Original research article

DIGITAL HEALTH

DIGITAL HEALTH
Volume 12: 1–21
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076261481297
journals.sagepub.com/home/dhj
Sage

Development and evaluation of a digitally based functional fitness self-assessment and monitoring tool for older adults

Jason Tallis[1], Darren Richardson[1], Sharn Shelley[1], Neil Clarke[2], David Broom[3], and Michael Duncan[4]

Abstract

Background: Maintaining physical function is central to healthy ageing and is emphasised in health policy aimed at shifting care from treatment to prevention. However, accessible tools that support routine monitoring of functional fitness remain limited despite evidence demonstrating links to adverse health outcomes. A Functional Fitness Monitoring Tool (FFMT) was developed as a digital, self-administered platform enabling older adults and healthcare professionals to assess and track physical function. The study aimed to; 1) evaluate perceived importance and usability; 2) compare between-session reliability of self-administered (SL) versus instructor-administered (IL) assessments. **Method:** Thirty-three community-dwelling older adults (68 ± 6 years) completed three SL assessments followed by two IL assessments, or vice versa. The FFMT incorporated nine validated functional tests with digital instructions, demonstration videos, and a results-monitoring interface. After all assessments, participants completed an online usability survey. **Results:** Between-session reliability across SL and IL assessments was good to excellent (ICC > 0. 816). Reliability remained across groups, though wider confidence intervals appeared for initial SL trials in some tests. Thresholds for measurable change were small. Participants found the FFMT easy to use, interpret, and valuable for long-term monitoring. **Conclusion:** The FFMT demonstrates high reliability and usability, with self-administered assessments performing comparably to instructor-led administration. These findings support the FFMT as an accessible, scalable tool capable of enabling accurate self-monitoring of functional fitness across clinical, community, and population settings, with potential to facilitate early detection of functional decline and promote prevention-focused healthy ageing.

Keywords

ageing, health promotion, physical function, muscle function, digital health

Received: 16 March 2026; Revised: 9 July 2026; Accepted: 11 August 2026

Introduction

There is global growth in the number and proportion of older adults living in the population,[1] and although people are living longer, increasing age is associated with a growing number of comorbidities.[2] Gains in health expectancy are therefore failing

[1]Research Centre for Discoveries in Life Science, Coventry University, Coventry, UK
[2]Research Centre for Life and Sport Science (CLaSS), School of Health Sciences, Birmingham City University, Birmingham, UK
[3]Institute for Health and Care Improvement, York St John University, York, UK
[4]Research Centre for Healthcare and Community Transformation, Coventry University, Coventry, UK

Corresponding author:
Jason Tallis, Research Centre for Discoveries in Life Science, Coventry University, Priory Street, Coventry CV1 5FB, UK.
Email: ab0289@coventry.ac.uk

to match gains in life expectancy.[3] Increasing age not only influences health, wellbeing and quality of life, it can also be associated with increased economic and societal costs arising from functional decline, multimorbidity, and the need for greater health and social care support.[4] Physical activity (PA) has a pivotal role in promoting successful ageing,[5] defined as multidimensional construct characterised by the maintenance of physical, cognitive, psychological, and social functioning, supporting continued independence, participation in meaningful activities, and overall wellbeing throughout later life.[6] PA for older adult populations has been firmly established to improve cardiovascular and muscle function, psychological health and wellbeing, and reduce prevalence or severity of disease.[7] Recent large-scale evidence supports the multidimensional health value of PA, demonstrating that older adults who engage in PA show better physical and cognitive functions.[8] As such, PA promotion is a primary focus of government strategy to improve older adult mental and physical health,[9] yet despite the well-known benefits, adults over the age of 75 years represent the least physically active population in the United Kingdom (UK).[10]

The basis for successful PA interventions that older adults will adhere to is valid, reliable, and meaningful evaluation of composites of healthy ageing, where assessment forms the basis of targeted intervention strategies, goal setting and provides opportunity for monitoring of intervention success. Within a research context, PA intervention success is typically evaluated by measuring changes in physical abilities and functional fitness which match tasks and activities of daily living (ADL),[11–15] both of which are prerequisites for successful ageing. Despite this, assessment of physical function is not commonplace in practice for monitoring and assessing successful ageing. Such evaluation may provide a more accurate assessment of population healthy ageing than PA alone, given the instability in PA as a construct which has been shown to fluctuate substantially over relatively short time periods.[16] More specifically, poor performance in functional fitness assessments is linked to increased risk of non-communicable disease, falls and premature mortality,[17–19] thus the value of assessment is without question. Despite the existence of several well-established functional fitness assessment batteries for older adults, these are mostly research oriented tools, which aim to identify severe ageing effects rather than healthy ageing trajectories (i.e. the assessment approach adopted by the European Working Group On Sarcopenia[20]) and are not easily accessible to older adults and healthcare professionals. Moreover, there is no evidence based objective assessment of functional fitness that is widely available for older adults to undertake on their own. This is a particular barrier to promoting healthy ageing as valid and reliable self-assessment of physical function would enable larger scale public health risk identification, intervention effectiveness and surveillance in this population.

Digital tools have wide-reaching potential for understanding and promoting healthy behaviours, but like many complex health promoting interventions should be developed in the context of appropriate theoretical frameworks, such as behaviour change theory, as recommended by the Medical Research Council.[21] With reference to the transtheoretical model,[22] developing a digital tool that provides older adults with the opportunity to evaluate and contextualise their own functional fitness may provide the impetus for increased PA that is targeted to their health need, with progress tracking vital for maintaining lifelong healthy habits. The Behaviour Change Wheel (BCW) is a commonly applied conceptual framework which has been developed from 19 behaviour change theories and considers interacting systems that influence behaviour change.[23] The BCW has been mapped to the Behaviour Change Technique Taxonomy (BCTT),[24] developed to provide an inventory of behaviour change strategies. With reference to the BCW and the BCTT, developing a digital tool that can provide older adults and healthcare professionals with an ability to understand, assess and monitor physical function could incite a change to positive health behaviours by facilitating goal setting and planning, providing opportunity for comparison, shaping knowledge, providing incentive, and by allowing feedback on the impact of behaviour change.

As such, our research group have developed a prototype of a digital functional fitness self-assessment and monitoring tool (FFMT) for older adults, which is the first of its kind allowing for accessible and age-appropriate monitoring of physical function for self-assessment or use in a variety of fitness and health care settings. Digital health tools have shown generally favourable effects in terms of costs and health outcomes,[25] and the development and application of digital tools has UK government support (e.g. NHS Weight Loss and NHS Couch to 5k apps[26,27]). By harnessing the wide-reaching potential of digital tools for health promotion, our digital functional fitness self-assessment monitoring tool for older adults has been designed based on well-established functional fitness and body composition assessments. The tool provides an opportunity to benchmark performance against fitness standards and evidence-based thresholds used to indicate and predict negative risks to health and wellbeing associated with ageing. As such, this tool has the potential to identify individuals at elevated risk of accelerated functional decline, loss of independence, and adverse health outcomes, thereby providing a platform for timely intervention. Furthermore, the ability to develop an assessment history would allow the FFMT to be used as a benchmarking tool to monitor healthy lifestyle changes and progress during rehabilitation over time.

Development of our digital FFMT can improve the health of older adults and will impact several transdisciplinary fields and stakeholders. Importantly, our FFMT promotes healthy ageing, improves the quality of life of our growing older adult population which can ultimately reduce the economic and societal impact associated with preventable functional decline, frailty, and loss of independence in later life. For healthcare professionals, the FFMT can provide a platform to promote self-monitoring of positive health outcomes and facilitate the early identification of adverse ageing trajectories characterised by declining physical function. This may support earlier intervention, aid disease detection, and enhance patient self-management following hospitalisation and during rehabilitation from musculoskeletal injury. This potential directly relates to the 'from sickness to prevention' and 'from hospital to community' pillars of the NHS Fit for the future 10 year health plan.[28] Furthermore, it aligns with the wider healthcare agenda of empowering individuals to take a more active role in managing their own health, shifting elements of monitoring and preventive care from healthcare institutions into community and home environments through accessible self-monitoring technologies. The wide reach of digital tools may further provide opportunity for population level surveillance aiding in the development of sex, age and ethnic specific normative data, understanding regional, social-economic status and seasonal trends, and in identifying trends specific to disease models. Such information would be of significance to the research community and policy makers promoting the design and implication of targeted intervention.

Importantly, an effective assessment of physical function should demonstrate a high degree of between session reliability in the target population to be able to effectively measure change over time and identify suitable change thresholds. As such, a foundational step in the development of the FFMT was to evaluate both the usability of the tool and the between session reliability of the assessments incorporated. Importantly, the between session reliability of self-administered versions of the functional fitness assessments included in the FFMT is yet to be established and may differ to that when administered by an experienced instructor. Therefore, within a population of community dwelling older adults, this study aimed to: 1) evaluate the perceived importance and usability of the FFMT; 2) To evaluate and compare the between-session reliability of self-administered versus instructor-administered FFMT assessments, considering the influence of assessment order.

Methods

Participants

Following ethics approval from the host institution [Ref: P138035] and written informed consent, 36 participants were recruited through opportunistic sampling and engagement with local social groups specifically for older adults. An a priori sample size calculation was conducted using the ICC-based reliability study calculator (wnarifin.github.io), which implements the methodology proposed by Walter et al.[29] Although five repeated assessments were conducted, aspects of analysis examined reliability between two distinct sessions and this information was used in the sample size calculation given the sample size estimation will reduce with increased number of rerepeated measures. Assuming a minimum acceptable ICC of 0.70, an expected ICC of 0.90, a significance level of $\alpha = 0.05$, statistical power of 80%, a minimum sample size of 23 participants was required.

Participants were randomly assigned to one of two groups who completed the planned assessments in one of two distinct orders. The FFMT was designed specifically for adults aged 60 yrs and over and to be inclusive of fitness capabilities and functional ability. However, participants were screened using the Physical Activity Readiness Questionnaire + (PARQ+)[30] and were excluded if they had been diagnosed with a chronic medical condition that prevented safe completion of the assessments or had a musculoskeletal injury/condition that had either not been fully rehabilitated or would be aggravated by the physical assessments. Three participants dropped out due to inability to complete all the assessments in the required timeframe (N=2) or illness (N=1). As such, the final study sample included 33 participants (14 females; Mean±SD: 68.8±5.8 yrs.; Range: 60-81 yrs.) with N=17 in Group 1 (Mean±SD: 68.4±5.4 yrs.; Range: 60-78 yrs.) and N=16 in Group 2 (Mean±SD: 69.3±6.3 yrs.; Range: 60-81 yrs.).

This study employed a repeated-measures, randomised crossover design to evaluate the between-session reliability and usability of the FFMT. Participants completed five trials of the assessments contained in the FFMT following completion of an online survey to understand functionality and useability. All assessments were completed within three weeks with a minimum of 48 hours between trials. As outlined in Figure 1, participants in Group 1 completed three self-led (SL) assessments followed by two instructor-led (IL) assessments, and participants in Group 2 completed three IL assessments followed by two SL assessments. In each case, the first trial was used as a familiarisation. This approach was used to determine the between session reliability of the measures, to understand if the between session reliability differed between SL and IL assessment, and to determine if the order in which SL and IL assessments were completed influenced the between session reliability. Data collection to place between August 2024 and August 2025.

Instructor-led assessments

All instructors were experienced and trained in the delivery of the test battery and followed standardised operating procedures for each test and measurement meticulously.

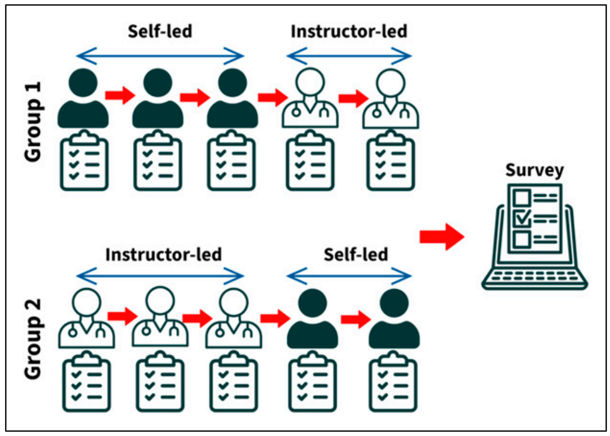

Figure 1. Summary of methodological approach. Participants were seperated into two groups and completed a series of Self- and Instructor-led functional fitness assessments. Upon completion all participants completed an online survey.

Participants attended the Human Performance Laboratory at the host institution on up to three separate occasions to complete IL assessments. Participants completed assessments at the same time of day to avoid circadian variation and were asked to abstain from intense physical activity in the 24hrs prior. Each trial consisted of nine established tests which were selected given their ease of application, validity, availability of sex and age specific normative data, and/or predictive ability. Body mass index (BMI) and Waist and hip circumference were determined as measures of body composition. Waist circumference and waist to hip ratio (WHR) have been shown to be positively associated with higher all-cause mortality risk independent of overall adiposity,[31] and have also been linked to cardiovascular disease, diabetes and cognitive impairment.[32,33] Therefore, such measures are useful in addressing some of the well purported limitations of BMI.[34] Moreover, it has been suggested that WHR may be a more valuable measure of risk stratification for older adults,[35] given mixed results regarding the association with BMI and mortality and an indication that in some instances an overweight BMI classification might be protective for older adults.[36]

Hand grip strength was measured as a global marker of muscular strength, and in older adult populations has been associated with mortality, disease, function in activities of daily living and disability, cognitive decline, risk of complications or prolonged length of stay after hospitalisation or surgery, and health related quality of life.[37–40] Performance in several functional fitness assessments (30s Arm Curl, 30s Chair Rise, 2-min Step Test, Timed Up and Go, Chair Sit and Reach), that replicate the movement patterns and physical facets of everyday activity (muscular strength, power and endurance; aerobic capacity; dynamic balance; flexibility, agility and coordination) were also determined. Unlike other assessments, these tests have been validated against gold standard assessments,[41] have established age and sex specific normative data expressed as percentiles to allow results to be more precisely contextualised,[42] and age and sex specific criterion referenced clinically relevant fitness standards for maintaining functional independence into the 10th decade of life have been determined.[43] Moreover, there is data to support that performance in specific tests (i.e. Timed Up and Go and 30s Chair Rise) is associated with falls, all-cause mortality, disease specific mortality, cardiovascular disease, reduced function in activities of daily living, and reduced physical performance following acute hospitalisation.[44–48]

Similar associations between gait speed and mortality, physical ability, cognitive decline, fall risk and disease have also been established,[49–51] justifying inclusion in the assessment battery. Fastest gait speed was measured given stronger association than usual gait speed for measures of physical function and balance confidence in community dwelling older adults.[49]

The personal, economic and societal impact of falls in older adult populations is a recognised public health priority.[52] It is estimated that 28-35% of adults aged over 65 yrs and 32-42% aged over 70 yrs will experience a fall annually,[53] with evidence indicating that falls represent the leading cause of accidental death in older adults.[54] Falls can lead to reduced physical function, loss of independence, and reduced quality of life.[55] As such, monitoring balance performance is essential. For this the single leg balance test was chosen and has been demonstrated to be a predictor of injurious falls in older adults.[56]

Each of these assessments were carried out based on the standardised procedures outlined below.

Body composition – Shoes and heavy clothing were removed and body mass (Digital Scales; Model 707, SECA Instruments Ltd., Germany) and stature (Stadiometer, SECA Instruments Ltd., Germany) assessed to the nearest 0.1kg and 0.1cm respectively. BMI (kg/m^2) was then determined.

Waist and hip circumference were also determined (to the nearest 0.1cm) using an anthropometric tape measure and in accordance with published protocols.[57,58] Measures were taken with participants in an upright standing position, with body mass evenly distributed, and following exhalation. The circumference of the smallest part of the waist and largest part of the hips were measured. Despite some disparity in methods for determining waist circumference, midpoint, umbilicus or minimal waist circumference measures appear to have limited effects on the association between waist circumference and all-cause mortality, cardiovascular disease and diabetes.[32]

Isometric Handgrip Strength – Isometric hand grip strength for the dominant side was measured following the procedures outlined by the American Association of Hand Therapists.[59] Participants were seated in a chair without arm support, with feet resting on the floor and hips flexed to 90°. The assessment was carried out with the elbow of the assessed arm flexed to 90°, the forearm neutral and the wrist withing 15-30° of extension (dorsiflexion). Initially the handle of the hand grip dynamometer (Takei 5401, Takei Scientific Instruments, Japan) was adjusted to a position where the participant felt able to squeeze as hard as possible following submaximal attempts. During measured assessments, participants were asked to squeeze the handle as hard as possible for ~5 s. Participants were permitted three attempts, separated by a minimum of 30 s of rest. The highest value was used for further analysis.

30s Arm Curl – The arm curl test is a recognised assessment of upper body strength.[60] Participants were seated on an armless chair, with feet flat on the floor, and using their dominant arm were asked to complete as many bicep curls as possible in 30 s whilst holding a dumbbell with a mass of 5 lbs (2.27 kg) or 8 lbs (3.63 kg) for females and males respectively. Only full repetitions were recorded. Repetitions where over 50% of the range of flexion was completed at the point time ceased were counted.[60]

30s Chair Stand – The chair stand test is a recognised assessment of lower body strength.[60] Participants sat on an armless chair with feet flat on the floor, shoulder width apart and back straight. With arms akimbo, participants were asked to rise to a full standing position and returning to the starting sitting position as many times as possible in 30 s. If time ceased and the participant had completed 50% of a chair rise, this was counted as a repetition.[60]

Timed up & go (TUG) – The TUG combines assessment of balance, agility and lower limb power.[60] Participants completed an 8-Foot TUG by moving from a seated position to standing, walking 8 feet, turning past a cone and returning to a seated position as quickly as possible. Time to complete the task was recorded using a stopwatch. Participants completed the task three times, with each attempt separated by 30 seconds of rest. Participants were instructed to complete the course as fast as possible, and the fastest time was recorded and used in later analysis.[60]

Chair Sit & Reach – The chair sit and reach was used to assess lower body flexibility. Sitting towards the edge of an armless chair with one leg fully extended, the foot of the opposing leg flat on the ground, and with one hand on top of the other, participants were asked to bend forward at the hip and attempt to touch the toes. The distance (cm) between the middle of the toes and the tip of the longest finger was recorded using an anthropometric tape measure. Using the midpoint of the toes as zero, if participants were able to reach beyond their toes, a positive distance was recorded. If the participants were unable to reach their toes a negative value was recorded. Given this measurement would be unfeasible for SL assessment, a four-zone scoring system was devised (Figure 2). A score of 1 was recorded if the extended fingertip of the longest finger could touch below the knee but above the top the ankle. A score of 2 was awarded if the reach was between the ankle and the big hallux, 3 if the hallux could be reached, and 4 if participants were able to reach beyond the hallux. Trials were discounted if knee flexion occurred during completion of the test or if any bouncing or jerking movements occurred. Participants were afforded 3 attempts on the dominant limb, with the best performance recorded.[60]

2-min Step Test (2MST) – The 2MST is a measure of aerobic endurance.[60] Participants were instructed to complete as many steps as possible within two minutes in a marching fashion. They began with their left leg, raising their knee to a predefined marker placed on a wall approximately 1.5 meters in front of them. Although it is recommended that the wall marker be positioned at a height corresponding to the midpoint between the patella and the iliac crest,[60] this was deemed impractical for SL assessment. Instead, before commencing the test, participants stood up right next to the wall with their shoulder and elbow at anatomical zero and their hand clenched into a fist. A wall marker was then placed at the midpoint of the hand, as this had been shown in preliminary studies to closely approximate the anatomically measured marker location. Participants were explicitly instructed that each step should reach the height of the marker and were reminded of this requirement during the test if necessary. They were informed of the elapsed time at 60, 90, and 110 seconds. The number of times the left knee reached the height of the marker during the test was recorded.

Fastest Gait Speed – Fastest gait speed was assessed by measuring the time required to walk a 4-m distance. Cones were placed at the start and end of the course. Participants began by standing with their feet shoulder-width apart next to the first cone. Following a countdown from the instructor, they were instructed to walk as fast as possible past the second cone. The time to complete the course, defined as the moment the participant passed the second cone, was recorded using a handheld stopwatch. Each participant completed three trials, and the fastest recorded gait speed (calculated as distance [m]/time [s]) was used for further analysis. Trials were excluded if there was any noticeable deceleration before passing the second cone.

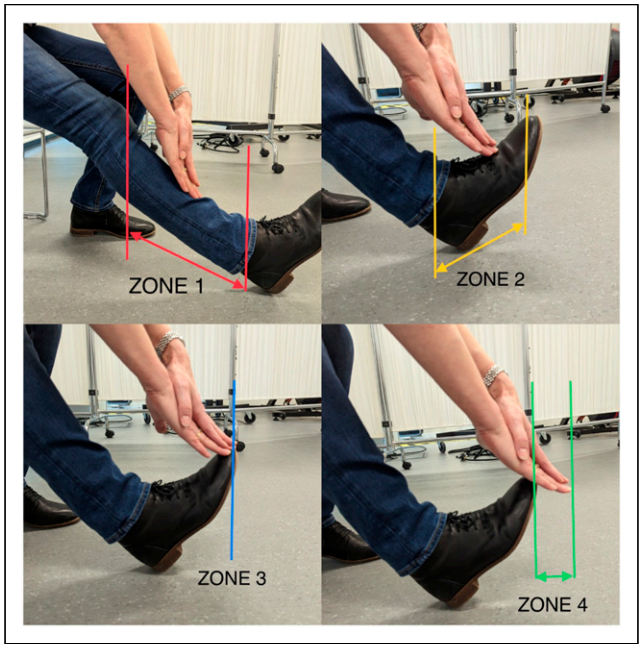

Figure 2. Chair sit and reach scoring. Zone 1 – If the extended fingertip of the longest finger touched below the knee but above the top of the ankle; zone 2 – between the ankle and the big toe; zone 3 – hallux; Zone 4 – beyond the hallux.

Single Leg Balance (SLB) – Participants completed a SLB on the dominant leg in accordance with previous literature.[61,62] Adjacent to a solid vertical surface, at approximately an arm's length away and with the standing leg placed furthest from the vertical surface, participants were asked to maintain balance on the dominant leg for as long as possible. Participants were asked to clasp their hands in front of them, maintain a neutral head position and raise the non-dominant leg a few centimetres from the ground via flexion of the knee. A handheld stopwatch was used to measure the time from when the participant lifted their non-standing foot and continued until balance was lost, defined as either touching the wall with their hand or contacting the ground with the non-standing foot. The test was terminated if participants were able to maintain balance for 120 seconds and three attempts were completed with the best performance used for further analysis.

FFMT & self-led assessment

For SL assessment, the same battery of tests as previously outlined was completed. Given the ambition to determine if these assessments could be competently performed by participants, there was no interference from the investigators. To support the delivery of SL assessment, participants were provided with the two-component digital FFMT via email (Figure 3) and a rucksack containing basic equipment to complete the tests (Figure 4).

The FFMT consists of a digital information package (provided as weblink) and a results monitoring and interpretation tool (provided as an email attachment). The digital information package provides details on the importance of functional fitness monitoring, instructions on how to use the FFMT, guidance on safe and accurate administration, and step-by-step directions for completing each test and inputting data into the results monitoring and interpretation tool. For each assessment, the package includes details on the test's purpose, the required equipment, and a short instructional video demonstrating test setup, execution, and scoring. The results monitoring and interpretation tool enables participants to understand age- and sex-specific target scores for each test, input test data, track and develop a performance history, and generate a report contextualising results with reference to age- and sex-specific percentiles and criterion referenced clinically relevant fitness standards for maintaining functional independence into the 10th decade of life. Where relevant, the tool also provides comparison against evidence-based thresholds for increased risk of ill health, falls, or other adverse outcomes.

The FFMT was distributed to participant' via email, and upon completion of all of the SL assessments, the data entered into the results monitoring and interpretation tool was returned to the investigators. All participants were informed that the FFMT was a prototype, with plans for future development into a web- and/or smartphone-based platform.

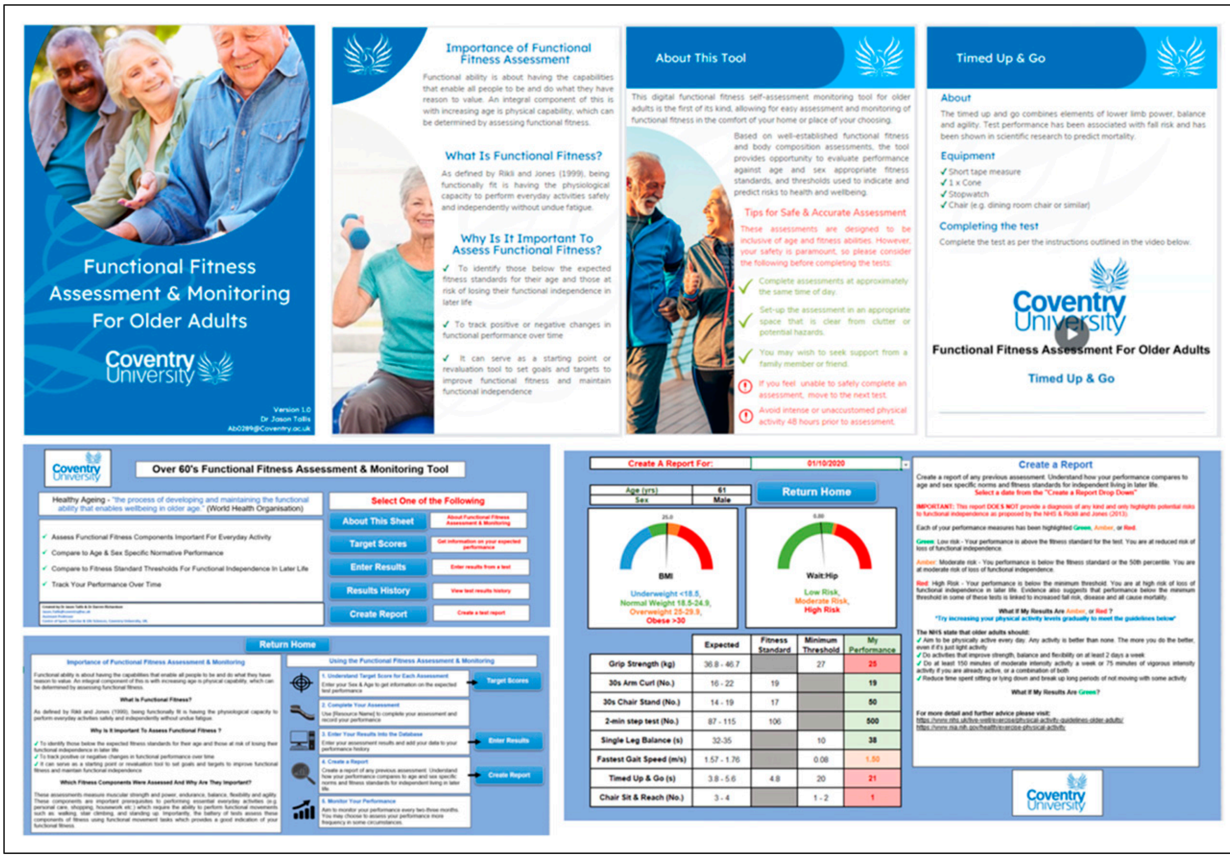

Figure 3. Images of the functional fitness monitoring tool for older adults. (Top panel = digital information package; Bottom panel = results monitoring and interpretation tool).

Online survey

Upon completion of the five trials, participants were asked to complete an online survey that was developed and distributed in JISC online surveys (https://www.onlinesurveys.ac.uk/). The survey aimed to understand the perceived importance of functional fitness monitoring, the value, useability and barriers to engagement of/with the FFMT, the suitability of the FFMT for long term functional fitness monitoring, and areas for improvement. These elements were evaluated using multiple choice, 10-point Likert scale and free text questions. Key themes from free text questions were identified and summarised in text. Prior to implementation, the survey was piloted by three older adults who met the study inclusion criteria (including one female participant) and by an independent researcher with extensive experience in survey design and evaluation of health promotion interventions. Feedback resulted in a small number of targeted amendments to improve the accessibility of terminology and to correct spelling and grammatical errors.

Statistical analysis of data

To assess whether performance in each assessment differed over time and between groups, a series of mixed-model ANOVAs (Analysis of Variance) were conducted. Group (Group 1 vs. Group 2) was included as the between-subjects factor, while Trial (1–5) was used as the within-subjects factor. Violations of sphericity were corrected using the Greenhouse-Geisser correction. Partial eta squared (ηp^2) was calculated as a measure of effect size and interpreted as small (>0.01), medium (>0.06), or large (>0.14) effects.[63] Significant main effects of Trial and interaction effects were further examined using Bonferroni-corrected pairwise comparisons and Cohen's d was calculated to estimate effect sizes interpreted as trivial (<0.2), small (<0.6), moderate (<1.2), or large (>1.2) effects.[64]

To determine the between session reliability of SL and IL assessment, a series of Intraclass Correlation Coefficients (ICC) and 95% confidence limits were calculated using an average measure two-way random effects model for absolute agreement (2,1).[65] The first familiarisation session was excluded. Subsequently, for each group separately, the between session

Figure 4. Equipment provided to participants to complete self-led functional fitness assessments. Participants were provided with a hand grip dynamometer, anthropometric tape measure, handheld stopwatch, stop clock, mechanical counter, 5 lbs (female) or 8 lbs (male) dumbbell, 4 x stickers, 2 x coloured cones, 4m cut to size measuring tape, cut to size measuring tape for the timed up & go.

reliability across trials was determined using a single-measure, two-way random effects model for absolute agreement (2,1).[65] Additionally, after excluding the first familiarisation trial, the reliability between IL and SL was evaluated using an average-measure, two-way random effects model for absolute agreement (2,k).[66] ICC's were interpreted as poor (<0.5), moderate (<0.75), good (<0.9) or excellent (>0.9) reliability.[66]

To establish a threshold for meaningful change and assess potential differences between groups, for each individual and assessment, the sum of the coefficient of variation (Mean/SD × 100) and the smallest worthwhile change (SD × 0.2) were calculated from trials 2-5 and averaged across participants. To understand if any difference in the between session reliability between groups effected the thresholds for change, a series of independent samples t-tests were conducted and Cohen's d was calculated and corrected for bias using Hedge's *g*.[67]

Statistical analysis was conducted using JASP (V 0.18.3.0, University of Amsterdam, Amsterdam) and graphical presentation of data was performed using GraphPad Prism (Version 8.3.1, San Diego, California). Statistical significance was *a priori* set at an alpha level of $P<0.05$. Data are presented as Mean±SD unless stated otherwise.

Results

Change in performance over time

For height, BMI, WHR, grip strength, 2MST, and the Chair sit & reach there was no main effect of Trial, Group or Trial*Group interaction (Table 1: $P>.104$; $\eta p^2<.084$). For the 30s Chair Stand, Fastest Gait Speed, TUG, and SLB, there was a main effect of Trial (Table 1: $P<.038$; $\eta p^2>.093$). Pairwise comparisons indicated that the time to complete the TUG in trial 1 was greater than all subsequent trials ($P<.013$; d = .230-.330), performance in the first trial of 30s Arm Curl was lower compared to subsequent trials ($P<.010$; d=.328-551), Fastest Gait Speed was lowest in the first assessment compared to trials 4 and 5 ($P<.006$; d=.684-.718), and SLB time was lower in the first trial compared to trial 5 ($P=.011$; d=.301).

Table 1. Mean ± SD performance outcomes & results from mixed model analysis of variance.

	Trial					ANOVA (P (ηp^2))		
	1	2	3	4	5	Time	Group	Group*time
Height (cm)	169.7 ± 10	169.7 ± 10.2	169.7 ± 10.1	169.6 ± 10.3	169.8 ± 10.2	.768 (<.005)	.894 (<.001)	.365 (.012)
Body Mass (kg)	73.9 ± 14.8	73.9 ± 14.9	73.8 ± 14.8	73.9 ± 14.7	73.8 ± 14.8	.971 (<.001)	.313 (.033)	.048 (.090)*
BMI (kg.m^2)	25.5 ± 3.7	25.5 ± 3.8	25.5 ± 3.7	25.5 ± 3.8	25.5 ± 3.8	.118 (.069)	.105 (.083)	.118 (.069)
Waist Circumference (cm)	91.3 ± 12	90.4 ± 10.8	91.4 ± 11.6	91.6 ± 12.3	91.6 ± 12	.349 (.036)	.448 (.019)	.048 (.085)*
Hip Circumference (cm)	101.4 ± 7.3	101.3 ± 7.3	101.4 ± 7.3	101.2 ± 7.5	101.2 ± 7.8	.890 (.006)	.213 (.051)	<.001 (.255)*
Waist:Hip	0.90 ± 0.08	0.89 ± 0.07	0.90 ± 0.07	0.90 ± 0.09	0.91 ± 0.08	.281 (.041)	.921 (.004)	.944 (.004)
Grip Strength (kg)	32.6 ± 8.4	33 ± 9.3	32.3 ± 9.2	33.0 ± 9.4	33.4 ± 8.7	.214 (.045)	.252 (.042)	.912 (.008)
30s Arm Curl (No.)	21.6 ± 5.2	22.1 ± 4.6	23.1 ± 4.7	24.0 ± 4.6	23.8 ± 4.5	<0.01 (.303)*	.369 (.032)	.032 (.093)*
30s Chair Stand (No.)	15.1 ± 2.8	16.0 ± 2.8	16.2 ± 3.1	16.6 ± 3.4	16.7 ± 3.1	<.001 (.252)*	.158 (.063)	.160 (.054)
2MST (No.)	100.9 ± 25	102.4 ± 16	108.2 ± 18	105.3 ± 16.3	105.7 ± 18.1	.126 (.070)	.464 (.019)	.402 (.030)
Fastest Gait Speed (m/s)	1.75 ± 0.36	1.84 ± 0.35	1.88 ± 0.4	1.92 ± 0.27	1.93 ± 0.31	<.001 (.160)*	.726 (.004)	.101 (.067)
TUG (s)	4.97 ± 1.02	4.76 ± 0.91	4.66 ± 0.93	4.72 ± 0.86	4.67 ± 0.85	<.001 (.150)*	.675 (.006)	.413 (.030)
SLB (s)	72.3 ± 48.1	80.2 ± 44.23	79.1 ± 45.17	83.6 ± 44.58	86.2 ± 46.64	.037 (.094)*	.452 (.018)	.192 (.051)
Chair Sit & Reach (Zone)	2.8 ± 1.19	2.9 ± 1.14	2.9 ± 1.14	2.9 ± 1.2	3 ± 1.2	.130 (.064)	.987 (<.001)	.279 (.040)

*N = 33; 2MST N = 31; ANOVA = Analysis of variance; BMI = Body Mass Index; 2MST = 2-min Step Test; TUG – Timed Up & Go; SLB = Single Leg Balance; * indicates P<.005.*

For Body Mass, Wait Circumference, Hip Circumference, and 30s Arm Curl, there was a significant Trial*Group interaction (Table 1: P<.049; ηp^2>.084). Pairwise comparisons indicated no differences between trials for either Group 1 or Group 2 (P>.052; d=.001-394) for the body composition measures. However, for the 30s Arm Curl, for Group 1, performance in the trial 1 was lower than performance in the trial 5 (P <0.001; d = .660), and performance in trial 2 was lower than in both trial 4 and trial 5 (P=<0.001; d=.535-585).

Between session reliability

When participants were considered collectively (Table 2), the between session reliability across IL and SL assessments was excellent (>.914) for all measures with confidence intervals ranging between good and excellent (.633 - .999). Fastest Gait Speed was the only exception, showing only good reliability (.817) and confidence intervals ranging between moderate and excellent.

When the participants that completed SL assessments first were considered separately (Tables 3 and 4), the between session reliability of SL assessments compared to IL assessments was excellent (>.915) for nearly all measures, and good (>.832) for 30s Arm Curl, 30s Chair Stand and Fastest Gait Speed. In all cases confidence intervals ranged between good and excellent, except for Fastest Gait Speed and 30s Chair Stand where confidence intervals ranged between moderate and excellent, and 30s arm curl where confidence intervals ranged from poor to excellent. To understand factors that may have influenced the wide-ranging confidence intervals for the 30s Arm curl, participants were median split separately by age, perceived computer literacy, and perceived understanding of the instructions in the FFMT instruction manual, and ICC analysis performed. There were no notable trends that differed that that presented in Tables 3 and 4 within this subsequent analysis.

Table 2. Between session reliability of self-led and instructor-led assessments.

	SL	IL	ICC (±95% CI's)
Height (cm)	169.8 ± 10.3	169.6 ± 10	.996 (.992-.998)
Body Mass (kg)	73.7 ± 14.9	74 ± 14.7	.999 (.998-.999)
BMI (kg.m^2)	25.5 ± 3.8	25.6 ± 3.7	.995 (.990-.998)
Wasit Circumference (cm)	90.8 ± 12.1	91.7 ± 10.9	.973 (.945-.987)
Hip Circumference (cm)	100.7 ± 7.5	101.8 ± 7.3	.983 (.934-.993)
Waist:Hip	0.9 ± 0.1	0.9 ± 0.1	.945 (.889-.973)
Grip Strength (kg)	32.8 ± 9.3	33.1 ± 8.9	.987 (.975.994)
30s Arm Curl (No.)	22.8 ± 4.7	23.7 ± 4.5	.928 (.845-.965)
30s Chair Stand (No.)	16.1 ± 2.9	16.7 ± 3.2	.937 (.886-.969)
2MST (No.)	104.3 ± 16.2	106.5 ± 17	.915 (.824-.959)
Fastest Gait Speed (m/s)	1.9 ± 0.4	1.9 ± 0.3	.817 (.633-.909)
TUG (s)	4.7 ± 0.8	4.7 ± 0.9	.959 (.918-.980)
SLB (s)	84.4 ± 44.7	80.1 ± 44.1	.961 (.921-.981)
Chair Sit & Reach (Zone)	3 ± 1.2	2.9 ± 1.2	.986 (.971-.993)

N = 33; 2MST N = 31; SL = Self-led assessment; IL= Instructor-led assessment; ICC = Intraclass Correlation Coefficients*; CI = Confidence Intervals; BMI = Body Mass Index; 2MST = 2-min Step Test; TUG – Timed Up & Go; SLB = Single Leg Balance.*

When the between session reliability of the specific attempts are considered (Tables 3 and 4), a similar trend to that of the between session reliability between IL and SL assessments was evident. However, in specific cases, (30s Chair Stand, 2MST, TUG) either the ICC or the range of confidence intervals were larger between trial 1 and trial 2 compared to subsequent trials.

When participants completed the IL assessments first (Table 5), the between session reliability of SL assessments compared to IL assessments was excellent (>.913) for all measures other than Fastest Gait speed which demonstrated good (>.813) reliability. In each case confidence intervals ranged between good and excellent, other than for 30s Chair stand which ranged between moderated and excellent, and Fastest Gait Speed where confidence intervals ranged between poor and excellent. For Fastest Gait speed, considering the reliability when participants were median split separately by age, perceived computer literacy, and perceived understanding of the instructions in the FFMT did reveal any notable trends that differed from the primary analysis. Typically, the between session reliability between specific trials followed the trends demonstrated for between session reliability of between the SL and IL reliability assessments.

Threshold for change

The threshold for a measurable change is presented in Table 5. When each group was considered separately, the threshold for change did not differ between assessments (P>.068; d=.030-.489). Waist circumference was the only exception where the change threshold was greater for Group 1 compared to Group 2 (P=.019; d=.877).

Participant perceptions of the FFMT

Results of the survey are presented in Figure 5. Self-rated computer literacy ranged between 1-10 (Median = 5 [IQR 5-8]). The perceived importance of functional fitness monitoring was high (Median = 9 [IQR 8-10]), although only 24% of participants engaged with some form of monitoring prior to participant in the current study. Free text responses indicated monitoring activity included perceived effort in habitual physical activity/daily tasks (N=3), perceived effort in habitual physical activity/daily tasks inclusive of organised exercise (N=4), use of a wrist worn physical activity monitor and performance monitoring based on outputs from gym equipment (N=1).

The perceived ease of using the FFMT digital resources (Median = 8 [IQR 7-9]), ease of interpretation of the guidance (Median = 9 [IQR 7.5-10]) and report (Median = 9 [IQR 7.5-10]), were all high. 82% of participants completed the SL assessments without support and perceived competency for SL assessment was high (Median = 9 [IQR 8-10]). Free text responses indicated that support was required or would be useful for inputting data (N=2); measuring performance of time bound assessments (N=3), both measuring performance of timed bound assessment and counting repetitions in tasks where this was required (N=1).

Eighty-two percent of participants outlined that they perceived the FFMT would be useful to track performance overtime, with a majority outlining the FFMT would be very useful in this context (Median = 7 [IQR 5-8]). Seventy-three% of

Table 3. Mean ± SD group 1 performance outcomes & between session reliability.

	Trial					ICC (±95% CI's)				
	1	2	3	4	5	1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	IL vs. SL
Height (cm)	169.5 ± 10	169.5 ± 10	169.5 ± 10	169.3 ± 9.7	169.5 ± 9.8	1.000 (1.000-1.000)	1.000 (1.000-1.000)	.994 (.983-.998)	.998 (.994-.999)	.997 (.993-.999)
Body Mass (kg)	76.3 ± 13.4	76.3 ± 13.7	76.3 ± 13.5	76.7 ± 13.4	76.5 ± 13.4	.999 (.997-1.000)	.999 (.997-1.000)	.996 (.990-.999)	.998 (.994-.999)	.998 (.996-.999)
BMI (kg.m^2)	26.5 ± 3.6	26.5 ± 3.7	26.5 ± 3.7	26.7 ± 3.7	26.6 ± 3.6	.998 (.994-.999)	.998 (.994-.999)	.994 (.977-.998)	.995 (.985-.998)	.997 (.992-.999)
Wasit Circumference (cm)	92.1 ± 12.9	91.5 ± 12.4	92.5 ± 13.1	93.9 ± 13.8	94.2 ± 12.2	.993 (.980-.998)	.967 (.910-.988)	.897 (.737-.962)	.929 (.813-.974)	.968 (.898-.989)
Hip Circumference (cm)	102.5 ± 7.5	102.5 ± 7.7	102.5 ± 7.5	103.4 ± 7.7	103.9 ± 7.8	.997 (.993-.998)	.997 (.993-.998)	.964 (.918-.983)	.959 (.919-.980)	.989 (.898-.994)
Waist:Hip	0.9 ± 0.09	0.89 ± 0.09	0.9 ± 0.09	0.91 ± 0.11	0.91 ± 0.09	.989 (.976-.995)	.952 (.904-.976)	.855 (.729-.926)	.882 (.776-.940)	.980 (.960-.980)
Grip Strength (kg)	30.8 ± 8.3	31 ± 9	30.6 ± 8.7	31.4 ± 8.7	31.8 ± 8	.906 (.765-.965)	.953 (.879-.983)	.978 (.937-.992)	.941 (.847-.978)	.988 (.964-.996)
30s Arm Curl (No.)	20.5 ± 3.3	21.1 ± 3.9	22.1 ± 3.9	23.9 ± 4	23.6 ± 3.7	.817 (.573-.929)	.886 (.670-.959)	.748 (.270-.912)	.964 (.906-.987)	.847 (.145-.957)
30s Chair Stand (No.)	14.8 ± 2.4	15.4 ± 1.9	15.5 ± 2.2	15.7 ± 2.3	15.9 ± 2.1	.583 (.172-.824)	.802 (.544-.923)	.754 (.452-.903)	.921 (.801-.970)	.841 (.684-.957)
2MST (No.)	100.1 ± 27.3	97.5 ± 14.5	106.8 ± 18.6	103.6 ± 17.3	104.4 ± 17.6	.444 (-.020-.754)	.663 (.146-.877)	.822 (.584-.931)	.966 (.911-.987)	.916 (.774-.969)
Fastest Gait Speed (m/s)	1.7 ± 0.4	1.83 ± 0.42	1.81 ± 0.46	1.95 ± 0.23	1.94 ± 0.29	.841 (.531-.944)	.836 (.612-.937)	.605 (.206-.834)	.788 (.516-.917)	.833 (.532-.940)
TUG (s)	5.11 ± 0.81	4.85 ± 0.78	4.71 ± 0.8	4.74 ± 0.77	4.69 ± 0.71	.729 (.394-.893)	.897 (.743-.962)	.874 (.694-.952)	.963 (.904-.986)	.932 (.817-.975)
SLB (s)	63.6 ± 47.4	75.3 ± 46.03	79.4 ± 48.7	77.4 ± 48.1	77.8 ± 50.3	.830 (.582-.936)	.951 (.873-.982)	.931 (.825-.974)	.915 (.785-.968)	.977 (.938-.992)
Chair Sit & Reach (Zone)	2.9 ± 1.2	2.9 ± 1.2	2.9 ± 1.2	2.9 ± 1.3	2.9 ± 1.3	.960 (.897-.985)	.980 (.947-.993)	.942 (.851-.978)	.944 (.856-.979)	.990 (.973-.996)
Chair Sit & Reach (cm)	-		-	-2.8 ± 15.26	-2.7 ± 15.14	-	-	-	.979 (.945-.992)	-

N = 17; ICC = Intraclass Correlation Coefficients*; CI = Confidence Intervals; BMI = Body Mass Index; 2MST = 2-min Step Test; TUG – Timed Up & Go; SLB = Single Leg Balance*.

participants indicated that they would be interested in using the FFMT to track performance over time, and 18% were unsure. For those that said they would not be interested in using the FFMT to track performance overtime, free text comments indicated that competition and encouragement was needed (N=1); that it would only be attractive if the frequency was once or twice per year (N=1); and that need to regularly record data was unattractive (N=1). Barriers to long term monitoring included access to the equipment (N=15), which were in the most part was specific to the hand grip dynamometer and the dumbbell, the need for a reminder (N=3), the need for an incentive or motivation (N=3); advice on the frequency of assessment (N=2), referral from a medical professional (N=1), opportunity to engage periodically with an appropriate professional (N=1).

In remaining free texts questions where participants were asked what further considerations were needed to improve the FFMT, the main theme was a need to develop a mobile application (N=6), web-based application (N=4), more accessible platform (N=3). Which aligns to the intention of the investigators that was communicated with participants at the outset. Specific improvements to layout and formatting were identified (N=3); the potential to link to interventions (N=2); the value

Table 4. Mean ± SD group 2 performance outcomes & between session reliability.

	Trial					ICC (±95% CI's)				
	1	2	3	4	5	1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	IL vs. SL
Height (cm)	169.8 ± 10.3	169.9 ± 10.6	169.9 ± 10.6	170 ± 11.1	170 ± 10.9	.997 (.992-.999)	.998 (.994-.999)	.989 (.971-.996)	.999 (.998-1.000)	.955 (.986-.998)
Body Mass (kg)	71.3 ± 16.2	71.3 ± 16.2	71.2 ± 16	70.9 ± 15.8	71 ± 16.1	1.000 (.999-1.000)	1.000 (.999-1.000)	.999 (.997-1.000)	.999 (.997-1.000)	.999 (.998-1.000)
BMI (kg.m^2)	24.5 ± 3.7	24.5 ± 3.7	24.4 ± 3.6	24.3 ± 3.6	24.3 ± 3.7	.998 (.994-.999)	.997 (.991-.999)	.986 (.962-.995)	.997 (.992-.999)	.992 (.978-.997)
Wasit Circumference (cm)	90.4 ± 11.3	89.3 ± 9.1	90.4 ± 10.2	89.2 ± 10.4	89.1 ± 11.6	.951 (.866-.983)	.962 (.892-.987)	.956 (.878-.984)	.961 (.895-.986)	.994 (.949-.994)
Hip Circumference (cm)	100.5 ± 7.1	100.5 ± 7.3	100.3 ± 7.1	98.9 ± 6.8	98.3 ± 7.3	.979 (.942-.993)	.973 (.927-.980)	.950 (.760-.985)	.953 (.873-.983)	.961 (.891-.986)
Waist:Hip	0.9 ± 0.06	0.89 ± 0.04	0.9 ± 0.06	0.9 ± 0.05	0.9 ± 0.06	.791 (.510-.921)	.813 (.555-.930)	.838 (.606-.940)	.872 (.679-.953)	.961 (.891-.986)
Grip Strength (kg)	34.4 ± 8.4	35.2 ± 9.4	34.1 ± 9.7	34.8 ± 10.1	35.1 ± 9.3	.955 (.878-.984)	.976 (.920-.992)	.965 (.906-.988)	.990 (.974-.997)	.987 (.994-.995)
3	22.8 ± 6.7	23.2 ± 5.2	24.3 ± 5.3	24.1 ± 5.3	24.1 ± 5.3	.874 (.685-.954)	.943 (.765-.983)	.953 (.875-.983)	.954 (.877-.984)	.988 (.938-.992)
30s Chair Stand* (No.)	15.3 ± 3.2	16.8 ± 3.5	17.1 ± 3.7	17.6 ± 4.2	17.6 ± 3.8	.795 (.293-.933)	.944 (.856-.979)	.903 (.758-.963)	.975 (.934-.991)	.894 (.737-.960)
2 min Step Test (No.)	101.8 ± 22.9	108.4 ± 16.2	109.9 ± 18	107.3 ± 15.4	107.3 ± 19.2	.823 (.473-.943)	.943 (.837-.981)	.759 (.415-.915)	.929 (.800-.976)	.914 (.740-.972)
Fastest Gait Speed (m/s)	1.8 ± 0.32	1.84 ± 0.28	1.96 ± 0.3	1.88 ± 0.31	1.92 ± 0.34	.863 (.659-.949)	.795 (.313-.934)	.629 (.229-.851)	.858 (.649-.948)	.814 (.482-.935)
TUG (s)	4.83 ± 1.22	4.66 ± 1.05	4.61 ± 1.08	4.7 ± 0.97	4.65 ± 1	.953 (.850-.984)	.942 (.847-.979)	.922 (.797-.972)	.964 (.902-.987)	.976 (.933-.992)
SLB (s)	98.7 ± 44.21	103 ± 27.93	87.6 ± 35.23	104.9 ± 29.24	111.1 ± 25.3	.892 (.725-.960)	.904 (.749-.965)	.804 (.519-.927)	.951 (.866-.983)	.940 (.791-.980)
Chair Sit & Reach (Zone)	2.7 ± 1.2	2.9 ± 1.09	2.9 ± 1.12	3 ± 1.15	3.1 ± 1.18	.881 (.699-.957)	.974 (.930-.991)	.976 (.934-.991)	.977 (.937-.992)	.981 (.945-.994)
Chair Sit & Reach (cm)	-6 ± 14.78	-4.8 ± 14.71	-4.5 ± 15.98	-	-	.971 (.913-.990)	.978 (.935-.993)	-	-	-

N = 16; 2MST N = 14; Chair Sit & Reach (cm) N = 14; ICC = Intraclass Correlation Coefficients*; CI = Confidence Intervals; BMI = Body Mass Index; 2MST = 2-min Step Test; TUG – Timed Up & Go; SLB = Single Leg Balance.*

of presenting normative data for individuals living with health conditions (N=1), and opportunity to engage with an appropriate professional (N=1) were identified.

Discussion

The present study highlights the usability and reliability of a novel, digitally delivered, functional fitness self-assessment and monitoring tool for older adults, designed to evaluate key domains of physical function associated with comorbidity, mortality, and wellbeing. Between-session reliability of the included assessments was predominantly good to excellent when comparing instructor-led and self-led administration. Moreover, reliability was largely equivalent regardless of whether self-led assessment preceded or followed instructor-led assessment. The findings also suggest that most of the included assessments required minimal familiarisation, however, specific tests may benefit from the exclusion of the initial trial and from further refinement of instructional guidance. Importantly, participants recognised the significance of functional fitness

Table 5. Change threshold for each test determined as the sum of the coefficient of variation and the smallest worthwhile change.

	All	Group 1	Group 2	Group 1 vs. group 2 (P= (*d*=))
Body Mass (kg)	0.69 ± 0.45	0.78 ± 0.48	0.59 ± 0.41	.239 (.418)
BMI (kg.m^2)	0.34 ± 0.24	0.30 ± 0.19	0.39 ± 0.29	.247 (.411)
Wasit Circumference (cm)	3.13 ± 1.8	3.86 ± 1.89	2.4 ± 1.41	.019 (.877)*
Hip Circumference (cm)	1.50 ± 1.06	1.37 ± 0.79	1.63 ± 1.29	.508 (.237)
Waist:Hip	0.03 ± 0.02	0.03 ± 0.02	0.03 ± 0.02	.177 (.489)
Grip Strength (kg)	1.85 ± 1.11	2.00 ± 1.24	1.69 ± 0.97	.432 (.277)
30s Arm Curl (No.)	1.71 ± 1.04	2.03 ± 1.14	1.37 ± 0.83	.069 (.366)
30s Chair Stand (No.)	1.12 ± 0.64	1.08 ± 0.53	1.16 ± 0.75	.713 (.129)
2 min Step Test (No.)	7.73 ± 4.67	8.71 ± 4.66	6.54 ± 4.57	.203 (.470)
Fastest Gait Speed (m/s)	0.20 ± 0.1	0.22 ± 0.09	0.18 ± 0.1	.262 (.397)
TUG (s)	0.29 ± 0.13	0.32 ± 0.1	0.27 ± 0.15	.312 (.358)
SLB (s)	10.76 ± 13.11	10.57 ± 12.06	10.97 ± 14.54	.931 (.030)

*N = 33; 2MST N = 31; BMI = Body Mass Index; 2MST = 2-min Step Test; TUG – Timed Up & Go; SLB = Single Leg Balance; * indicates P<.005.*

monitoring and perceived the FFMT as valuable for benchmarking and tracking performance over time. The consistency of the reliability outcomes and participant experiences demonstrate the tool's ease of use and clarity of data interpretation, supporting its suitability for long-term functional fitness monitoring. Given participants limited prior engagement in self-monitoring and a lack of widely accessible tools, the FFMT may offer a practical and accessible means for health-related fitness assessment. Collectively, these findings position the FFMT as a scalable digital strategy for accurate benchmarking and longitudinal monitoring, with potential to support healthy ageing, enhance quality of life, and reduce healthcare costs. Clinically, the FFMT could be effective in supporting practitioners in enabling earlier identification of adverse ageing trajectories, support prescription of personalised intervention, and support patient self-management during rehabilitation and post-hospitalisation recovery.

Several studies have examined the between-session reliability of the functional fitness assessments utilised in the present study.[43,68–71] However, this evidence has predominantly been generated under conditions in which test instructions, administration, and interpretation were delivered by experienced assessors. The findings of the present study substantially advance this literature by demonstrating, for the first time, that with appropriately designed digital instructions, older adults can administer these functional fitness assessments both reliably and accurately on their own, without prior familiarisation or direct supervision from trained professionals. Taken collectively and considering the wide age range represented in the present sample, these findings support the broader applicability of functional fitness assessments beyond controlled laboratory environments. They support the use of these assessments in home-based, community, and a wider range of clinical contexts. By enabling reliable self-assessment, the FFMT has the potential to substantially increase access to functional fitness monitoring, reduce dependence on specialist personnel, and facilitate more frequent, scalable evaluation of physical function in ageing populations.

Whilst a high degree of reliability was evident when participants were stratified by group, and when examined both holistically and across individual trials, performance in the 30-s arm curl, 30-s chair stand, fastest gait speed, TUG, and single-leg balance tests was consistently lower in the first trial compared with subsequent attempts, highlighting the need for familiarisation. Although the point estimates for the ICC remained high, the reliability between the first and second attempts was characterised by wider confidence intervals, indicating increased variability during the initial exposure. This suggests that, although some individuals may perform adequately without a familiarisation trial, and the practical impact on percentile classification or risk categorisation may be limited due to relatively small between-trial differences, this initial performance decrement becomes more meaningful if the assessments are intended for long-term monitoring. Although the analysis demonstrated no group effect, indicating that the initial performance deficit was comparable between instructor-led and self-administered conditions, closer inspection of the reliability indices revealed narrower confidence intervals between the first and second trials in the group initially exposed to instructor-led assessments. Collectively, these findings indicate that familiarisation is necessary and may be particularly important when assessments are first delivered in a digital or remote format. Despite appearing relatively simple, these tasks likely impose a cognitive load associated with simultaneously understanding instructions, executing the test, and scoring performance which may account for the reduced reliability

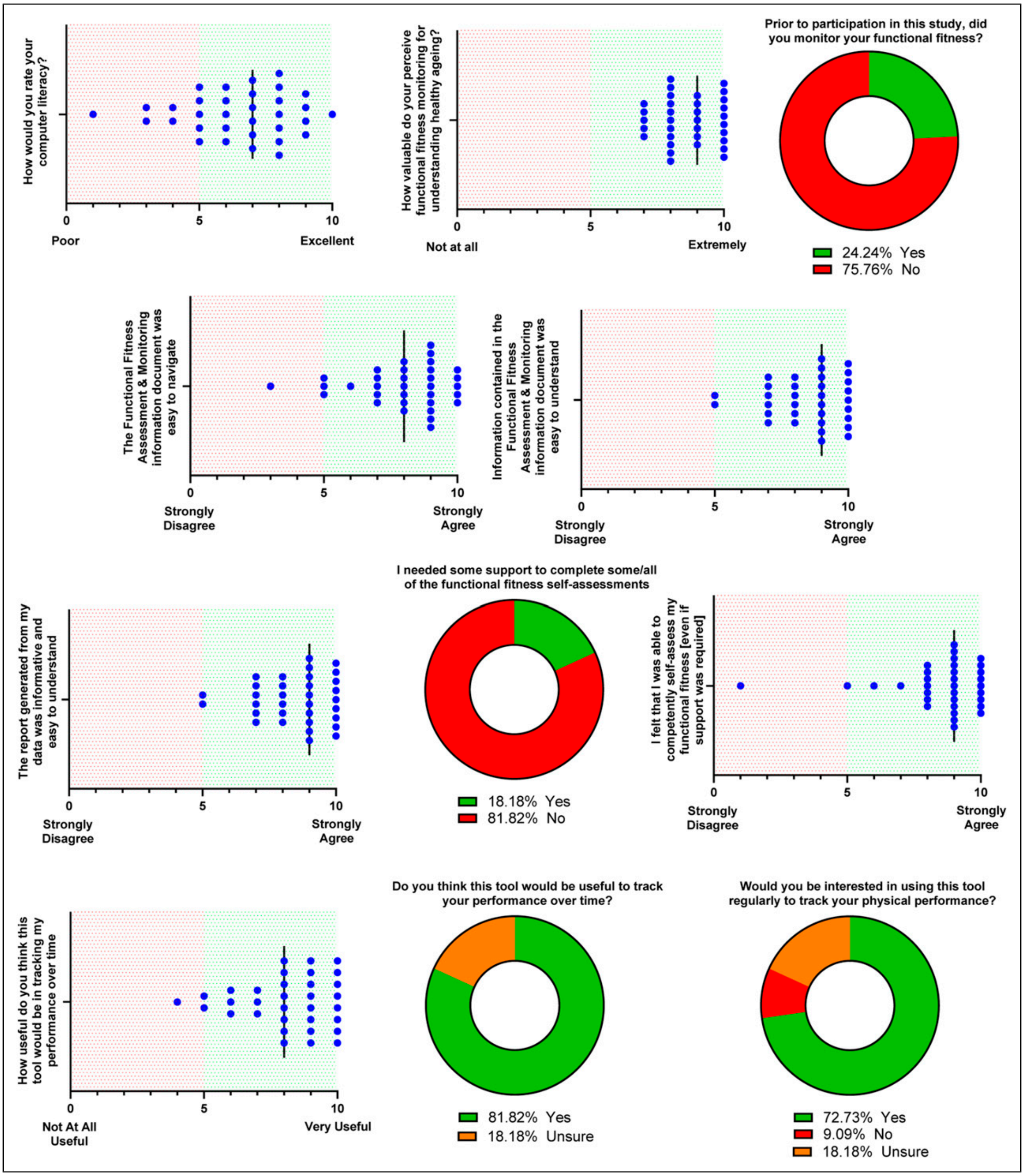

Figure 5. Summary of survey results.

observed in the first self-administered attempt. Therefore, future application of the FFMT in practice would benefit from incorporating a dedicated practice trial or pre-assessment familiarisation phase.

Given its simplicity and association with[49–51] in older adult population, gait speed assessments are widely implemented. Our findings provide needed insight into the reliability of fastest gait speed in older adult populations, where this assessment maybe useful in overcoming the poor consistency of usual gait speed assessment.[68,72] Whilst collectively, the results infer high between session reliability, when individual trials were examined, the confidence intervals ranged from poor to excellent, with no discernible pattern differentiating the two groups. Although the overall reliability and the distribution of confidence intervals were similar when instructor-led repetitions were compared with self-administered counterparts, a noticeably wider spread of confidence intervals was observed when comparing the instructor-led repetition to the first self-administered repetition. The data therefore indicate that even within the same testing environment, the breadth of the confidence intervals reflects a degree of inconsistency in repeatability for certain individuals. This may be further exacerbated when assessments are self-administered without an independent assessor to confirm accurate completion of the walking course. Although refinement of written or video-based instructions may help mitigate these issues, the manual timing approach used in the present study remains susceptible to error when compared with more precise performance quantification methods, such as photocell timing systems.[73,74] The literature has shown that device-based technologies, including those derived from smartphones can provide accurate and reliable evaluations of gait performance.[75] Incorporating such technology into self-administered testing protocols may therefore represent a pragmatic and accessible solution to enhance the reliability of gait speed assessment in remote or unsupervised contexts.

The current results also contribute meaningfully to the limited and conflicting evidence regarding the between-session reliability of single-leg balance assessments in older adults. While consistency in performance, similar to that observed in the present study, has been reported previously,[68] other studies have shown poorer reliability, particularly among individuals with lower baseline balance ability and among the oldest-old (≥75 years).[68,76] This phenomenon warrants further investigation and may not be specific to the single-leg balance test.

Whilst the between-session reliability of the 30-s arm curl was high across both self-led and instructor-led assessments, examining the data at the individual trial level revealed important differences in performance stability. Specifically, the range of confidence intervals was typically wider for participants who completed the initial trial in the self-led condition compared with those who commenced testing under instructor-led supervision. The present findings therefore suggest that appropriately conducted instructor-led assessments may reduce the need for a separate familiarisation trial, as performance in this group stabilised more rapidly. In contrast, participants who began in the self-led condition appeared to require a longer familiarisation period, evidenced by progressive improvements in performance during the early trials. This disparity between groups indicates that the provision of clearer guidance, particularly regarding correct execution and the need for maximal effort, may help mitigate these early inconsistencies in self-administered testing. Although it is theoretically plausible that social influences such as those associated with external evaluation, which has been demonstrated to improve performance of strength based assessment,[77] may underpin this disparity, this explanation is unlikely in the current context. No comparable improvement was observed in other maximal performance assessments within the study, suggesting that arousal-related effects were not a major contributing factor. Instead, the observed pattern is more plausibly attributed to increased cognitive and procedural demands during the first self-administered attempt, which may reduce the consistency of performance until participants become better acquainted with the test requirements.

Whilst targeted refinements to the tool are required, the present data support the FFMT as a viable and promising innovation for functional fitness monitoring. Our comprehensive approach to establishing between-session reliability represents a cornerstone of robust intervention design models such as the MRC framework for developing and evaluating complex interventions, which emphasises refining intervention on the basis of consistent, rigorous measurement,[78] where the use of these reliability estimates to determine measurable change thresholds enhances the future capability of the FFMT to detect meaningful longitudinal shifts in performance. In nearly all assessments, the thresholds for detectable change were small, and where available aligned with those derived from instructor led assessments in previous work.[79] This development strengthens key principles outlined in the BCTT,[24] particularly by improving the FFMT's goal-setting functionality beyond simple age- or risk-based thresholds, enabling more precise and individualised performance targets to support progress monitoring.

Beyond between session reliability, the broader potential of the FFMT is reinforced by overwhelmingly positive user perceptions, which extended beyond ease of use and clarity of interpretation, even among individuals reporting lower perceived digital literacy. Despite participants clearly recognising the value of structured functional fitness monitoring, only a minority reported engaging in any form of regular tracking before involvement in this study, and when such monitoring occurred, it was primarily based on subjective perceptions during daily activities or structured exercise. Within the context of the BCW,[23] this disconnect is likely attributable to reduced physical opportunity (i.e., limited access to suitable tools) and diminished psychological capability (i.e., limited understanding of how to perform and interpret such assessments). The FFMT is strategically designed to address these barriers. In addition to its primary role in providing opportunities for benchmarking against population norms and established risk thresholds, the FFMT offers a platform for longitudinal mapping of performance trajectories and for identifying thresholds indicative of measurable change. Together, these features align with multiple clusters of the BCTT, including goal setting and planning, feedback and monitoring, natural consequences, comparison of behaviour, and comparison of outcomes.[24] The data presented in the study with respect to reliability and perceived value and usability collectively position the FFMT as a powerful tool for the early identification of risk and facilitating positive health behaviour change both independently and as part of broader intervention initiatives.

Limitations & future directions

Whilst this study represents an important initial first step in reaching the wide research benefits of accessible functional fitness monitoring, it is not without limitation. This study included older adults with a wide range of functional abilities. However, most participants were above key risk thresholds, limiting generalisability to individuals with more pronounced mobility impairments, multimorbidity, cognitive decline, or lower digital literacy. Whilst participants reported generally positive experiences with the FFMT, successful engagement with the tool requires not only basic digital skills to access and navigate the platform, but also the ability to interpret assessment outcomes and utilise this information to support health-related decision making. Future research should therefore examine usability and effectiveness in populations with a broader range of digital health literacy levels, alongside the development of algorithms and decision-support systems that facilitate data interpretation and translate assessment outcomes into personalised intervention recommendations. Future work should also examine the FFMT in more diverse and clinically complex populations, and particularly in a larger population of older adults aged >75 years. Whilst the age of participants (60–78 years) in the present study reflects an important target population for healthy ageing strategies aimed at maintaining functional independence, the findings may not fully generalise to the oldest-old (≥80 years), who are more likely to experience greater functional impairment, multimorbidity, cognitive decline, and technology-related barriers. As such, this work should be viewed as a focused first step towards supporting the wider scalability and application of accessible functional fitness monitoring across the ageing population.

Further refinement of the tool would benefit from continued stakeholder engagement to ensure that instructions, accessibility features, and interface design remain appropriate for real-world use. Incorporating accessible device-based technologies, such as smartphone video or inertial sensors, may reduce equipment barriers, enhance timing accuracy, and provide additional insight into movement strategy, which outcome-only metrics may overlook. This is particularly relevant as movement-strategy changes often precede overt declines in performance as seen in other fields[80–82] and may therefore strengthen early identification of functional deterioration.

Longitudinal studies are now required to determine the FFMT's sensitivity to meaningful change, its predictive value for adverse outcomes, and its integration within clinical decision-making pathways. Implementation studies across primary care, rehabilitation, and community health settings will be important for understanding how the FFMT can support preventive care, enable self-management, and complement existing service structures. An essential part of this would be to evaluate he suitability and safety of remote, unsupervised assessment in individuals with greater functional limitations or cognitive impairment and identify circumstances where caregiver or professional support may be required. Finally, formal health-economic evaluations are needed to quantify potential cost savings associated with earlier detection, reduced hospitalisation, fewer specialist appointments, and the delivery of scalable monitoring.

Conclusion

This study developed and evaluated a novel FFMT as the first digitally delivered, self-administered assessment platform designed specifically to support functional fitness monitoring in older adults. The FFMT demonstrated good-to-excellent between-session reliability across most assessments, with performance largely comparable to instructor-led administration. Although a small number of tests exhibited small inconsistencies during the initial trial, these can be mitigated through refined guidance or exclusion of the first attempt. By enabling accurate and repeatable assessment outside specialist settings, the FFMT aligns closely with key public health priorities. It provides a practical mechanism for the early detection of functional decline, supports a shift from reactive care to proactive prevention, and expands access to functional monitoring across clinical, community, and population-level contexts. Collectively, these findings position the FFMT as a scalable digital solution with the potential to enhance healthy ageing, improve quality of life, and reduce the broader burden of age-related functional deterioration.

Acknowledgements

The authors would like to thank Mrs Susie Wilson for providing technical assistance to support the completion of the project.

ORCID iD

Jason Tallis https://orcid.org/0000-0001-8904-2693

Ethical considerations

Ethics approval was sought from the host institute [Ref: P138035] prior to study commencement.

Consent to participate

All participants provided written informed consent.

Author contributions

Conceptualisation (JT, DR); Methodology (JT, DR), Software (JT), Investigation (JT, DR, NC, DB), Formal analysis (JT, MD), Visualisation (JT), Project administration (JT, DR), Writing - Original Draft (JT), Writing - Review & Editing (All).

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

Data will be made available upon reasonable request by contacting the corresponding author.

References

1. WHO. *Ageing and health*. World Health Organisation. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health#:~:text=By_2030%2C_1_in_6,will_double_(2.1_billion
2. Divo MJ, Martinez CH and Mannino DM. Ageing and the epidemiology of multimorbidity. *Eur Respir J* 2014; 44(4): 1055–1068. https://doi.org/10.1183/09031936.00059814
3. Spiers GF, Kunonga TP, Beyer F, et al. Trends in health expectancies: a systematic review of international evidence. *BMJ Open* 2021; 11(5): e045567. https://doi.org/10.1136/bmjopen-2020-045567
4. Janssen I, Shepard DS, Katzmarzyk PT, et al. The healthcare costs of sarcopenia in the United States. *Journal of the American Geriatrics Society* 2004; 52(1): 80–85. https://doi.org/10.1111/j.1532-5415.2004.52014.x
5. Daskalopoulou C, Stubbs B, Kralj C, et al. Physical activity and healthy ageing: A systematic review and meta-analysis of longitudinal cohort studies. *Ageing Res Rev* 2017; 38: 6–17. https://doi.org/10.1016/j.arr.2017.06.003
6. Urtamo A, Jyväkorpi SK and Strandberg TE. Definitions of successful ageing: a brief review of a multidimensional concept. *Acta Biomed* 2019; 90(2): 359–363. https://doi.org/10.23750/abm.v90i2.8376
7. Chodzko-Zajko WJ, Proctor DN, Singh MAF, et al. Exercise and physical activity for older adults. *J Medicine Science in Sports Exercise and Sport Sciences Reviews* 2009; 41(7): 1510–1530.
8. Li X and Li C. Promoting Healthy Aging: Physical Activity and Its Dual Effects on Physical Health and Cognitive Function in Chinese Older Adults. *Front Public Health* 2025; 13: 1561060. https://doi.org/10.3389/fpubh.2025.1561060
9. PHE. *Wider impacts of COVID-19 on physical activity, deconditioning and falls in older adults*. Public Health England, 2021. https://www.gov.uk/government/publications/covid-19-wider-impacts-on-people-aged-65-and-over
10. NHS. *Statistics on Obesity, Physical Activity and Diet*. NHS England, 2020. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-obesity-physical-activity-and-diet/england-2020
11. Yamamoto S, Hotta K, Ota E, et al. Effects of resistance training on muscle strength, exercise capacity, and mobility in middle-aged and elderly patients with coronary artery disease: A meta-analysis. *J Cardiol* 2016; 68(2): 125–134. https://doi.org/10.1016/j.jjcc.2015.09.005
12. Zhang Y, Zhang Y, Du S, et al. Exercise interventions for improving physical function, daily living activities and quality of life in community-dwelling frail older adults: A systematic review and meta-analysis of randomized controlled trials. *Geriatric Nursing* 2020; 41(3): 261–273. https://doi.org/10.1016/j.gerinurse.2019.10.006
13. Wright J and Baldwin C. Oral nutritional support with or without exercise in the management of malnutrition in nutritionally vulnerable older people: A systematic review and meta-analysis. *Clin Nutr* 2018; 37(6 Part A): 1879–1891. https://doi.org/10.1016/j.clnu.2017.09.004
14. Taylor LM, Kerse N, Frakking T, et al. Active Video Games for Improving Physical Performance Measures in Older People: A Meta-analysis. *J Geriatr Phys Ther* 2018; 41(2): 108–123. https://doi.org/10.1519/JPT.0000000000000078
15. Di Lorito C, Long A, Byrne A, et al. Exercise interventions for older adults: A systematic review of meta-analyses. *Journal of Sport and Health Science* 2021; 10(1): 29–47.
16. Strain T, Sharp SJ, Spiers A, et al. Population level physical activity before and during the first national COVID-19 lockdown: A nationally representative repeat cross-sectional study of 5 years of Active Lives data in England. *The Lancet regional health Europe* 2022; 12: 100265. https://doi.org/10.1016/j.lanepe.2021.100265

17. Pavasini R, Guralnik J, Brown JC, et al. Short Physical Performance Battery and all-cause mortality: systematic review and meta-analysis. *BMC Med* 2016; 14(1): 215. https://doi.org/10.1186/s12916-016-0763-7
18. Welch SA, Ward RE, Beauchamp MK, et al. The Short Physical Performance Battery (SPPB): A Quick and Useful Tool for Fall Risk Stratification Among Older Primary Care Patients. *J Am Med Dir Assoc* 2021; 22(8): 1646–1651.
19. Sebastião E, Henert S and Siqueira V. Physical Activity and Physical Function in Older Adults Living in a Retirement Community: A Cross-Sectional Analysis Focusing on Self-Rated Health. *Am J Lifestyle Med* 2021; 15(3): 279–285. https://doi.org/10.1177/1559827620942720
20. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age and ageing* 2019; 48(1): 16–31. https://doi.org/10.1093/ageing/afz046
21. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ* 2021; 374: n2061. https://doi.org/10.1136/bmj.n2061
22. Hashemzadeh M, Rahimi A, Zare-Farashbandi F, et al. Transtheoretical Model of Health Behavioral Change: A Systematic Review. *Iran J Nurs Midwifery Res* 2019; 24(2): 83–90. https://doi.org/10.4103/ijnmr.IJNMR_94_17
23. Michie S, van Stralen MM and West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011; 6: 42. https://doi.org/10.1186/1748-5908-6-42
24. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013; 46(1): 81–95. https://doi.org/10.1007/s12160-013-9486-6
25. Gentili A, Failla G, Melnyk A, et al. The cost-effectiveness of digital health interventions: A systematic review of the literature. *Front Public Health* 2022; 10: 787135. https://doi.org/10.3389/fpubh.2022.787135
26. NHS. *Lose weight*. NHS. https://www.nhs.uk/better-health/lose-weight/?WT.mc_ID=PaidSearch_LoseWeight_Tools&wt.tsrc=paid_search&gclsrc=aw.ds&gad_source=1&gad_campaignid=23131764877&gbraid=0AAAAAC7M3f3w9jAOpDs2Wk4OzGcTRI_X5&gclid=CjwKCAjw687NBhB4EiwAQ645dol4ZiP6-BZf7STXhXtSjndQ6DRpBpWzd_X2V716JfTzeBXMr7VNYBoCBKQQAvD_BwE2026
27. NHS. *Get running with Couch to 5K*. NHS, 2026. Available from: https://www.nhs.uk/better-health/get-active/get-running-with-couch-to-5k/
28. NHS-England. *Fit for the Future: 10 Year Health Plan for England*. NHS England, 2025. https://www.england.nhs.uk/long-term-plan/
29. Walter SD, Eliasziw M and Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998; 17(1): 101–110. https://doi.org/10.1002/(sici)1097-0258(19980115)17:1<101::aid-sim727>3.0.co;2-e
30. Warburton DER, Jamnik VK, Bredin SSD, et al. The Physical Activity Readiness Questionnaire for Everyone (PAR-Q+) and Electronic Physical Activity Readiness Medical Examination (ePARmed-X+). *Health & Fitness Journal of Canada* 2011; 4(2): 3–23.
31. Jayedi A, Soltani S, Zargar MS, et al. Central fatness and risk of all cause mortality: systematic review and dose-response meta-analysis of 72 prospective cohort studies. *BMJ* 2020; 370: m3324. https://doi.org/10.1136/bmj.m3324
32. Ross R, Berentzen T, Bradshaw AJ, et al. Does the relationship between waist circumference, morbidity and mortality depend on measurement protocol for waist circumference? *Obes Rev* 2008; 9(4): 312–325. https://doi.org/10.1111/j.1467-789X.2007.00411.x
33. Liu Z, Yang H, Chen S, et al. The association between body mass index, waist circumference, waist–hip ratio and cognitive disorder in older adults. *Journal of Public Health* 2018; 41(2): 305–312. https://doi.org/10.1093/pubmed/fdy121
34. Ross R, Neeland IJ, Yamashita S, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. *Nat Rev Endocrinol* 2020; 16(3): 177–189. https://doi.org/10.1038/s41574-019-0310-7
35. Srikanthan P, Seeman TE and Karlamangla AS. Waist-hip-ratio as a predictor of all-cause mortality in high-functioning older adults. *Ann Epidemiol* 2009; 19(10): 724–731. https://doi.org/10.1016/j.annepidem.2009.05.003
36. Javed AA, Aljied R, Allison DJ, et al. Body mass index and all-cause mortality in older adults: A scoping review of observational studies. *Obes Rev* 2020; 21(8): e13035. https://doi.org/10.1111/obr.13035
37. Halaweh H. Correlation between Health-Related Quality of Life and Hand Grip Strength among Older Adults. *Exp Aging Res* 2020; 46(2): 178–191. https://doi.org/10.1080/0361073X.2020.1716157
38. Ling CH, Taekema D, de Craen AJ, et al. Handgrip strength and mortality in the oldest old population: the Leiden 85-plus study. *CMAJ* 2010; 182(5): 429–435. https://doi.org/10.1503/cmaj.091278
39. Bohannon RW. Hand-grip dynamometry predicts future outcomes in aging adults. *J Geriatr Phys Ther* 2008; 31(1): 3–10. https://doi.org/10.1519/00139143-200831010-00002
40. Cui M, Zhang S, Liu Y, et al. Grip Strength and the Risk of Cognitive Decline and Dementia: A Systematic Review and Meta-Analysis of Longitudinal Cohort Studies. *Front Aging Neurosci* 2021; 13: 625551. https://doi.org/10.3389/fnagi.2021.625551
41. Rikli RE and Jones CJ. Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. *J Aging Phys Act* 1999; 7(2): 129–161. https://doi.org/10.1123/japa.7.2.129

42. Rikli RE and Jones CJ. Functional Fitness Normative Scores for Community-Residing Older Adults, Ages 60-94. *J Aging Phys Act* 1999; 7(2): 162–181. https://doi.org/10.1123/japa.7.2.162
43. Rikli RE and Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. *Gerontologist* 2013; 53(2): 255–267. https://doi.org/10.1093/geront/gns071
44. Asai T, Oshima K, Fukumoto Y, et al. Association of fall history with the Timed Up and Go test score and the dual task cost: A cross-sectional study among independent community-dwelling older adults. *Geriatr Gerontol Int* 2018; 18(8): 1189–1193. https://doi.org/10.1111/ggi.13439
45. Shumway-Cook A, Brauer S and Woollacott M. Predicting the Probability for Falls in Community-Dwelling Older Adults Using the Timed Up & Go Test. *Phys Ther* 2000; 80(9): 896–903.
46. Ascencio EJ, Cieza-Gómez GD, Carrillo-Larco RM, et al. Timed up and go test predicts mortality in older adults in Peru: a population-based cohort study. *BMC Geriatr* 2022; 22(1): 61. https://doi.org/10.1186/s12877-022-02749-6
47. Chun S, Shin DW, Han K, et al. The Timed Up and Go test and the ageing heart: Findings from a national health screening of 1,084,875 community-dwelling older adults. *European Journal of Preventive Cardiology* 2019; 28(2): 213–219.
48. Bruun IH, Maribo T, Nørgaard B, et al. A prediction model to identify hospitalised, older adults with reduced physical performance. *BMC Geriatr* 2017; 17(1): 281. https://doi.org/10.1186/s12877-017-0671-5
49. Tiernan C, Schwarz DJ and Goldberg A. Associations of Usual and Fast Gait Speed With Physical Performance and Balance Confidence in Community-Dwelling Older Adults: Implications for Assessment. *J Geriatr Phys Ther* 2024; 47(4): 192–201. https://doi.org/10.1519/JPT.0000000000000397
50. Veronese N, Stubbs B, Volpato S, et al. Association Between Gait Speed With Mortality, Cardiovascular Disease and Cancer: A Systematic Review and Meta-analysis of Prospective Cohort Studies. *J Am Med Dir Assoc* 2018; 19(11): 981–988.e7. https://doi.org/10.1016/j.jamda.2018.06.007
51. Mielke MM, Roberts RO, Savica R, et al. Assessing the Temporal Relationship Between Cognition and Gait: Slow Gait Predicts Cognitive Decline in the Mayo Clinic Study of Aging. *The Journals of Gerontology: Series A* 2012; 68(8): 929–937. https://doi.org/10.1093/gerona/gls256
52. WHO. *Falls*. World Health Organisation, 2018. Available from: https://www.who.int/mediacentre/factsheets/fs344/en/
53. Organization WH. *WHO global report on falls prevention in older age*. World Health Organisation, 2008. Available from: https://www.who.int/publications/i/item/9789241563536
54. Melzer I, Benjuya N and Kaplanski J. Postural stability in the elderly: a comparison between fallers and non-fallers. *Age and ageing* 2004; 33(6): 602–607. https://doi.org/10.1093/ageing/afh218
55. Vaishya R and Vaish A. Falls in Older Adults are Serious. *Indian J Orthop* 2020; 54(1): 69–74. https://doi.org/10.1007/s43465-019-00037-x
56. Vellas BJ, Wayne SJ, Romero L, et al. One-Leg Balance Is an Important Predictor of Injurious Falls in Older Persons. *J Am Geriatr Soc* 1997; 45(6): 735–738. https://doi.org/10.1111/j.1532-5415.1997.tb01479.x
57. Welborn TA and Dhaliwal SS. Preferred clinical measures of central obesity for predicting mortality. *Eur J Clin Nutr* 2007; 61(12): 1373–1379. https://doi.org/10.1038/sj.ejcn.1602656
58. Dagenais GR, Yi Q, Mann JFE, et al. Prognostic impact of body weight and abdominal obesity in women and men with cardiovascular disease. *Am Heart J* 2005; 149(1): 54–60. https://doi.org/10.1016/j.ahj.2004.07.009
59. Wearing J, Konings P, Stokes M, et al. Handgrip strength in old and oldest old Swiss adults - a cross-sectional study. *BMC Geriatr* 2018; 18(1): 266. https://doi.org/10.1186/s12877-018-0959-0
60. Rikli RE and Jones CJ. Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. *J Aging Phys Act* 1999; 6: 127–159.
61. Omaña H, Bezaire K, Brady K, et al. Functional Reach Test, Single-Leg Stance Test, and Tinetti Performance-Oriented Mobility Assessment for the Prediction of Falls in Older Adults: A Systematic Review. *Phys Ther* 2021; 101(10): pzab173. https://doi.org/10.1093/ptj/pzab173
62. Blodgett JM, Hardy R, Davis D, et al. One-Legged Balance Performance and Fall Risk in Mid and Later Life: Longitudinal Evidence From a British Birth Cohort. *Am J Prev Med* 2022; 63(6): 997–1006. https://doi.org/10.1016/j.amepre.2022.07.002
63. Richardson JT. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 2011; 6(2): 135–147. https://doi.org/10.1016/j.edurev.2010.12.001
64. Hopkins WG, Marshall SW, Batterham AM, et al. Progressive statistics for studies in sports medicine and exercise science. *Med Sci Sports Exerc* 2009; 41(1): 3–13. https://doi.org/10.1249/MSS.0b013e31818cb278
65. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005; 19(1): 231–240. https://doi.org/10.1519/15184.1
66. Koo TK and Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016; 15(2): 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

67. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol* 2013; 4: 863. https://doi.org/10.3389/fpsyg.2013.00863
68. Beauchamp MK, Hao Q, Kuspinar A, et al. Reliability and Minimal Detectable Change Values for Performance-Based Measures of Physical Functioning in the Canadian Longitudinal Study on Aging. *The Journals of Gerontology: Series A* 2021; 76(11): 2030–2038. https://doi.org/10.1093/gerona/glab175
69. Barrios P, Martin-Biggers J, Quick V, et al. Reliability and criterion validity of self-measured waist, hip, and neck circumferences. *BMC medical research methodology* 2016; 16: 49. https://doi.org/10.1186/s12874-016-0150-2
70. Benton MJ, Spicher JM and Silva-Smith AL. Validity and reliability of handgrip dynamometry in older adults: A comparison of two widely used dynamometers. *PLoS One* 2022; 17(6): e0270132. https://doi.org/10.1371/journal.pone.0270132
71. Christopher A, Kraft E, Olenick H, et al. The reliability and validity of the Timed Up and Go as a clinical tool in individuals with and without disabilities across a lifespan: a systematic review. *Disabil Rehabil* 2021; 43(13): 1799–1813. https://doi.org/10.1080/09638288.2019.1682066
72. Bohannon RW and Wang Y-C. Four-Meter Gait Speed: Normative Values and Reliability Determined for Adults Participating in the NIH Toolbox Study. *Arch Phys Med Rehabil* 2019; 100(3): 509–513. https://doi.org/10.1016/j.apmr.2018.06.031
73. Mehmet H, Robinson SR and Yang AWH. Assessment of Gait Speed in Older Adults. *J Geriatr Phys Ther* 2020; 43(1): 42–52. https://doi.org/10.1519/JPT.0000000000000224
74. Stuck AK, Bachmann M, Füllemann P, et al. Effect of testing procedures on gait speed measurement: A systematic review. *PLoS One* 2020; 15(6): e0234200. https://doi.org/10.1371/journal.pone.0234200
75. Brognara L. Gait Assessment Using Smartphone Applications in Older Adults: A Scoping Review. *Geriatrics (Basel)* 2024; 9(4): 95. https://doi.org/10.3390/geriatrics9040095
76. Goldberg A, Casby A and Wasielewski M. Minimum detectable change for single-leg-stance-time in older adults. *Gait Posture* 2011; 33(4): 737–739. https://doi.org/10.1016/j.gaitpost.2011.02.020
77. Rhea MR, Landers DM, Alvar BA, et al. The effects of competition and the presence of an audience on weight lifting performance. *J Strength Cond Res* 2003; 17(2): 303–306. https://doi.org/10.1519/1533-4287(2003)017<0303:teocat>2.0.co;2
78. Connell L, Ackerley S and Rycroft-Malone J. Applying the updated MRC framework for developing and evaluating complex interventions with integrated implementation conceptual knowledge: an example using NeuroRehabilitation OnLine. *Frontiers in Health Services* 2025; 2025: 5. https://doi.org/10.3389/frhs.2025.1562627
79. Hesseberg K, Bentzen H and Bergland A. Reliability of the senior fitness test in community-dwelling older people with cognitive impairment. *Physiother Res Int* 2015; 20(1): 37–44. https://doi.org/10.1002/pri.1594
80. Kotsifaki A, Whiteley R, Van Rossom S, et al. Single leg hop for distance symmetry masks lower limb biomechanics: time to discuss hop distance as decision criterion for return to sport after ACL reconstruction? *Br J Sports Med* 2022; 56(5): 249–256. https://doi.org/10.1136/bjsports-2020-103677
81. Labban W, Manaseer T, Golberg E, et al. Jumping into recovery: A systematic review and meta-analysis of discriminatory and responsive force plate parameters in individuals following anterior cruciate ligament reconstruction during countermovement and drop jumps. *Journal of Experimental Orthopaedics* 2024; 11(2): e12018. https://doi.org/10.1002/jeo2.12018
82. Bishop C, Turner A, Jordan M, et al. A Framework to Guide Practitioners for Selecting Metrics During the Countermovement and Drop Jump Tests. *Strength & Conditioning Journal* 2022; 44(4): 95–103. https://doi.org/10.1519/ssc.0000000000000677

Appendix

Appendix 1. Online survey questions

1. Participant Number Please enter your participant number.

2. Computer Literacy *(10-point Likert scale)* How would you rate your computer literacy? (1 = Poor, 10 = Excellent)

3. Perceived Value of Functional Fitness Monitoring *(10-point Likert scale)* How valuable do you perceive functional fitness monitoring for understanding healthy ageing? (1 = Not at all valuable, 10 = Extremely valuable)

4. Previous Engagement in Functional Fitness Monitoring Prior to participation in this study, did you monitor your functional fitness? (Yes/No)

  4a. Previous Monitoring Methods *(Free-text response)* If you answered yes, could you please describe how you have monitored your functional fitness?

5. Ease of Navigation of the FFMT Information Document *(10-point Likert scale)* The Functional Fitness Assessment and Monitoring information document was easy to navigate. (1 = Strongly Disagree, 10 = Strongly Agree)

6. Ease of Understanding the FFMT Information Document *(10-point Likert scale)* The information contained within the Functional Fitness Assessment and Monitoring information document was easy to understand. (1 = Strongly Disagree, 10 = Strongly Agree)

7. Ease of Understanding the FFMT Report *(10-point Likert scale)* The report generated from my data was informative and easy to understand. (1 = Strongly Disagree, 10 = Strongly Agree)

8. Requirement for Support During Self-Assessment I needed some support to complete some or all of the functional fitness self-assessments. (Yes/No)

8a. Support Requirements *(Free-text response)* If you needed some support, please let us know what support you needed and from whom.

9. Perceived Competence for Self-Assessment *(10-point Likert scale)* I felt that I was able to competently self-assess my functional fitness, even if some support was required. (1 = Strongly Disagree, 10 = Strongly Agree)

10. Usefulness for Long-Term Monitoring Do you think this tool would be useful to track your performance over time? (Yes/No/Unsure)

10a. Reasons for Perceived Lack of Usefulness *(Free-text response)* If you do not think it would be useful, please tell us why.

10b. Perceived Usefulness for Long-Term Monitoring *(10-point Likert scale)* How useful do you think this tool would be in tracking your performance over time? (1 = Not at all Useful, 10 = Very Useful)

11. Interest in Future Use Would you be interested in using this tool regularly to track your physical performance? (Yes/No/Unsure)

11a. Reasons for Lack of Interest *(Free-text response)* If you would not be interested in using this tool regularly, please tell us why.

11b. Interest in Future Use *(10-point Likert scale)* How interested would you be in regularly using this tool to track your physical performance? (1 = Not at all Interested, 10 = Very Interested)

12. Requirements for Ongoing Self-Monitoring *(Free-text response)* To carry out regular functional self-assessments, I would need…

Please use this space to inform us of anything you would need or want in order to perform these assessments regularly. This could be equipment, support, motivation, reminders, professional guidance, or anything else you feel would be required.

13. Suggestions for Improvement *(Free-text response)* Do you have any suggestions on how we could improve the digital tools provided for Functional Fitness Assessment and Monitoring?

14. Additional Comments *(Free-text response)* Do you have any other comments or suggestions?