-administer the tests (Multimedia Appendix 3). Another interesting insight is how going from the lab to a home-setting influenced the type of usability problems observed, in particular those related to the test setup. In the lab setting, the setup was prepared beforehand, whereas in the home setting, participants needed to follow the instructions in the app describing how to measure the walkway in the TUG, secure the chair for Self-STS, and perform the Self-Tandem in a spot with a secure object within hands reach, without any guidance from the assessor.

The next step in the developmental process of the apps is to implement the solutions proposed in Multimedia Appendix 3

to address the remaining usability problems and conduct new usability tests to ensure that the apps are ready to be used by the target group to self-administer the tests safely and correctly. Furthermore, the algorithms used for signal processing in Self-TUG and Self-STS need to be validated with the changed placement of the smartphone from the lower back (version 1) to the thigh (versions 2 and 3). Although we experienced some issues with this new placement in the usability testing (eg, some trousers were too loose), with the smartphone tilting down on either side of the thigh and making the trial invalid, we nevertheless believe that this solution offers the best trade-off between motion detection ability on one side and ease of use on the other side.

Our app-based tests of physical function could be applicable in many contexts, and different contexts may require different test outcomes. In the current version of the apps, the results presented to the user after performing the tests are total durations for the Self-TUG and Self-STS and sway path distance in the Self-Tandem. As discussed, however, these might not be feasible to exploit from an unsupervised test, where correct test set-up cannot be verified. The data processed by the inertial sensors within the smartphone provide additional features, and we aim to review existing literature to identify which of the signal features from instrumented versions of the TUG, Five Times STS, and standing tandem are the most predictive of functional decline in seniors. Given that these features can be reliably measured with the smartphone worn in the trouser pocket, they will be exploited as outcomes presented to the app users.

Although many tests of physical function have been instrumented by the use of smartphones, the authors are only aware of one other app that is developed for unsupervised self-assessment, the Steady app [14]. What separates the Steady app from ours is the type of tests implemented in the app. In addition to static balance and repeated sit-to-stands, we integrated the instrumented TUG. Furthermore, we performed usability assessments of the app in the participants' own homes, in contrast to Steady, where an unoccupied apartment was used for all non-lab test sessions in order to mimic a home environment.

Limitations

In our first 2 usability tests, the apps were tested by 189 and 134 participants, respectively. Although this was very useful for identifying what did and did not work well, we might have achieved similar results with fewer participants. Earlier studies have suggested that as few as 12 test users can be sufficient to detect the majority of usability problems [16]. Thus, with shorter and faster test cycles, the apps could potentially have been at a more mature stage today.

The participants in the summative usability evaluation differed to those from the PreventIT study in terms of age and smartphone experience. This makes it more difficult to say something about the impact that each app improvement had, as opposed to testing all app versions in 3 different, but homogenous, cohorts. However, we see it as an advantage that the apps are also tested in a slightly older cohort, as these participants can help us identify problems that could be more relevant to how they experience the usability of the apps, as compared to seniors that are younger or more experienced with technology in their daily life. Furthermore, the self-tests have not been validated in persons with tremor or pathologies; thus, the results do not necessarily generalize to these populations.

ISO's definition of usability comprises 3 main aspects: effectiveness, which is the accuracy and completeness with which users achieve certain goals; efficiency, which is the relationship between the accuracy and completeness with which users achieve certain goals and the resources expended in achieving them; and satisfaction, which is the user's comfort with and positive attitudes towards the use of the system [17]. Efficiency was not measured in our usability studies. It is often measured as task completion time or learning time, but in the context of testing physical function, where the time spent on completing a task also depends on the person's physical abilities, we did not consider task completion time to be an appropriate outcome measure of usability, but rather of functional level of the participant.

Although we have assessed the usability of these apps and identified solutions to the remaining usability problems, the validity of the outcome measures from the tests also needs to be further investigated before being made available to end users. Another point worth mentioning is that the correct use of the apps and, accordingly, valid test results could be ensured by giving the end users a one-time demonstration of how to use apps and perform the tests correctly. This could be given in a home visit or in an appointment at the lab or clinic, depending on the context of use.

Conclusion

The study results suggest that the apps have the potential to be offered as a solution for self-testing of physical function in a nonsupervised, home-based setting. Participants found the apps easy to use. The summative user evaluation in a home setting revealed important usability problems that were not identified in the lab, highlighting the importance of utilizing both test settings when assessing app usability. The current version of the apps has some remaining usability problems that can affect the test results, indicating that the apps need to be further improved and then validated before being made available to end users.

Acknowledgments

The authors wish to thank Andreas Hansen Schille and Jonas Langset Hustad for their contribution to the editing of the instruction videos and Gemma Goodall for proofreading the manuscript. We thank all participants for being part of the study and providing us with valuable input.

This study was supported by PreventIT, which received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 689238. The content is solely the responsibility of the authors and does not necessarily

represent the official views of the European Commission. The European Commission had no role in the design, execution, interpretation, or writing of the study.

Conflicts of Interest

RB, BV, MC, JLH, and KT declare no conflict of interest. SM holds a share of the mHealth Technologies srl company, which owns the rights to some of the movement analysis algorithms.

Multimedia Appendix 1

Video instruction available within version 3 of the Self-TUG app, demonstrating how to perform the Self-TUG, with voiceover in Norwegian.
[MP4 File (MP4 Video), 21012 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Detailed description of usability problems.
[DOCX File , 16 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Sub-themes within perceived ease of use and sample quotes from participant interviews following the third iteration of usability testing and proposed solutions.
[DOCX File , 15 KB-Multimedia Appendix 3]

References

1. Olesen K, Rod NH, Madsen IEH, Bonde JP, Rugulies R. Does retirement reduce the risk of mental disorders? A national registry-linkage study of treatment for mental disorders before and after retirement of 245,082 Danish residents. Occup Environ Med 2015 May 26;72(5):366-372 [FREE Full text] [doi: 10.1136/oemed-2014-102228] [Medline: 25814269]
2. Choy NL, Brauer S, Nitz J. Changes in postural stability in women aged 20 to 80 years. J Gerontol A Biol Sci Med Sci 2003 Jun 01;58(6):525-530. [doi: 10.1093/gerona/58.6.m525] [Medline: 12807923]
3. Isles R, Choy N, Steer M, Nitz J. Normal values of balance tests in women aged 20-80. J Am Geriatr Soc 2004 Aug;52(8):1367-1372. [doi: 10.1111/j.1532-5415.2004.52370.x] [Medline: 15271128]
4. Mellone S, Tacconi C, Chiari L. Validity of a Smartphone-based instrumented Timed Up and Go. Gait Posture 2012 May;36(1):163-165. [doi: 10.1016/j.gaitpost.2012.02.006] [Medline: 22421189]
5. Joe J, Demiris G. Older adults and mobile phones for health: a review. J Biomed Inform 2013 Oct;46(5):947-954 [FREE Full text] [doi: 10.1016/j.jbi.2013.06.008] [Medline: 23810858]
6. Anderson MP, A. Pew Research Center. 2017. Technology use among seniors URL: https://www.silvergroup.asia/wp-content/uploads/2017/07/Technology-use-among-seniors-_-Pew-Research-Center.pdf [accessed 2019-08-01]
7. Weiss A, Herman T, Plotnik M, Brozgol M, Giladi N, Hausdorff J. An instrumented timed up and go: the added value of an accelerometer for identifying fall risk in idiopathic fallers. Physiol Meas 2011 Dec;32(12):2003-2018. [doi: 10.1088/0967-3334/32/12/009] [Medline: 22094550]
8. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
9. ISO. Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability. Geneva: International Organization of Standardization; 1998. Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability Genewa, Switzerland 1998; URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en [accessed 2019-08-01]
10. Taraldsen K, Mikolaizak AS, Maier AB, Boulton E, Aminian K, van Ancum J, et al. Protocol for the PreventIT feasibility randomised controlled trial of a lifestyle-integrated exercise intervention in young older adults. BMJ Open 2019 Mar 20;9(3):e023526 [FREE Full text] [doi: 10.1136/bmjopen-2018-023526] [Medline: 30898801]
11. Brooke J. SUS - A quick and dirty usability scale. In: Jordan P, Thomas B, Weerdmeester B, McClelland I, editors. Usability Evaluation In Industry. London: Taylor and Francis; 1996:189-194.
12. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
13. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
14. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. JMIR Aging 2018 Nov 20;1(2):e11569 [FREE Full text] [doi: 10.2196/11569] [Medline: 31518234]

15. Bergquist R, Weber M, Schwenk M, Ulseth S, Helbostad J, Vereijken B, et al. Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review. BMC Geriatr 2019 Jan 09;19(1):9 [FREE Full text] [doi: 10.1186/s12877-018-1011-0] [Medline: 30626340]
16. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT. 169166: ACM; 1993 Presented at: Conference on Human Factors in Computing Systems; 1993; Amsterdam, The Netherlands p. A-13. [doi: 10.1145/169059.169166]
17. Frøkjær E, Hertzum M, Hornbæk K. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? 332455: ACM; 2000 Presented at: The SIGCHI conference on Human Factors in Computing Systems; 2000; The Hague, The Netherlands p. a-52. [doi: 10.1145/332040.332455]

Abbreviations

ISO: International Organization for Standardization.
mHealth: mobile health.
STS: Sit-to-Stand.
SUS: Systems Usability Scale.
TUG: Timed Up and Go.
UEQ: User Experience Questionnaire.

Edited by C Dias; submitted 09.10.19; peer-reviewed by F Lamers, K Hsieh, E Sadeghi-Demneh, JP Leuteritz, J Knitza; comments to author 19.11.19; revised version received 07.01.20; accepted 21.02.20; published 27.04.20

Please cite as:
Bergquist R, Vereijken B, Mellone S, Corzani M, Helbostad JL, Taraldsen K
App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study
JMIR Mhealth Uhealth 2020;8(4):e16507
URL: http://mhealth.jmir.org/2020/4/e16507/
doi: 10.2196/16507
PMID:

Appendix 2. Detailed description of usability problems

Problem ID	Problem category: performance of tasks	Description
2	Incorrect performance of task	Does not walk around object (TUG), does not sit down when returning to the chair (TUG), STS clearly not performed as fast as possible, not 5 reps, feet not fully extended
6	Starts performing test during instruction video	When seeing the phone being inserted into pocket in instruction video, participant start doing it themselves
12	Not able to keep tandem position for 15s	Moves feet and/or holds onto something during tandem
Problem ID	**Problem category: set-up**	
1	Pocket not suitable for testing / real-time feedback not working	Pocket too short or too loose, preventing the phone to be kept in place on the thigh in TUG and STS. Leads to one or more of the following scenarios: 1) a seating position is not detected, which is required for the test sequence to start, and in order to detect that the user is finished (have sat down), 2) repetitions in STS not counted accurately, causing the participant to stop or slow down, and/or perform more than five repetitions.
3	Heels not in contact with floor in STS	
5	TUG walkway incorrect	Walkway is curved, too long/short, or partly obstructed by furniture, or there's no object on the floor
17	No wall or chair for safety in tandem	
20	Chair not secured in STS	
Problem ID	**Problem category: app usability**	
7	Did not perceive instruction	Did not hear clearly, e.g. due to speaker being squeezed, or impaired hearing
8	Presses the app icon more than once	Due to the delay of opening the app, the participant presses repeatedly, with the risk of pressing "instructions" or "start test" without knowing it

9	Struggles somewhat with opening the app	Pressing the button too long, moving the finger while pressing, pressing too lightly
10	Repeats instruction video instead of pressing "Back to test"	After having seen the instruction video the participant does not recognise how to start the test, and presses the "how to perform the test"-button again, as if he/she forgot that this is only the video, and that the test needs to be started in the previous menu. Category chosen when user clearly is looking for a way to stop the video, and not when participant watches video again on purpose.
11	Presses home button instead of back-button during video	participant presses home-button to stop the video, instead of the back-button
13	Accidentally presses "lock"-button while wearing phone in pocket	
14	Not able to get back from results-screen without support	
15	Presses "start test" without pressing "instructions" first	When opening app, participant ignores the "instruction"-button and directly presses the "start test"-button
16	Starts test without watching instructions first	Got warning, but started test anyway. "*" means warning did not show up (due to bug).
18	Accidentally cancels the test	E.g. pressing home-button
19	Other	Problems which can't be explained, bugs etc
21	Did not see the entire instruction video	

Appendix 3. Sub-themes within perceived ease of use and sample quotes from participant interviews following the third iteration of usability testing and proposed solutions.

Sub-themes of perceived ease of use	Sample quotes	Solution
Apps were easy to use	*"I think it went surprisingly well." ... "When you take the time to read the instructions, it's okay" (M, 72)*	
Unclear written instructions	*"The first app I think had too much text." … "You have to read a lot before you understand what you are supposed to do"(M, 72)*	Video of set-up
Trying to start the test without reading and watching instructions first	*"I just, I think the start symbol was overwhelming. A green, large [arrow], so you became motived to push it right away" (M, 69)* *"If you think of this as something you do regularly, I think it's the first time you fumble with this, and then it will go very automatically" (M, 61)*	Insert text: "Step 1", "Step 2"and "Step 3" above buttons for set-up, performance instructions and test, respectively. Also make the "Start test"-button grey until both instructions have been opened.

Real-time verbal counting of repetitions in Self-STS was unhelpful	*"It took a second or two before she said it [announced the repetition]. So it was she who controlled the tempo, in a way." (M, 72)* *"It seemed almost as if you had to slow down in order for the woman to keep up" (F, 63)* *"I think it was unnecessary, actually. Because you are able to count to five." (M, 63)* *"If that was the goal [to do it as fast as possible], it turned out wrong" (M, 62)*	Enhance algorithm to detect and announce repetition sooner, or make the TTS announce only after all 5 repetitions have been performed.
Self-TUG walkway not accurately pre-measured	*"I thought that three meters, approximately three meters, that it was not a big case. If it is supposed to be exactly three meters, it should say so." (M, 63)*	State in the set-up instructions that it is important that the walkway is exactly 3 meters.

Getting from instructions to test	*"..."go back", then I think "have I done something wrong?", then I have to go back to the beginning, in a way" (F, 72)* *"To begin with I didn't quite get it. Then, after familiarising myself with it, it wasn't a problem" (F, 79)*	Place the "Start the test" button on the same menu as instruction buttons are. Make the button grey and unclickable, until both instructions have been opened; then it turns green and clickable.
Verbal instructions	*"I think it took a very, you were standing for a very long time in that [tandem] position before he said "now you are supposed to stand still"" (M, 65)* *"I was instructed to place my feet in tandem, and by the time I was told that the test is starting, I had already been standing and keeping my balance" (M, 72)* *"..because it took such a long time, that it actually, that I started to wonder whether I, if something was wrong, that I hadn't pressed [the button] right" (M, 65)*	Enhance algorithms to reduce the waiting time. For the Self-Tandem: State in the instructions that the participant can use one hand to hold onto a chair or table until the TTS instructs the participant to let go.
App structure	*"So, you have three apps (sighs), will I scroll through them (?). If you have one app, which is easy to navigate within. There's no reason for anything else, in my opinion" (M, 70)*	Integrate the three apps into a single app.

Paper IV

Research Article

Predicting advanced balance ability and mobility with an instrumented Timed Up and Go-test

Ronny Bergquist[1,*], Corinna Nerz[2], Kristin Taraldsen[1], Sabato Mellone[3], Espen A. Ihlen[1], Beatrix Vereijken[1], Jorunn L. Helbostad[1], Clemens Becker[2] and A. Stefanie Mikolaizak[2]

1 Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway
2 Department for Clinical Gerontology, Robert-Bosch-Hospital, Stuttgart, Germany
3 Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Bologna, Italy
* Correspondence: ronny.bergquist@ntnu.no

Abstract: Extensive test batteries are often needed to obtain a comprehensive picture of a person's functional status. Many test batteries are not suitable for active and healthy adults due to ceiling effects, or require a lot of space, time and training. The Community Balance and Mobility Scale (CBMS) is considered a gold standard for this population, but it is complex, time- and resource intensive. There is a strong need for a faster, yet sensitive and robust test of physical function in seniors. We sought to investigate whether an instrumented Timed Up and Go (iTUG) could predict the CBMS score in 60 outpatients and healthy community-dwelling seniors, which features of the iTUG were predictive, and how the prediction of CBMS with the iTUG compared to standard clinical tests. A partial least squares regression analysis was used to identify latent components explaining variation in CBMS total score. The model with iTUG features was able to predict the CBMS total score with an accuracy of 85.2% (84.9-85.5%), while standard clinical tests predicted 82.5% (82.2-82.8%) of the score. These findings suggest that a fast and easily administered iTUG could be used to predict CBMS score, providing a valuable tool for research and clinical care.

Keywords: iTUG; physical function; functional assessment; balance; mobility; older adults; partial least squares

1. Introduction

Physical function can be measured using self-report questionnaires or supervised clinical tests that quantify the observable ability to perform tasks, such as standing up from a chair, walking, turning, or standing on a differing base of support, such as a single leg stance [1]. Single tests can rarely capture multiple aspects of mobility, so a battery of tests is often administered to obtain a comprehensive picture of functional status. Many test batteries commonly used in geriatric testing and ageing research are not sensitive to change due to floor and ceiling effects [2, 3], or they require considerable assessor training.

The Community Balance and Mobility Scale (CBMS) has recently been identified as a valid, reliable and comprehensive performance-based assessment for measuring physical function in seniors [2, 4]. It contains a range of challenging balance and physical tasks. The scale is complex, time- and resource intensive. This limits its feasibility for use in large scale public health approaches [5], or in daily clinical practice. While difficult to administer, the CBMS can be considered a current gold standard for the measurement of simple to advanced balance tasks in seniors [2, 4, 6].

There is a need for quicker yet sensitive and robust measures to assess balance, strength and functional decline in young and older seniors. An alternative to developing new tests is to use instrumented versions of existing validated measures. In clinical environments and ageing research, three procedures are commonly used: the Short Physical Performance Battery [7], the Timed Up and

Go (TUG) [8] and the measure of gait speed [9]. The TUG consists of sit-to-stand transitions, walking and turning in one test that is deployable and scalable. The test is face valid to most populations and can be easily taught to health care professionals; it is widely recognized and quick to administer. The outcome measure of the TUG is the time taken to complete the whole task, measured in seconds. The 'score' does not discriminate fallers from non-fallers [3], identify frailty, or accurately predict falls in higher-functioning seniors [10]. Instrumenting the TUG (iTUG), using inertial sensor signals, allows measurement of spatial and temporal features from different segments of an iTUG trial, such as sit-to-stand transitions, walking, and turning. The iTUG has shown improved performance compared to the original TUG at assessing seniors at risk of falling or with Parkinson's disease, disability or cognitive impairment [10]. Here, we hypothesized that an iTUG performed several times back-to-back might be a comprehensive, robust, quick and feasible way to assess and extract advanced balance and mobility scores in seniors.

In order to obtain a quick and meaningful measure of functional decline in young and older seniors, we aimed to evaluate how well the averaged inertial sensor features from 5 iTUG repetitions can predict the CBMS total score within a group of geriatric outpatients and healthy community-dwelling seniors. Further, we sought to investigate whether the iTUG and which components of the iTUG can predict the CBMS total score accurately, compared to standard clinical tests used in routine assessments today.

2. Materials and Methods

2.1 Population

Sixty participants from two different cohorts (40 community dwelling healthy seniors and 20 geriatric patients from an outpatient clinic) in Stuttgart, Germany, were invited to participate in this cross-sectional method study. Participants were included if they were a) aged between 60 and 85 years and b) able to walk 30 meters independently. Exclusion criteria were any patient-reported cardiovascular, pulmonary, neurological, or mental diseases. The study was approved by the local medical ethical committee (Germany, no: 850/2018BO1), and adhered to the Declaration of Helsinki. All participants gave their written informed consent prior to inclusion.

2.2 Measurements

Demographic data and medical history were obtained from all participants. The following measurements were completed (in order): Late Life Function and Disability Index (LLFDI) [11], Montreal Cognitive Assessment (MoCA) [12] Short Falls Efficacy Scale-International (FES-I) [13], Eight-level balance scale (8-LBS) [14], Community Balance and Mobility Scale (CBMS) [15], 7-meters walk test (habitual and fast), 30-second chair-stand test (30-CST) [16], Short Physical Performance Battery (SPPB) [7]. A complete description of test administration and -outcomes can be found in Appendix Table A.

2.3 Procedures

Participants underwent assessments in a hospital gait lab, which were administered by trained research assistants (physiotherapists or sport scientists). The assessment battery consisted of self-reported and objectively measured tests of physical ability, including the iTUG. The test order was randomised, with participants starting with either the iTUG or traditional non-instrumented clinical tests. The entire assessment battery took on average 1.5 hours and participants could take breaks between tests when needed.

2.4 TUG and iTUG

The TUG was performed as five consecutive repetitions of the original TUG, with 30 seconds break between each repetition. We used a chair that was 46 centimeters high, with armrests. A cone was

placed at a mark 3 meters from the front legs of the chair. Instructions were given in accordance with those from the original TUG [8].

During the trials the participants wore a Huawei smartphone (Huawei Technologies Co, Ltd, Shenzhen, Guangdong, China) running a custom-made Android application, originally developed within the FARSEEING project [17]. The smartphone was worn on their lower back in a belt case. The assessor controlled a second smartphone which was connected via Bluetooth to the smartphone worn by the participant, to manually time each trial according to the original guidelines [8]. The set-up is illustrated in **Error! Reference source not found.**.

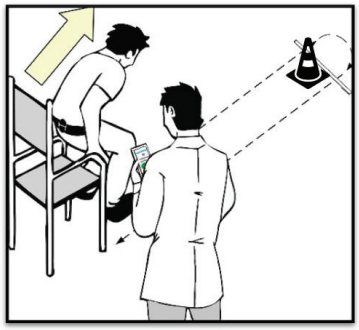

Figure 1. Illustration of the test set-up for the iTUG.

Sensor signals were recorded from the triaxial gyroscope and accelerometer embedded within the smartphone worn by the participants during the iTUG. The procedures have been described elsewhere [18], but in short we divided the iTUG into four segments: Sit-to-walk (StW), Walking (W), First turn (FT), and Turn-to-sit (TtS). Anterior-posterior (AP) acceleration and angular velocity around the medio-lateral (ML) axis were used to identify the sit-to-stand (StS) and walking segments. To identify the turning segments, we used the angular velocity around the vertical (V) axis. We computed 78 features from the inertial sensor signals (see Appendix Table A2), including segment durations, intensity measured as root mean square (RMS), and the smoothness of the signal measured as normalized jerk scores (NJS). Mean and maximum angular velocities were computed from the turning segments, as well as number of steps from the walking and turning segments.

2.5 Statistical Analysis

The descriptive information was calculated and reported as number of participants, mean age, height, weight, BMI, education, and gender distribution. Descriptive variables were included in the PLSR (partial least squares regression) analysis in both models (Figure 2).

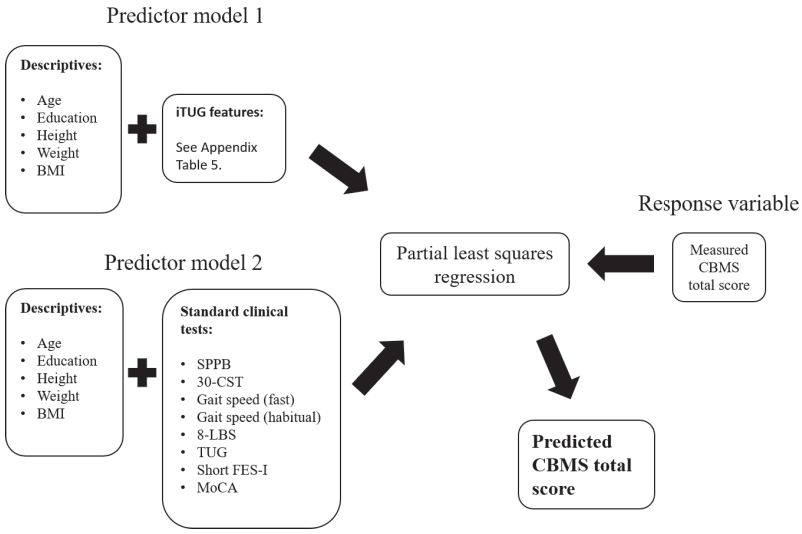

Figure 2. Illustration of predictor and response variables used as input for the two separate PLSR models.

The iTUG was performed with five repetitions to eliminate variance in performance across trials (16). The average value for each feature across the five repetitions was used for maximal robustness in the final model presented.

2.5.1 Partial least squares regression (PLSR)

To find the variables that most accurately described the variation in CBMS total score, we used a PLSR analysis. This analysis finds components that explain as much of the covariance as possible between the predictor variables *X* (i.e. the iTUG features or the standard clinical tests) and the response variable *Y* (CBMS total score).

The PLSR model was validated in a 7-step cross-validation procedure. The iTUG features that were significantly ($p<0.05$) correlated with the training data, *Xtrain* and *Ytrain*, were selected for the PLSR model. Data were then cross-validated to identify the robust components without overfitting the model, using a Monte Carlo simulation procedure with 100 repeated random iterations of repartitioning the *Xtrain* and *Ytrain* data. The optimal number of components (see Figure 1B and Figure **3**B) were chosen by calculating the root mean square error of prediction (RMSEP), by which we chose the number with the minimum RMSEP and RMSEP confidence interval for the PLSR. We then calculated the variable importance in projection (VIP), which is an accumulated measure of the importance of each variable from each component in the PLSR model. The most common VIP cut-off for variable selection is a VIP value of >1, but variables with VIP values between 0.83 and 1.21 are also used in some situations [19], hence we chose to illustrate how the variables selected in our PLSR method align with all these three cut-offs (Figure 2 and Figure **4**). Finally, Z-scores were obtained to analyze the statistical difference in RMSEP between the iTUG model and the model of standard clinical tests. Alpha was set to 0.05. All computations were performed in Matlab 2019b [20].

3. Results

Sixty participants were included in the study (mean age 74.2 years ± 7.6), 32 females (53.3%)). The characteristics and scores of physical function are presented in Table 1.

Table 1. Participants' characteristics. Mean and standard deviation (SD) for all variables except sex distribution.

	Community-dwellers (n=40)	Outpatients (n=20)
Age, in years	71.8 (7.3)	78.9 (5.9)
Sex (M/F)	17/23	11/9
Years of education	14.6 (3.7)	11.5 (4.3)
BMI (kg/m^2)	25.0 (3.6)	27.3 (4.3)
MoCA (0-30)	27.1 (4.9)	23.7 (2.2)
CBMS (0-96)	66.7 (18.3)	15.0 (17.2)
LLFDI (0-100)	75.5 (9.9)	51.3 (14.4)
TUG (s)	8.3 (1.2)	13.9 (4.0)
SPPB (0-12)	11.7 (0.9)	9.0 (2.4)
8-LBS (1-8)	5.4 (1.5)	4.3 (1.4)
30-CST (no. of repetitions)	15.3 (2.9)	8.8 (3.3)
Gait speed, habitual (m/s)	1.36 (0.20)	0.88 (0.21)
Gait speed, fast (m/s)	1.83 (0.27)	1.18 (0.30)
Short FES-I (7-28)	8.1 (1.5)	10.7 (4.0)

M, male; F, female; BMI, body mass index; MoCA, Montreal Cognitive Assessment; CBMS, Community Balance and Mobility Scale; LLFDI, Late Life Function and Disability Index; TUG, Timed Up and Go; SPPB, Short Physical Performance Battery; 8-LBS, Eight Level Balance Scale; 30-CST, 30-second Chair Stand; FES-I, Falls Efficacy Scale-International

3.1. PLSR of iTUG features versus the CBMS

Using the PLSR model with iTUG-features as predictors and CBMS total score as the response variable, we found that the first three components of iTUG features predicted the CBMS total score with an R^2 of 0.852 (95% CI 0.849-0.855, see Table 2).

Table 2. Loading scores, R^2 and VIP scores for all variables selected and R^2 of the first three components in the PLSR analysis for iTUG features and descriptive variables.

Variables selected by PLSR	Loading scores, component 1	Loading scores, component 2	Loading scores, component 3	VIP	R^2
iTUG features					
Mean velocity first turn	0.215	0.032	-0.091	1.123	0.739
Walk duration	-0.222	0.024	0.028	1.123	0.739
Mean velocity TtS	0.207	0.094	-0.069	1.124	0.722
Total duration	-0.218	0.030	0.015	1.104	0.706
Total number of steps	-0.213	-0.039	0.079	1.099	0.706
Peak velocity TtS	0.209	0.045	-0.035	1.092	0.700
Peak velocity first turn	0.207	-0.001	-0.056	1.074	0.674
Average step length	0.204	0.067	-0.054	1.064	0.654
TtS turning duration	-0.205	-0.032	0.065	1.055	0.647
Turn duration	-0.208	0.000	0.130	1.056	0.643
Gait speed	0.207	-0.009	0.009	1.055	0.634
Number of steps in first turn	-0.201	-0.020	0.200	1.032	0.601
NAJS first turn	-0.182	-0.121	0.205	1.004	0.560
RMS acc. walking V	0.196	-0.112	0.074	0.991	0.548

NAJS TtS	-0.173	-0.177	0.130	0.985	0.523
RMS acc. walking AP	0.191	-0.266	0.080	1.014	0.449
RMS angular velocity walking V	0.185	-0.188	0.012	0.953	0.443
Range angular velocity walking V	0.181	-0.182	0.042	0.929	0.433
TtS duration	-0.173	0.061	0.091	0.887	0.419
Jerk score walking AP	0.174	-0.137	-0.048	0.884	0.416
StW duration	-0.140	-0.076	-0.203	0.891	0.373
Step regularity V	0.138	0.163	-0.070	0.899	0.366
Jerk score walking V	0.143	0.073	-0.047	0.785	0.338
Descriptive					
Age	-0.129	-0.401	-0.327	1.404	0.472
Education	0.113	0.496	0.047	1.188	0.352
	Component 1	**Component 2**	**Component 3**		**Total**
Mean explained variation (R^2)	0.771	0.058	0.023		0.852
95% CI	0.769-0.772	0.054-0.061	0.020-0.027		0.849-0.855

The RMSEP (see Figure 3) was found to be lowest with 3 and 4 components, where 4 was slightly lower than 3 (11.79 vs 11.81), albeit not significantly so (p=0.9).

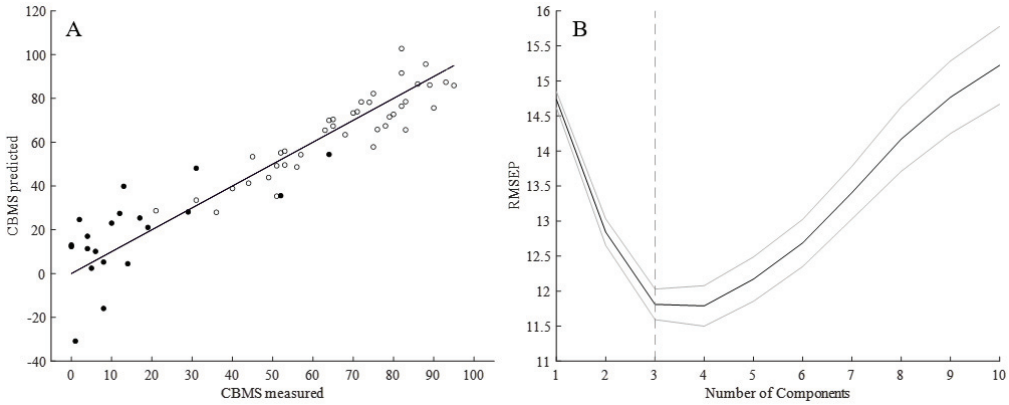

Figure 1. (**a**) Predicted vs. measured CBMS total score from iTUG. Outpatients presented by dots and community-dwellers by circles; (**b**) RMSEP (black line) with +/- one standard deviation (grey lines) across 10 components.

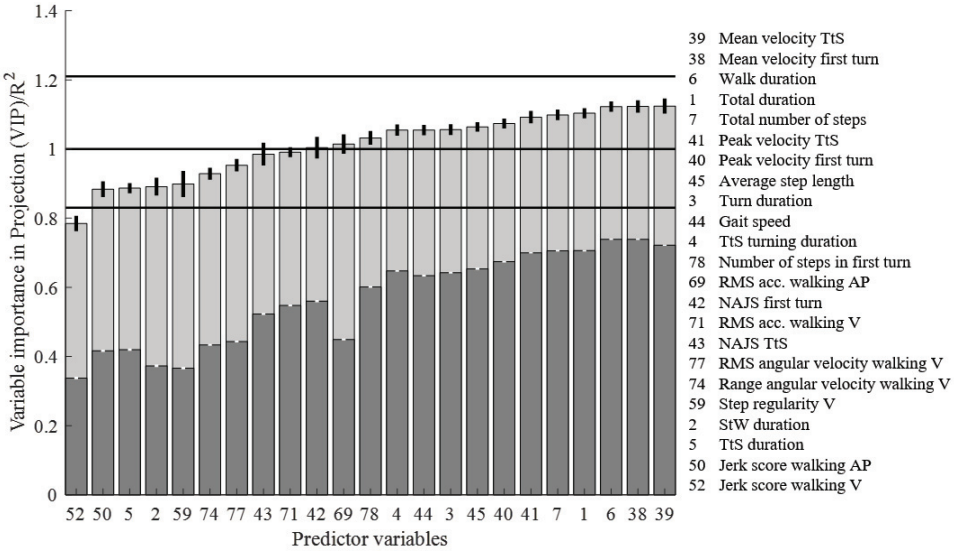

Figure 2. The VIP scores (light grey) and R^2 (dark grey) of the iTUG features selected in the PLSR model. The horizontal lines represent the lower (0.83), middle (1) and upper (1.21) cut-off values used for interpreting the VIP of individual predictor variables. TtS, Turn to sit; RMS, Root mean square; acc, acceleration; AP, Anterior-posterior; NAJS, Normalised angular jerk score; V, vertical axis; StW, Sit to walk.

3.2. PLSR of standard clinical tests vs. the CBMS

In the PLSR model with standard clinical test scores as predictors and CBMS total score as the response variable, we found that the first two components predicted the CBMS total score with an R^2 of 0.825 (95% CI 0.822-0.828, see **Error! Reference source not found.**).

Table 3. Loading scores, R2 and VIP scores for all variables selected and R2 of the first two components in the PLSR analysis for standard clinical tests and descriptive variables.

Variables selected by PLSR	Loading scores, component 1	Loading scores, component 2	VIP	R^2
Clinical				
TUG	-0.398	0.102	1.133	0.698
Gait speed, fast	0.385	-0.057	1.115	0.676
Gait speed, habitual	0.381	-0.034	1.113	0.672
SPPB	0.371	-0.260	1.010	0.533
30-CST	0.346	-0.320	0.985	0.504
8-LBS	0.257	0.215	0.814	0.335
Short FES-I	-0.314	0.451	0.855	0.307
Descriptive				
Age	-0.250	-0.728	1.065	0.472
Education	0.270	0.331	0.843	0.352

	Component 1	Component 2	Total
Mean explained variation (R^2)	0.798	0.027	0.825
95% CI	0.796-0.801	0.024-0.029	0.822-0.828

The RMSEP was not significantly lower with more than two components, and mean RMSEP with two components was 12.85 (Figure 5). RMSEP of CBMS prediction with the iTUG model was significantly lower than with standard clinical tests (Figure 3B and 5B, p=<0.0001).

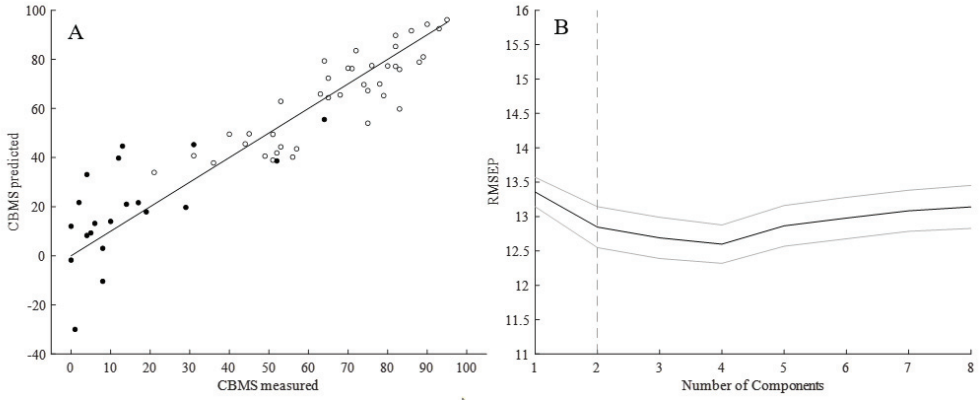

Figure 3. (**a**) Predicted vs. measured CBMS total score from clinical variables. Outpatients presented by dots and community-dwellers by circles; (**b**) RMSEP (black line) with +/- one standard deviation (grey lines) across 8 components.

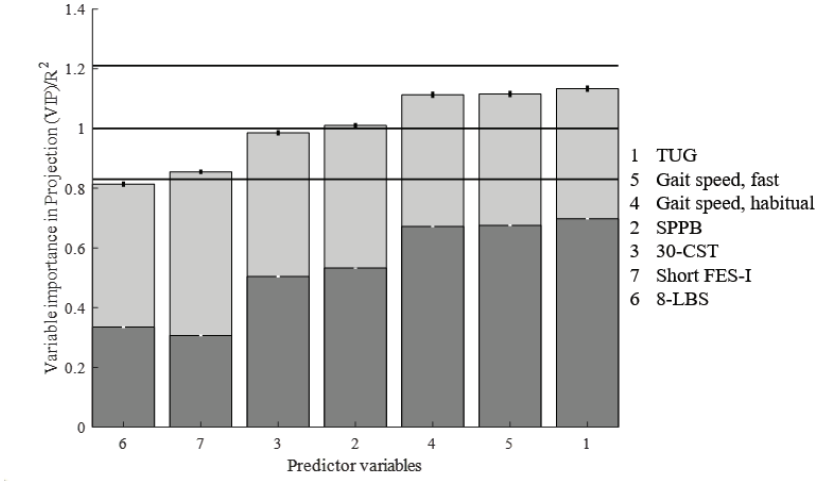

Figure 4. The VIP scores (light grey) and R2 (dark grey) of the clinical variables selected in the PLSR model. The horizontal lines represent the lower (0.83), middle (1) and upper (1.21) cut-off values used for interpreting the VIP of individual predictor variables.

4. Discussion

This study aimed to assess whether signal features averaged from 5 iTUG trials could predict CBMS scores as a ground truth model in community-dwelling seniors and geriatric outpatients, using a PLSR analysis. In addition, this study sought to investigate whether the predictive ability of iTUG was superior to standard clinical tests in predicting CBMS scores.

The PLSR model of iTUG features could predict the CBMS score with a substantial level of predictive accuracy (mean explained variation of 85.2%) [21]. The iTUG model had a similar predictive ability of CBMS scores as a battery of clinical tests, and significantly less error of prediction.

CBMS evaluates high-level gait, balance and mobility, required for safe and independent living in the community [4, 6]. Our findings suggest that a 5-times repeated iTUG, which requires little floor space, a smartphone, and about 5 minutes in the clinic or lab, can accurately predict a person's score on the CBMS, which otherwise would require 20-25 min and larger facilities and resources to administer. While the standard clinical tests were also able to predict the CBMS score with high accuracy, the test battery took approximately 35 minutes to administer, and staff needed specific to be able to collect those data. For iTUG, testing can be completed within 5 minutes and minimal training is required to administer the test.

The signal features with highest loading scores on the first component are features that represent several different segments of the iTUG, including walking, turning and turn-to-sit. We also found that they represent different units, such as velocity, duration, number of steps, and step length. These findings indicate that no specific signal features stand out from the others in terms of how much of the variation in CBMS they describe, but rather that a good prediction of CBMS relies on several complementary pieces of information. However, six of the ten features with highest R^2 scores were features obtained from the two turning phases of the iTUG, perhaps not coincidental, as the importance of turning for predicting balance have been recognized in several other studies. For example, an earlier study on older adults found that those who had poorer scores on the Berg Balance Scale and the Fullerton Advanced Balance scale, exhibited slower turns in the iTUG [22]. In a study of high-functioning young seniors, the features 'Walk duration' and 'TtS maximum velocity' both had significant discriminative ability on self-reported physical function as measured by the LLFDI [18]. Turning features of the iTUG have also been found to be sensitive for testing people with impaired motor control due to neurological conditions [23-25], fallers [26], and in persons with mild cognitive impairment [27], which could also be explained by motor control impairment [27, 28].

4.1 Limitations

We acknowledge that this study has some limitations. First, the sample size was relatively small. It is generally recognized that machine learning-based prediction models trained on small sample-sizes are vulnerable to biased performance estimates [29]. This study was intended as a pilot study, and a larger study with additional or larger cohorts is necessary to confirm the findings. In addition, the model described here has not been validated on an external dataset in which the same procedures have been applied. Therefore, the presented results need to be interpreted with caution and cannot be generalised to other populations without further confirmation of this work.

4.2 Implications for clinical practice and future research

We found that five trials of iTUG, which require very little time, space and training to administer, could predict the CBMS with a substantial and slightly higher accuracy than a battery of standard clinical tests. The potential implications of these findings are that the use of instrumented tests would save time for the individual and clinicians, and avoid fatigue in the patients related to comprehensive test batteries. Furthermore, self-administrable iTUGs are on the rise, which would allow seniors to assess their ability from the comfort of their own living room [30]. The adoption of electronic technologies has been recognized as a key strategy for cutting costs in healthcare [31], and with the iTUG, healthy seniors as well as patients could use their own smartphone and have their physical function monitored remotely by clinicians or researchers. Furthermore, access to physical activity levels

objectively measured by the smartphone could increase the predictive ability of not only CBMS, but other outcomes as well.

In future work, the iTUG PLSR model should be trained on a larger dataset and validated externally in new, unseen data collected using the same procedures as the training data. The same procedure used in the current analysis could also be used to evaluate how well the iTUG can predict relevant outcomes for other populations with impaired physical function, such as Parkinson's disease, multiple sclerosis, chronic obstructive pulmonary disease, stroke, and others.

5. Conclusions

In this study, we demonstrated that averaged signal features from a smartphone worn during a 5-times repeated iTUG could predict the CBMS score in community-dwellers and outpatients with 85.2% accuracy, while more elaborate standard clinical tests could predict it with 82.5% accuracy. The results suggest that an iTUG, which is potentially cost-saving, fast and easy to administer, may be used to predict a person's score on the CBMS in research, clinical care and from remote.

Author Contributions: Conceptualization, C.B., R.B., A.S.M., and K.T.; methodology, R.B., C.B., E.A.I., K.T., S.M., A.S.M.; software, E.A.I., S.M., R.B.; formal analysis, R.B., E.A.I., S.M.; investigation, C.N., A.S.M., R.B.; resources, C.B.; data curation, R.B., S.M., C.N.; writing—original draft preparation, R.B., A.S.M., C.N., and K.T.; writing—review and editing, R.B., A.S.M., K.T., C.B., C.N., B.V., J.L.H., S.M., E.A.I.; visualization, R.B., E.A.I.; supervision, A.S.M., K.T., C.B.; project administration, A.S.M., C.N., R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by PreventIT, which received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreement No. 689238. The content is solely the responsibility of the authors and does not necessarily represent the official views of the European Commission. The European Commission had no role in the design, execution, interpretation, or writing of the study.

Acknowledgments: The authors would like to thank Christoph Endress, Rebekka Leonhardt and Antonia Stolz for carrying out the clinical assessments, and Lindsay Wu for proof-reading the manuscript. Thanks also to the participants who volunteered in this study.

Conflicts of Interest: R.B., C.N., K.T., E.A.I., B.V., J.L.H., C.B. and A.S.M. declare no conflict of interest. S.M. holds a share of the mHealth Technologies srl company, which owns the rights to some of the movement analysis algorithms.

Appendix A

Table A1. Complete list of standard clinical tests with description.

Late Life Function and Disability Index (LLFDI) The LLFDI [11] consists of two parts; function, and disability with 32 and 16 items respectively. Both parts of the LLFDI were administered, but for the purpose of this study, only the functional scores will be used in the analysis. The items are regarding how much difficulty the participant experiences with carrying out different activities of daily life with a rating scale of 1-5, ranging from no difficulty to cannot do. The questions span across three dimensions; upper extremity, lower extremity, and advanced lower extremity. The total score is scaled, resulting in scores ranging from 0-100 (higher score indicating better performance), allowing comparison to other trials and cohorts.
Short Falls-Efficacy Scale International (Short FES-I) The Short FES-I (9) is a 10-item questionnaire developed to assess the fear of falling in community-dwelling older adults. The outcome is a sum score ranging from minimum 7 (no concern about falling) to maximum 28 (severe concern about falling).
Montreal Cognitive Assessment (MoCA)

MoCA [12] is a screening tool used identify mild cognitive impairment (MCI) that assesses short term memory, visuospatial abilities, executive functions, attention, concentration and working memory, language, and orientation to time and place. The score ranges from 0 to a maximal score of 30, with a higher score indicating better cognitive function.

Eight-level balance scale (8-LBS)

The 8-LBS (11) is a test of static balance in which the participants attempt increasingly difficult positions for 15 seconds. The test ends when (if) the participants are not able to hold the position for 15 seconds. The positions are 1) side-by-side standing, narrow base, eyes open; 2) side-by-side standing, narrow base, eyes closed; 3) semi-tandem, eyes open; 4) tandem, eyes open; 5) tandem, eyes closed; 6) one-leg standing, eyes open; 7) one-leg standing, eyes closed; 8) one-leg stand, eyes closed + cognitive distraction (mentioning of the months of the year in a backwards order). The outcome is the number (in order) of the most difficult position attempted, ranging from 1 (least difficult) to 8 (most difficult).

Community Balance and Mobility Scale (CBMS)

The CBMS (14) is a test battery of balance and mobility consisting of 13 tasks, of which six are assessed unilaterally. It has been shown to be a promising performance-based test of physical function in higher-functioning seniors (2, 6). Each task is rated at the assessor's discretion, and a score given from 0 (unable) to 5 (coordinated and controlled, without excessive equilibrium reactions). The scores are summed and the total score ranges from 0 to 96, where a higher score indicates better performance. The bonus point (95 +1) is given if the participants can descend a staircase while holding a weighted basket in front of them, allowed only intermittently to look at the steps.

7-meters walk test (habitual and fast)

Participants are timed over 7 meters within a 9-meter track, allowing one meter in each end for acceleration/deceleration. The best time from two trials in both habitual and fast walking conditions were used to calculate respective gait speeds (m/s).

30-second chair-stand test (30-CST)

In the 30-CST the assessor counts the number of repetitions of sit-to-stands the participants can perform in 30 seconds. The test was developed to overcome the floor-effect associated with the Five times sit-to-stand, and is originally a part of the Fullerton Functional Fitness Test battery (12).

Short Physical Performance Battery (SPPB)

SPPB is a test of physical functioning of the lower extremities in older adults (13). The test consists of three parts, where the participants 1) attempts to keep their balance in three different feet-positions for 10 seconds in each, 2) walk four meters in habitual pace (performed twice), and 3) perform five repeated sit-to-stands as fast as possible. Each part is scored, and a combined score from 0-12 is given, where a higher score indicates better performance.

Table A2. Complete list of extracted iTUG features used in the PLSR analysis.

Total Duration	RMS ML gyro StW	Peak Velocity TtS	Stride Regularity ML
StW Duration	RMS V gyro StW	NAJS 180° Turn	Stride Regularity V
180° Turn Duration	range AP acc TtS	NAJS TtS	Gait Symmetry AP
TtS Turning Duration	range ML acc TtS	Gait Speed	Gait Symmetry ML
TtS Duration	range V acc TtS	Average Step Length	Gait Symmetry V
Walk Duration	RMS AP acc TtS	Step Duration	Range Acceleration Walking AP

Total Number of Steps	RMS ML acc TtS	Standard Dev. of Step Duration	Range Acceleration Walking ML
range AP acc StW	RMS V acc TtS	Coef. Variation of Step Duration	Range Acceleration Walking V
range ML acc StW	Jerk Score AP TtS	Coordination Index	RMS Acceleration Walking AP
range V acc StW	Jerk Score ML TtS	Jerk Score Walking AP	RMS Acceleration Walking ML
RMS AP acc StW	Jerk Score V TtS	Jerk Score Walking ML	RMS Acceleration Walking V
RMS ML acc StW	range AP gyro TtS	Jerk Score Walking V	Range Angular Velocity Walking AP
RMS V acc StW	range ML gyro TtS	Normalised Jerk Score Walking AP	Range Angular Velocity Walking ML
Jerk Score AP StW	range V gyro TtS	Harmonic Ration AP	Range Angular Velocity Walking V
Jerk Score ML StW	RMS AP gyro TtS	Harmonic Ration ML	RMS Angular Velocity Walking AP
Jerk Score V StW	RMS ML gyro TtS	Harmonic Ration V	RMS Angular Velocity Walking ML
range AP gyro StW	RMS V gyro TtS	Step Regularity AP	RMS Angular Velocity Walking V
range ML gyro StW	Mean Velocity 180° Turn	Step Regularity ML	Number of Steps in 180° Turn
range V gyro StW	Mean Velocity TtS	Step Regularity V	
RMS AP gyro StW	Peak Velocity 180° Turn	Stride Regularity AP	

References

1. Freiberger, E.; De Vreede, P.; Schoene, D.; Rydwik, E.; Mueller, V.; Frändin, K.; Hopman-Rock, M., Performance-based physical function in older community-dwelling persons: a systematic review of instruments. *Age Ageing* **2012,** *41*, (6), 712-721.
2. Bergquist, R.; Weber, M.; Schwenk, M.; Ulseth, S.; Helbostad, J. L.; Vereijken, B.; Taraldsen, K., Performance-based clinical tests of balance and muscle strength used in young seniors: a systematic literature review. *BMC Geriatr.* **2019,** *19*, (1), 9.
3. Schoene, D.; Wu, S. M.; Mikolaizak, A. S.; Menant, J. C.; Smith, S. T.; Delbaere, K.; Lord, S. R., Discriminative ability and predictive validity of the timed up and go test in identifying older people who fall: systematic review and meta-analysis. *J. Am. Geriatr. Soc.* **2013,** *61*, (2), 202-8.
4. Weber, M.; Van Ancum, J.; Bergquist, R.; Taraldsen, K.; Gordt, K.; Mikolaizak, A. S.; Nerz, C.; Pijnappels, M.; Jonkman, N. H.; Maier, A. B.; Helbostad, J. L.; Vereijken, B.; Becker, C.; Schwenk, M., Concurrent validity and reliability of the Community Balance and Mobility scale in young-older adults. *BMC Geriatr.* **2018,** *18*, (1), 156.
5. Gordt, K.; Mikolaizak, A. S.; Taraldsen, K.; Bergquist, R.; Van Ancum, J. M.; Nerz, C.; Pijnappels, M.; Maier, A. B.; Helbostad, J. L.; Vereijken, B.; Becker, C.; Schwenk, M., Creating and Validating a Shortened

Version of the Community Balance and Mobility Scale for Application in People Who Are 61 to 70 Years of Age. *Phys. Ther.* **2020,** *100*, (1), 180-191.
6. Balasubramanian, C. K., The community balance and mobility scale alleviates the ceiling effects observed in the currently used gait and balance assessments for the community-dwelling older adults. *J. Geriatr. Phys. Ther.* **2015,** *38*, (2), 78-89.
7. Guralnik, J. M.; Simonsick, E. M.; Ferrucci, L.; Glynn, R. J.; Berkman, L. F.; Blazer, D. G.; Scherr, P. A.; Wallace, R. B., A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J. Gerontol.* **1994,** *49*, (2), M85-94.
8. Podsiadlo, D.; Richardson, S., The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J. Am. Geriatr. Soc.* **1991,** *39*, (2), 142-148.
9. Studenski, S.; Perera, S.; Patel, K.; Rosano, C.; Faulkner, K.; Inzitari, M.; Brach, J.; Chandler, J.; Cawthon, P.; Connor, E. B.; Nevitt, M.; Visser, M.; Kritchevsky, S.; Badinelli, S.; Harris, T.; Newman, A. B.; Cauley, J.; Ferrucci, L.; Guralnik, J., Gait Speed and Survival in Older Adults. *JAMA* **2011,** *305*, (1), 50-58.
10. Sprint, G.; Cook, D. J.; Weeks, D. L., Toward Automating Clinical Assessments: A Survey of the Timed Up and Go. *IEEE Rev. Biomed. Eng.* **2015,** *8*, 64-77.
11. Haley, S. M.; Jette, A. M.; Coster, W. J.; Kooyoomjian, J. T.; Levenson, S.; Heeren, T.; Ashba, J., Late Life Function and Disability Instrument: II. Development and evaluation of the function component. *J. Gerontol. A Biol. Sci. Med. Sci.* **2002,** *57*, (4), M217-22.
12. Nasreddine, Z. S.; Phillips, N. A.; Bédirian, V.; Charbonneau, S.; Whitehead, V.; Collin, I.; Cummings, J. L.; Chertkow, H., The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **2005,** *53*, (4), 695-699.
13. Kempen, G. I.; Yardley, L.; van Haastregt, J. C.; Zijlstra, G. A.; Beyer, N.; Hauer, K.; Todd, C., The Short FES-I: a shortened version of the falls efficacy scale-international to assess fear of falling. *Age Ageing* **2008,** *37*, (1), 45-50.
14. Clemson, L.; Fiatarone Singh, M. A.; Bundy, A.; Cumming, R. G.; Manollaras, K.; O'Loughlin, P.; Black, D., Integration of balance and strength training into daily life activity to reduce rate of falls in older people (the LiFE study): randomised parallel trial. *BMJ.* **2012,** *345*, e4547.
15. Howe, J. A.; Inness, E. L.; Venturini, A.; Williams, J. I.; Verrier, M. C., The Community Balance and Mobility Scale--a balance measure for individuals with traumatic brain injury. *Clin. Rehabil.* **2006,** *20*, (10), 885-95.
16. Rikli, R. E.; Jones, C. J., Development and validation of a functional fitness test for community-residing older adults. *J. Aging Phys. Act.* **1999,** *7*, (2), 129-161.
17. Mellone, S.; Tacconi, C.; Schwickert, L.; Klenk, J.; Becker, C.; Chiari, L., Smartphone-based solutions for fall detection and prevention: the FARSEEING approach. *Z. Gerontol. Geriatr.* **2012,** *45*, (8), 722-727.
18. Coni, A.; Van Ancum, J. M.; Bergquist, R.; Mikolaizak, A. S.; Mellone, S.; Chiari, L.; Maier, A. B.; Pijnappels, M., Comparison of Standard Clinical and Instrumented Physical Performance Tests in Discriminating Functional Status of High-Functioning People Aged 61–70 Years Old. *Sensors* **2019,** *19*, (3), 449.
19. Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbø, S., A review of variable selection methods in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems* **2012,** *118*, 62-69.
20. Matlab, *9.7.0.1190202 (R2019b)*. The MathWorks Inc.: Natick, Massachusetts, 2019.
21. Hair, J. F.; Ringle, C. M.; Sarstedt, M., PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice* **2011,** *19*, (2), 139-152.
22. Chen, T.; Chou, L.-S., Effects of Muscle Strength and Balance Control on Sit-to-Walk and Turn Durations in the Timed Up and Go Test. *Arch. Phys. Med. Rehabil.* **2017,** *98*, (12), 2471-2476.
23. Caronni, A.; Sterpi, I.; Antoniotti, P.; Aristidou, E.; Nicolaci, F.; Picardi, M.; Pintavalle, G.; Redaelli, V.; Achille, G.; Sciumè, L.; Corbo, M., Criterion validity of the instrumented Timed Up and Go test: A partial least square regression study. *Gait Posture* **2018,** *61*, 287-293.
24. Palmerini, L.; Mellone, S.; Avanzolini, G.; Valzania, F.; Chiari, L., Quantification of motor impairment in Parkinson's disease using an instrumented timed up and go test. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2013,** *21*, (4), 664-673.
25. Salarian, A.; Horak, F. B.; Zampieri, C.; Carlson-Kuhta, P.; Nutt, J. G.; Aminian, K., iTUG, a sensitive and reliable measure of mobility. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2010,** *18*, (3), 303-10.
26. Greene, B. R.; O'Donovan, A.; Romero-Ortuno, R.; Cogan, L.; Scanaill, C. N.; Kenny, R. A., Quantitative Falls Risk Assessment Using the Timed Up and Go Test. *IEEE Trans. Biomed. Eng.* **2010,** *57*, (12), 2918-2926.

ISBN 978-82-326-5048-4 (printed ver.)
ISBN 978-82-326-5049-1 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

27. Mirelman, A.; Weiss, A.; Buchman, A. S.; Bennett, D. A.; Giladi, N.; Hausdorff, J. M., Association between performance on Timed Up and Go subtasks and mild cognitive impairment: further insights into the links between cognitive and motor function. *J. Am. Geriatr. Soc.* **2014,** *62*, (4), 673-678.
28. Buchman, A. S.; Boyle, P. A.; Leurgans, S. E.; Barnes, L. L.; Bennett, D. A., Cognitive function is associated with the development of mobility impairments in community-dwelling elders. *Am. J. Geriatr. Psychiatry* **2011,** *19*, (6), 571-80.
29. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A. J., Machine learning algorithm validation with a limited sample size. *PLoS One* **2019,** *14*, (11), e0224365-e0224365.
30. Bergquist, R.; Vereijken, B.; Mellone, S.; Corzani, M.; Helbostad, J. L.; Taraldsen, K., App-based Self-administrable Clinical Tests of Physical Function: Development and Usability Study. *JMIR Mhealth Uhealth* **2020,** *8*, (4), e16507.
31. Kvedar, J.; Coye, M. J.; Everett, W., Connected Health: A Review Of Technologies And Strategies To Improve Patient Care With Telemedicine And Telehealth. *Health Aff. (Millwood)* **2014,** *33*, (2), 194-199.