

# Automated fall risk assessment of elderly using wearable devices

Marian Haescher<sup>1,2</sup> , Wencke Chodan<sup>1</sup>, Florian Höpfner<sup>1</sup>, Gerald Bieber<sup>1</sup> , Mario Aehnelt<sup>1</sup>, Karthik Srinivasan<sup>3</sup> and Margit Alt Murphy<sup>4</sup>

Journal of Rehabilitation and Assistive Technologies Engineering  
Volume 7: 1–13  
© The Author(s) 2020  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/2055668320946209  
journals.sagepub.com/home/jrt



## Abstract

**Introduction:** Falls cause major expenses in the healthcare sector. We investigate the ability of supporting a fall risk assessment by introducing algorithms for automated assessments of standardized fall risk-related tests via wearable devices.

**Methods:** In a study, 13 participants conducted the standardized 6-Minutes Walk Test, the Timed-Up-and-Go Test, the 30-Second Sit-to-Stand Test, and the 4-Stage Balance Test repeatedly, producing 226 tests in total. Automated algorithms computed by wearable devices, as well as a visual analysis of the recorded data streams, were compared to the observational results conducted by physiotherapists.

**Results:** There was a high congruence between automated assessments and the ground truth for all four test types (ranging from 78.15% to 96.55%), with deviations ranging all well within one standard deviation of the ground truth. Fall risk (assessed by questionnaire) correlated with the individual tests.

**Conclusions:** The automated fall risk assessment using wearable devices and algorithms matches the validity of the ground truth, thus providing a resourceful alternative to the effortful observational assessment, while minimizing the risk of human error. No single test can predict overall fall risk; instead, a much more complex model with additional input parameters (e.g., fall history, medication etc.) is needed.

## Keywords

Fall risk assessment, fall prevention, wearable computing, mobile assisted living (MAL), ambient assisted living (AAL)

Date received: 25 February 2020; accepted: 10 July 2020

## Introduction

Many care facilities, such as nursing homes or geriatric wards, experience high costs due to follow-up treatments related to injuries caused by fall.<sup>1</sup> Nursing homes experience around 1.5 falls per bed per year,<sup>2</sup> resulting in average costs of 944 Euro per fall.<sup>3</sup> Approximately 25% to 46% of all patients in stroke rehabilitation wards have been reported to fall at least once during their stay.<sup>4–6</sup> Alarming, 75% of deaths in the age group of 65+ in the USA are caused by falls,<sup>2</sup> thus concerning 13% of the country's population. In different regions including Europe, USA, and Australia, between 0.9% and 1.5% of the total health care expenditures of each region are related to fall.<sup>7</sup> About 0.2% of the USA's gross domestic product (547 USD PPP) is paid for costs related to fall per inhabitant per year (ib.).

Factors that contribute to the risk of falling include injuries or pathologic conditions (e.g., hip fracture, stroke) or even unfamiliar environments and

<sup>1</sup>Fraunhofer Institute for Computer Graphics Research IGD, Competence Center Visual Assistance Technologies, Rostock, DE, Germany

<sup>2</sup>Institute for Visual and Analytic Computing, Department of Multimedia Communication, University of Rostock, DE, Germany

<sup>3</sup>Next Step Dynamics AB, SE, Germany

<sup>4</sup>Institute of Neuroscience and Physiology, Department of Clinical Neuroscience, Rehabilitation Medicine, Sahlgrenska Academy, University of Gothenburg, SE, Germany

### Corresponding author:

Marian Haescher, Fraunhofer IGD, Joachim-Jungius-Str. 11, 18059 Rostock, Germany.  
Email: marian.haescher@igd-r.fraunhofer.de



medication.<sup>8,9</sup> A literature review investigated risk factors in a variety of settings, and found gait instability, agitated confusion, urinary incontinence and urinary frequency, and falls history to be most relevant.<sup>9</sup>

Many review articles emerged on fall risk prevention in recent years.<sup>10,11</sup> While it is prudent to identify the individual's fall risk first to allocate resources according to needs, less research has been conducted on the assessment of fall risk. The last systematic reviews on assessment tools for fall risk have included studies up to 2016.<sup>12–14</sup> Lusardi et al.<sup>13</sup> identified five performance-based measures, including the Timed Up and Go test ( $\geq 12$  seconds) and Sit-to-Stand test, that were rated as “currently the most evidence-supported functional measures to determine individual risk of future falls” in ambulatory older adults.<sup>13</sup> Aranda-Gallardo et al.<sup>12</sup> conducted a meta-analysis on three fall risk assessment tools used for prediction of falls in acute hospitalized patients. Unfortunately, the authors did not include any behavioural measures, and the validity of the three questionnaires that the review compared varied considerably depending on the population and the environment.<sup>12</sup>

The manual assessment using standardised physical tests is an established procedure for the fall risk estimation. Muscle strength, static and dynamic balance (postural control), walking behaviour and mobility are often examined and provide quantifiable parameters that are commonly used in combination with questionnaires and anamnestic information to determine the likelihood of fall risk. Since manual tests are time-consuming and allow for human error and bias, a technology-assisted assessment would be preferred.

The possibility of assessing human behaviour and classifying human motions has been extensively shown in the field of human activity recognition.<sup>15–18</sup> Within the broad spectrum of activity recognition, behaviours that are important for fall risk detection, such as gait analysis,<sup>9</sup> can also be detected using wearable sensors,<sup>19</sup> including smartwatches.<sup>20–22</sup> Haescher et al.<sup>1</sup> investigate fall risk prevention using a wrist-worn smartwatch and present a multifactorial model that take into account basal fall risk (using questionnaires, medical history, e.g., asking for the need of a permanent walking aid), environmental fall risk (challenging outdoor conditions such as weather, familiarity of the environment), and variable fall risk (assessed by manual tests, such as the Timed Up and Go test).<sup>1</sup>

Thus, our overall ambition is to develop an objective and automated fall-risk assessment by using wearable technology in combination with a multi-parametrical model to provide an estimate of fall risk and in this way could be used to support fall prevention in elderly people. The objective of the study was to investigate to what degree the mobile wrist-worn device, equipped

with sensors, provides comparable and accurate results when compared to observational clinical assessments in a group of older adults living in the nursing home.

## Method

### Participants

In this cross-sectional study, participants were recruited from a nursing home in west of Sweden, out of a pool of attendees of the routine balance training program. Data collection took place between April and June, 2019. Inclusion criteria were: (i) age  $\geq 65$  years; (ii) able to perform the clinical assessment tests without any walking aid. To clarify, this did not exclude any individual who would use a walking aid in everyday life; it merely means that the walking aid was not used during the tests. Exclusion criteria comprised Parkinson's disease and/or any other disease that might hinder the performance of the selected assessments or compromise the safety of the patient during assessments. Written informed consent was obtained from all participants or participant's legal representative. Participants did not receive any compensation for their participation. The Declaration of Helsinki for medical research was followed, but no ethical approval was applied prior the study since the data was collected during the clinical routine and handled fully anonymized by the research team (authors of the study).

In total, 13 participants were included (6 male, 7 female). The mean age of the participants was 79.61 years (SD 4.55 yrs). The participants had a mean weight of 71.23 kg (SD 15.07 kg). The average height of the participants was 168.31 cm (SD 8.95 cm).

### Standardized clinical assessments

The data was collected during the ordinary balance training session according to clinical routine. The participants performed four standardized tests in a controlled clinical environment: 6-Minutes Walk Test (6MWT),<sup>23</sup> Timed Up and Go Test (TUG),<sup>24</sup> 30-Second Sit-to-Stand Test (STS, also called 30-Second Chair Stand Test, 30STS),<sup>25,26</sup> and 4-Stage Balance Test (4SBT).<sup>27–29</sup> The tests were performed by a physiotherapist who instructed the participants, observed the tests, and recorded the observed test results in means of time via stopwatch or repetitions via counting. The participants performed the tests up to seven times with a break of at least seven days in between. Physiotherapist were always close to the participants to avoid accidental loss of balance or fall during testing.

**6-Minutes Walk Test (6MWT).** In this test, participants were instructed to walk as far as possible during six

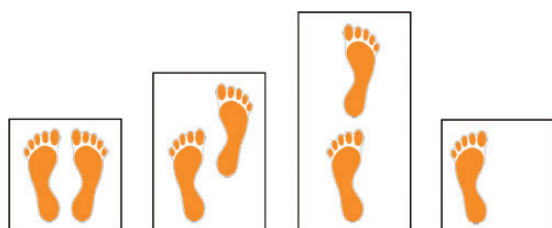
minutes. The walking distance was marked by two turning points 20 m apart from each other. After six minutes of circling the turning points, the participant was asked to stop, and the number of turns was counted and multiplied by the distance (20 m). The remaining distance walked (between the turning points) was measured by using a tape measure.

**Timed up and Go Test (TUG).** In this test, participants had to stand up and walk a distance of 3 meters. At the 3-meter mark, participants made a turn and went back to the starting point in order to sit down again. The time taken for the whole test was measured using a stopwatch.

**Sit-to-Stand Test (30STS).** In this test, the participants had to transition between standing up and sitting down as often as possible during a 30-second period. During the test, the arms had to be crossed in front of the chest. This way, the ability of standing up without using the arms was tested. The number of transitions within 30 seconds was counted by the therapist. The 30-second limit was controlled via a stopwatch.

**4-Stage Balance Test.** In this test, the participants had to stand as still as possible in four different postures (see Figure 1) for the duration of 10 seconds each (measured with stopwatch). The postures were ordered hierarchically starting with the easiest (standing feet together) and ending with the most difficult (standing on one leg). Participants could choose freely which foot to position forward or upward in the three postures ‘half step’, ‘tandem’ and ‘balancing on one foot’ (see Figure 1). In case a participant lost balance in one of the postures, the test was stopped and marked as failed. All remaining postures were marked as failed as well.

For the detection of the different postures with the wearable, the subjects were asked to perform a clap at



**Figure 1.** The Balance Test involves four different feet positions (postures) that have to be performed for ten seconds each, from left to right: a) standing with feet together, b) half step, c) tandem, d) balancing on one foot. For the detection of the different postures with the wearable, the subjects were asked to perform a clap at the beginning and end of each posture. In case of a failure, the participant was asked to clap twice.

the beginning and end of each posture. In case of a failure, the participant was asked to clap twice.

### Apparatus

The participants performed the standardized clinical tests while wearing a CE-certified smartwatch (Huawei smart watch 2) at their left wrist to enable an automated assessment (see Figure 2). The watch recorded raw data of the acceleration sensor and gyroscope. The data was sampled at 50 Hz. The smartwatch data was sent to a cloud platform for processing the results. An app was programmed to provide an easy-to-use interface for the therapists (see an example in Figure 2).

At the beginning of every recording session, the physiotherapist selected an anonymized 6-digit identifier for each participant. Subsequently, each test was selected by choosing the particular test icon on the wearable screen (see Figure 2). After selecting the test, a start button was pushed to initiate the recording of accelerometer and gyroscope raw data. As soon as the patient finished a test and the physiotherapist had logged the ground truth, a stop button was pressed in order to stop the recording and upload the data.

### Fall risk

All test results were analysed with regard to their inference on the individual fall risk. To this end, the fall risk for both the observed test results recorded by the physiotherapist and for the automated algorithm-based test results was classified into a binary scoring of “high” versus “low” risk of falling, referring to each test’s standardized cut-off values. In both cases, defined cut-off



**Figure 2.** The Huawei Watch 2 worn on the left wrist, showing the interface. The interface shows four icons, one icon for each of the four tests. 1) 4SBT (upper left icon), 2) TUG (upper right icon), 3) 30STS (lower left icon), 4) 6MWT (lower right icon).

points described in the literature were used.<sup>30–33</sup> For the 6MWT, a distance below the reference values (adjusted for age and sex) indicated an increased fall risk.<sup>30,31</sup> Each participant who took  $\geq 12$  seconds to perform the TUG was scored with increased risk of falling.<sup>30</sup> For the 30SCT, reference values adjusted for age and sex for number of sit-to-stand were used to define the increased risk of falling.<sup>31</sup> Each participant who failed to perform the third balance position (tandem standing, 4SBT) for ten seconds was scored with increased risk of falling.<sup>32,33</sup>

In addition, the physiotherapists appraised the risk of falling as either “low” or “high” based on the checklist of risk for falling,<sup>34</sup> originally developed as the Fall Risk Questionnaire (FRQ).<sup>35</sup> A score of 4 points or more on the FRQ indicated a high risk of falling. This label was used as a theoretical ground truth to which both the observational data and the data derived by the algorithm were compared to assess each of their convergent validity.

### Data analysis

We conducted three types of data analysis for the data assessed with the wearable device: 1) visual data analysis, 2) automated data analysis (via an algorithm), 3) automated data analysis after excluding participants that did not follow the protocol (outliers). Data preprocessing included converting the movement signal to a representation in the frequency domain, followed by a peak detection. Visual analysis was performed by having at least two data scientists analyse the recorded raw data, thriving for consensus between interrater judgements. In the automated data analysis, the algorithm, which was developed for automatic detection of the relevant parameters, was used. The algorithm includes of Butterworth filtering for reducing noise and artefacts and enables the automatic detection of the start position and the end position of the test as well as relevant parameters identified for each test (e.g., the turning points in the 6MWT). The last method used the aforementioned algorithm, which was applied to a data subset that had been cleaned of outliers by a data analyst.

Data from visual and automated analysis (via an algorithm) were compared to observational testing (ground truth, recorded by the physiotherapists who conducted the tests). The accuracy and mean deviation from the ground truth were calculated for three measures: distance walked during the 6MWT, time taken to perform the TUG, and number of transitions performed during the 30SCS. For these three tests, descriptive properties were calculated as the main analysis.

When comparing the fall risk classified by using the defined cut-off points of the clinical standardized tests, the observational data was used as ground truth and the data derived by the algorithm was tested against the observational data.

Due to an error in designing the interface for the physiotherapists, the final scoring of the 4SBT was binary (either having passed the fourth position (standing on one leg) or not, with no data available on positions 2 and 3). One-leg standing balance have shown to be a predictive marker for injurious falls in elderly population.<sup>29,35</sup> Thus, the accuracy was computed for each test between the physiotherapist’s observation and the algorithm’s appraisal, and an F1 score was computed for the binary result of the 4SBT. The F1 score is the harmonic mean of precision and sensitivity. It is a measure of accuracy, considering both precision (positive predictive value) and recall (sensitivity). The score ranges from its best value at 100% (perfect precision and recall) to its worst value at 0%.

Additional variables derived from sensor data are presented descriptively, since there is no data from observational testing that can be used for comparison. For the 6MWT, step length, number of turns at the cones, step impact, and walking endurance were computed. Step impact indicates the speed of treading, i.e., how fast the foot hits the ground and endurance factor, computed as slope ( $\Delta t$ ), indicates the speed changes over the course of the six minutes. For the TUG, time for standing up, time for sitting down, and, while walking, the number of steps, step length, and step impact were calculated. A lower number of steps might be a result of a higher step length, indicating more gait stability. For 30STS, the number of transitions (sit-to-stand and stand-to-sit) and the durations for the transitions (i.e., the time it took to stand up and the time it took to sit down) were obtained. For the 4SBT, the failing or passing of each of the four postures and the duration of holding each posture, as well as the compensatory arm movement during each posture were calculated. The arm movement serves to keep the balance, as is it used to equalize the movement caused by instability. Stability can thus be defined as the reverse of the equalizing motion.

### Results

A total of 226 individual tests were recorded across the 13 participants. Each participant repeated each test between one and eight times with an average of 4.35 times (SD 1.82 times), uniformly distributed across the four test types (6MWT, 4SBT, TUG, 30STS). Due to eight missings in the handwritten records of the physiotherapists and two missings that resulted from

forgetting to start the recording of the smartwatch, a total of 216 test results were used for data analysis.

**Accuracy of visual and automated data analysis.** The accuracy and mean deviation based on the visual data analysis, automated data analysis (via an algorithm) and automated data analysis after excluding outliers for 6MWT, the TUG and the 30STS are shown in Table 1. Due to the above-mentioned error, there was no sufficient ground truth data available for a similar analysis for the 4SBT.

With regard to visual raw data analysis, we were able to achieve an accuracy of 70–98% (depending on the test) in comparison to the physiotherapists' clinical assessments (Table 1). An accuracy of 86–97% was reached when the automated algorithm was used.

### 6-Minutes walk test

An example of the Gyroscope data recorded with the smartwatch during one complete 6MWT is shown in Figure 3. Exemplary, the Figure shows the gyroscope data that corresponds to the accelerometer's x-axis, but the pattern can be seen in all rotations around all three axes. The 18 high amplitude blocks indicate the

walking and the spikes between the blocks signify the turns. The last (19th) block is shorter, indicating the remaining distance until the test was stopped after exactly 6 minutes.

The visual analysis of the 6MWT data showed high accuracy (97.57%). Only two tests had an error due to a hardware malfunction and were subsequently eliminated from analysis. The automated evaluation of the 6MWT provided an accuracy of 90.75% for the whole data set (including outliers) and 96.55% (excluding outliers). Outliers include participants who did not follow the instruction of walking in circles but instead crossed the middle line and walked in the shape of an eight ( $\infty$ ) around the two markers, as did the participant shown in Figure 3. This led to walking the wrong way around a cone once, which results in a flip of the regarding peak (see Figure 4).

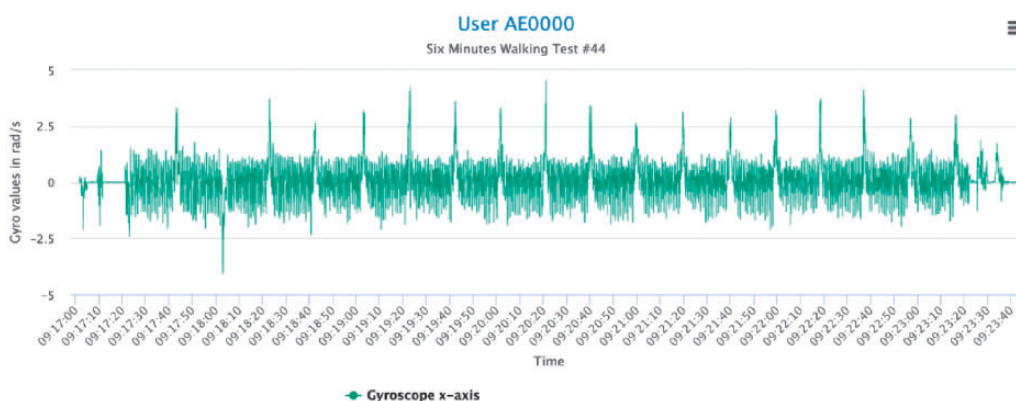
Further descriptive analysis reveals that the automated analysis differs from the ground truth by 30.50 m on average. The difference decreases to 17.64 m when excluding the outliers. This equals the difference between the visual analysis and the ground truth (16.68 m), meaning that the algorithm performs as well as an experienced professional rating the raw data. Moreover, an additional analysis that did not use

**Table 1.** Results for visual data analysis, automated data analysis, and automated data analysis after removing the outliers.

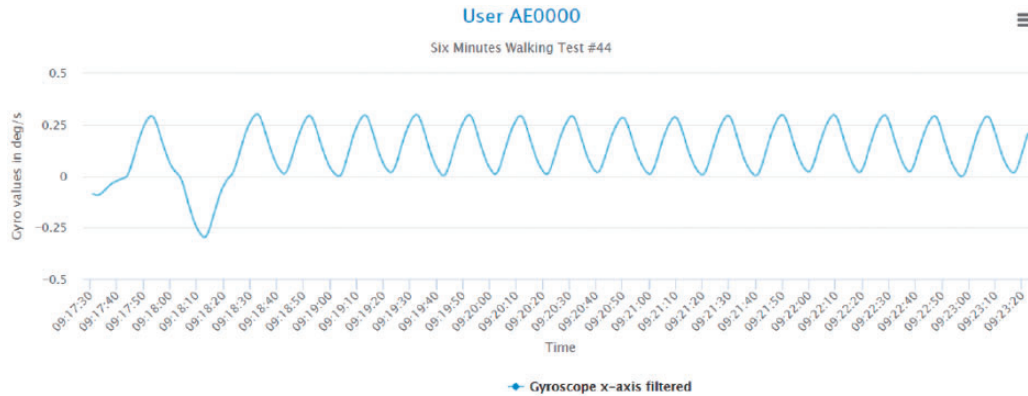
Test	Results					
	Visual Data Analysis		Automation Algorithm (on whole data set)		Automation Algorithm on valid data only (excluding outliers)	
	Accuracy	Mean Difference	Accuracy	Mean Difference	Accuracy	Mean Difference
6MWT	97.57%	2.31%	90.75%	9.67%	96.55%	3.05%
TUG	69.22%	21.16%	84.64%	11.00%	86.42%	9.23%
30STS	91.78%	7.16%	78.15%	14.28%	78.15% <sup>a</sup>	14.28% <sup>a</sup>

Each was compared to the ground truth, i.e., the physiotherapist's observations.

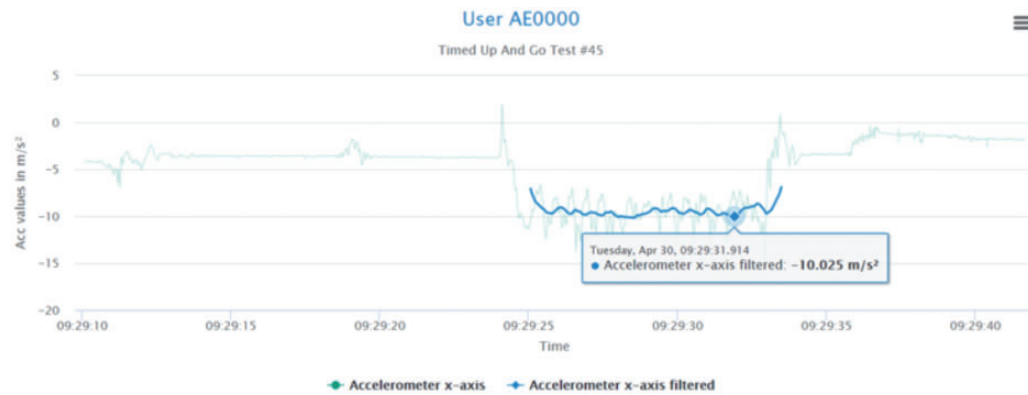
<sup>a</sup>No outliers identified.



**Figure 3.** Gyroscope signal of a complete 6-Minutes Walk Test. This pattern of longer segments indicating straight walking separated by short segments indicating turns is observed in all three rotations.



**Figure 4.** Turning points of the 6-Minutes Walk Test (6MWT). The signal is recorded by a gyroscope. The participant changed turning directions once at 09:18:10, as visible in the graph by the peak that is flipped upside down.



**Figure 5.** Filtered and cropped signal of accelerometer TUG recording. The starting event is getting up, whereas the stopping event is sitting down.

absolute numbers showed that the deviations neither tend to systematically over- or underestimate the ground truth, as the deviation decreased to 5.08 m when overestimations and underestimations were juxtaposed.

As for the fall risk, the 6MWT showed an accuracy of 94.64% for the comparison between the physiotherapist's observations and the fall risk determined by the automatic algorithm. The sensitivity reached 97.29% whereas the specificity reached 89.47%. The  $F_1$  score achieved 96%.

### Timed up and go test

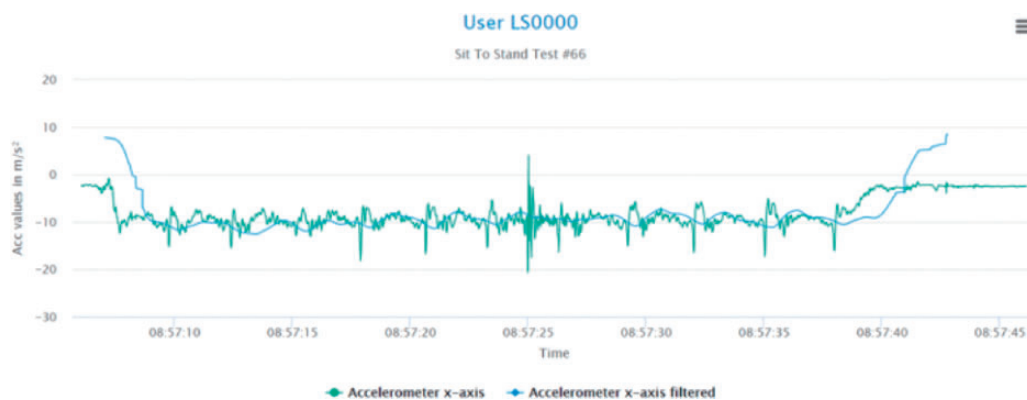
Figure 5 shows the process of cropping and filtering the raw data (accelerometer data). After the pre-processing, the total time taken for the test was detected. The time taken to stand up and sit down was calculated since the process of standing up and sitting down is visible in the orientation with regard to the earth's gravity.

The results of the automated analysis differ by 1.75 seconds from the ground truth. This low deviation is reduced even further when excluding outliers (yielding a difference of 1.24 s from the ground truth). As the visual analysis of the recorded data differed from the ground truth by 3.01 seconds on average, the algorithm outperforms an experienced professional in analysing the recorded data. This result explains the lower accuracy reported in Table 1. Despite the deviations, however, the accuracy still lies at 69.22% for visual analysis and above 84% for the automated analysis.

As for the fall risk, the TUG showed an accuracy of 94.64% for the comparison between increased fall risk determined by the physiotherapists and the automatic algorithm. The sensitivity reached 77.77% whereas the specificity reached 97.87%. The  $F_1$  score achieved 82.35%.

### 30-Second sit-to-Stand test

Figure 6 shows the result of a 30-Second Sit-to-Stand Test. The graph shows the accelerometer raw data



**Figure 6.** Acceleration sensor data of a 30-Second Sit-to-Stand Test (LS0000).

(green graph) and the filtered and cropped data (blue graph).

The visual analysis of the recorded data shows a high congruence with the ground truth (91.78%). The automated data analysis and the ground truth share a lower congruence (78.15%), as they differ on average by 1.79 sit-to-stand-transitions. The recorded data allows for a better analysis, as the visual analysis only differs by 0.68 transitions from the ground truth.

As for the fall risk, the 30STS showed an accuracy of 76.78% for the comparison between increased fall risk determined by the physiotherapists and the automatic algorithm. The sensitivity reached 86.20% whereas the specificity reached 66.66%. The  $F_1$  score achieved 79.36%.

#### 4-Stage Balance Test

Since the ground truth of the Balance Test (4SBT) is binary, we chose the F1 score over the mean deviation. An F1 score of 86.59% was achieved. The congruence between the automated analysis and the ground truth was higher for the first part of the 4SBT (namely, 92.73%) than for the fourth part (78.18%). Moreover, it was exceeded by the congruence of the visual analysis with the ground truth (100% for the first part, 89.09% for the fourth part), meaning the recorded data is a promising basis for accurate analyses when improving the algorithm.

No comparison between the visual, automated and observational analysis could be computed due to missing ground truth (observational data). An error in the interface of the digital evaluation sheets only allowed the physiotherapists to mark whether a participant had passed or failed the first posture and whether a participant had passed or failed the last posture. No information was given on posture two and three, the latter being used for determining the fall risk. Since no participant had failed the first posture and only in 20 percent of the balance tests posture four was passed,

sample size was too small to conduct inferential analyses on the informative cases only. Instead, additional variables have been identified and extracted from the recordings (see section on additional parameters).

As for the fall risk, the 4SBT showed an accuracy of 76.78% for the comparison between increased fall risk determined by the physiotherapists and the automatic algorithm. The sensitivity reached 95.45% whereas the specificity reached 8.33%. The  $F_1$  score achieved 86.59%.

#### Additional information provided by the recorded smartwatch data

The additional variables obtained from the smartwatch during the clinical tests, for which no ground truth data was available, are presented descriptively in Table 2. Data is presented for a prototypical representative for each of the two classifications of fall risk: one participant with a low risk of falling and one participant with a high risk of falling risk (as classified by the questionnaire).

The participants with high risk of falling demonstrated shorter step length, lower step impact and had a larger decrease of speed over the 6MWT, indicating a lower endurance factor during the 6MWT than the participants with low risk. In addition, in the TUG and 30STS, the participants with high risk of falling showed longer time in transfers between sitting and standing. In the 4SBT, the amount of arm compensatory movement was larger for the participants with high risk of falling.

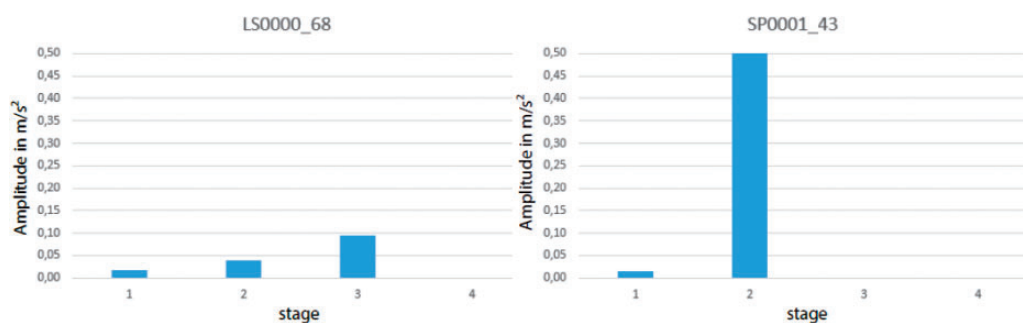
Figure 7 illustrates the equalizing motion for two participants while holding positions that were both simply marked as “passed the first posture” in the physiotherapist’s observation. Participant LS (low risk of falling according to the questionnaires) passed the first three postures while participant SP (high risk of falling according to the questionnaires) only passed the first two postures. In the recorded data, we can see that

**Table 2.** Additional parameters computed for a participant with low risk of falling (LS) and a participant with high risk of falling (SP) in the four clinical tests during one testing session.

Results			
Test	Parameter	Participant LS has not (low risk)	Participant SP (high risk)
6MWT	Distance in m	492.44	300.19
	Number of steps	784	618
	Step length in m	0.63	0.48
	Step impact in $m/s^2$	4.48	4.37
	Number of turns	24	14
	Endurance factor (slope, $\Delta t$ ) <sup>a</sup>	0.11	0.20
TUG	Stand up duration in s	0.9	1.5
	Sit down duration in s	0.7	1.9
	Number of steps	14	19
	Step length in m	0.5	0.4
	Step impact in $m/s^2$	4.32	4.07
30STS	Number of transitions “sit to stand”	11	7
	Number of transitions “stand to sit”	12	6
	Stand up duration in s	1.0	1.9
	Sit down duration in s	1.1	1.9
4SBT <sup>b</sup>	Arm movement (variance averaged over all 4 tests)	0.12	4.31

<sup>a</sup>The factor denotes how quickly the participant loses speed; the smaller the value, the higher the endurance.

<sup>b</sup>Whether each test part was passed or failed in the 4SBT cannot be listed in the table, as this information was not logged in the study due to an error in implementation.

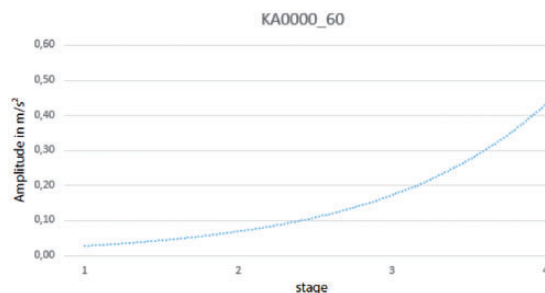


**Figure 7.** Comparison of two Balance Test results (LS0000) vs. (SP0001). The graph shows an increase in variance (motion scoring) for participant LS0000 and a fail in the last posture. Participant SP0001 has a much higher increase in variance (motion score) between posture one and two and failed in the third posture. Equalizing arm motion increases with the more difficult postures, as illustrated in Figure 8 (data of participant KA who completed all four postures), which shows a continuous increase in motion within the four parts of the 4SBT.

participant LS used much less equalizing arm motion to balance than participant SP in the second posture, which serves as an additional indicator for fall risk.

## Discussion

The goal of this study was to evaluate if four individual tests can be conducted in a digital system. In this study, after removing the outliers, the 6MWT could be automated with an accuracy of 97%, the TUG with an accuracy of 85%, and the 30STS with an accuracy of 78%. To ensure the algorithm neither underestimates



**Figure 8.** Change of three axes variance (motion score) over the four balance postures (KA0000). As the level of difficulty is increasing, a rise of variance is measured.



nor overestimates the participant's fall risk, we tested whether the algorithm yields results comparable to the physiotherapist's observations. Descriptive analyses confirmed this balance for the 30SCS and the TUG, but showed a trend to overestimate the results of the 6MWT (though this trend diminished once outliers were excluded). The 6MWT and the 30STS showed no significant differences compared to the classification based on the physiotherapist's observations. The TUG did yield a significant difference between automated detection and the physiotherapist's observations, despite its comparably high accuracy and despite the low average deviation of 1.24 seconds, which lies well within the standard deviation of the ground truth (3.28 seconds). In the analysis of the motion recordings, it was difficult to determine the beginning and end of the test (in case the user did not sit still, which was annotated by the physiotherapists for 5 incidences) as well as the moment of "sitting down" during the test, as the algorithm captures the moment the user sits still, not when the user touches the chair.

A computation of accuracy and accordance was not possible for the 4SBT, since the ground truth data was incomplete, and the additional task involved in the test procedure (clapping between the tests and double clap in case of a failed or missed test) was too demanding for the participants. The digital tests provide the potential to compute additional parameters for evaluating the individual fall risk that could not be recorded before. In addition to this, the digital test is objective and independent from the human bias that might occur in classical settings with physiotherapists.

### **6-Minutes Walk Test (6MWT)**

Very high accuracy was achieved for the visual analysis of the 6MWT data (98%). The accuracy, however, was lower for the automated analysis (91%). This was mainly caused by the two participants who did not walk in circles, but instead crossed the midline between turning points (walked in the figure of eight;  $\infty$ ). When excluding these outliers, the performance of the automated analysis increased to the performance of the visual analysis. The algorithm relies on a proper performance of the test and is afflicted if participants do not walk in circles but change directions within the test or cross the centre line when walking from turning point to turning point (i.e., walked in the shape of an eight;  $\infty$ ). Because of this, the algorithm might miss a turn around the cones, resulting in a discrepancy of 20 meters (one segment) in the final distance. In the current state of algorithmic development, these changes in position influence the accuracy in detection. In future versions of the algorithm, this type of behaviour must be considered, as taken up in the section future scope.

### **Timed up and Go Test (TUG)**

For the TUG, the accuracy of the automated algorithm (85%) outperformed the visual data analysis. The low accuracy on the visual analysis (69%) was caused by difficulties to distinguish where the test started or ended. After the recording had been started on the device, the participant might have moved an arm before the test was started, for example. To deal with these confounding influences (motion artefacts in the recordings prior to the actual test or after the actual test), participants were asked not to move for one second before starting the test. Some participants struggled to comply. Therefore, the algorithms had to be adjusted towards more constant parameters that provide information about the process of standing up.

In the current development state, the automated detection depends on a standardized starting and ending posture (similar posture, e.g., hands on the lap). Nonetheless, we were able to calculate the time taken to stand up and sit down using the orientation with regard to the earth's gravity. It is worth noting that the automated analysis differed from the ground truth only by a mere 1.24 s on average.

### **30-Second Sit-to-Stand Test (30STS)**

For the 30-second Sit-to-stand Test, the accuracy was found to be higher for the visual analysis (92%) than for the automated analysis (78%). Factors that negatively influenced the automated computation quality were identified as the speed of performance and the resulting variance in distinct posture transition states (e.g., sitting or standing). Participants who performed the transitions quickly did not stay in the posture very long, whereas patients who performed transitions slowly stayed in a particular state (e.g., sitting) for a longer period. When the participants were too fast, it occurred that the algorithm misinterpreted two transitions as one. This led to detection errors and must be addressed in the future when designing the process of filtering the data. In the future, filters can be designed so that fast-performing participants can still be detected accurately (e.g., define cut-off frequencies accordingly). Moreover, strong variations in starting posture (especially arm posture) and end posture led to varying levels of accuracy between different recordings that should also be considered in the further development of the algorithm.

### **4-Stage Balance Test (4SBT)**

It is very difficult to interpret the 4SBT data because there are no ground truth data except for the passing or failing of the test stages one and four. Information on stages two and three was not available due to an error

in designing the interface for the physiotherapists. There was no place for the physiotherapists to enter the data for stages two and three, and while sometimes an optional comment was added to list this information, it was not done so throughout. This makes it impossible to give an accuracy of the developed algorithm; however, the congruence of the mere marking as "passed" or "failed" between the automated analysis and the ground truth lay between 78.18 and as high as 92.73%, indicating a promising approach. The visual analysis of the recordings even corresponded with the ground truth to up to 100%, hinting at potential for even more improvement.

However, if we consider the detection of the different parts as a tool for calculating additional and clinically relevant parameters, the automated analysis can generate very interesting results from the recorded data; for example, information about the amount of motion that is necessary to keep the balance (e.g., arm movement). As shown in the results, this equalizing motion increases with the more difficult postures.

#### *Additional parameters*

Each of the four tests in the current study was evaluated by relying on one parameter only (e.g., number of steps in the 6MWT, total completion time in the TUG). To overcome this limitation of the clinical tests, additional parameters that can be easily extracted from the sensor data were computed, especially, when using a smartwatch to record the tests. Step impact, for example, indicates whether the foot is placed confidently and pointedly on the floor, indicating stability. Participants that are walking insecurely show a lower impact, since they are carefully placing their feet, most likely due to shorter step length and not raising their feet very high from the ground as a consequence of low postural control. Low endurance, i.e., when the walking speed decreases over time, can indicate whether the test wears the participant. Quantification of arm movement during the 4SBT would allow a more detailed qualitative analysis of static postural control instead of simply looking at whether the patient was able to hold one posture for a specified time (yielding a binary result). Without a ground truth, it was impossible in this study to determine the accuracy and clinical validity of these parameters, but the descriptive analysis comparing the results from one participant with low and one with high risk of falling demonstrates the additional potential of these parameters. Since the behavioural tests are conducted as standard, sufficient data could be collected in a timely manner - both in those with low and high risk of falling, simply by routinely incorporating the smartwatch in those standard examinations. Once thresholds will have been determined according to the gathered

data, the smartwatch will allow for a much more detailed and qualitative and thus more valid classification of the fall risk.

#### *Strengths, challenges, and limitations*

As shown, wearable wrist-worn sensors can record and compute parameters related to gait, postural control, mobility and strength. By analyzing these parameters, a personalised risk assessment can be enabled via a fall risk model. Yet, even though the comparison of the test outcomes and the fall risk according to the questionnaires shows that the data correlates, the whole fall risk model is more complex and needs additional input, such as patient anamnesis information. It is not possible to infer on the overall fall risk by solely looking at the test results of any one test (TUG, 30STS, 4SBT, or 6MWT). This is in line with the literature, that suggests that a single tests should not be used in isolation to identify individuals at high risk of falls, e.g., in community-dwelling older adults.<sup>36,37</sup> While it is known that no single measure is an accurate diagnostic tool, there is limited information on which anamnestic question, self-report measure, or performance-based measure, or combination of measures, best predicts future falls. Thus, guidelines recommend that patients aged 65 years and older who are admitted to a hospital should undergo a multifactorial falls risk assessment, resulting in a vast number of required testing with limited staff.<sup>38</sup> The proposed automated analysis can save time and resources, producing relief and simultaneously improving the diagnostic efficiency, indicating patients at risk, and enabling selective preventive measures.

It not only saves time and resources but also limits the chance for human error. Human error – and human bias – can distort test results. Since the results of the observational testing were considered as the ground truth (i.e., used as the actual correct value) in the calculations, they could also influence the results within this study. For example, the reaction time of the supervisor using a manual stopwatch might influences the time measured. It is also possible that the supervisor forgets the number of turns or repetitions while counting. In addition to these errors, the human observer could be unconsciously biased, which could influence his perception, e.g., of the defined start and stop positions, such as the actual moment a subject sat down on a chair.

In case an automated algorithm recognizes the defined starting point and ending point more precisely than a human supervisor, the bias (human error/inaccuracy) in the ground truth data could lead to a higher error for the algorithm (in comparison to the ground truth). While, usually, a higher error would indicate

that the algorithm yields a lower accuracy than the observation, in these cases, it would mean that the algorithm actually outperforms the observation. This needs to be considered when interpreting the correspondence between the smartwatch data and the observational data reported in this study.

However, there are limitations and challenges to the proposed approach. One main challenge in developing a wearable system for automatic detection of standardized tests is the design of a proper user interface. An ideal solution would require as little input as possible. Instead, an implicit detection approach would identify the performed test automatically without a need of explicit user inputs. This could improve the compliance with regard to long-term usage.

Another main challenge is rooted in the fact that the results of the current approach rely on an accurate cropping of motions that either happen prior to the actual test or post recording. Erroneous cropping leads to errors in the automatic detection algorithms and therefore to lower accuracy rates. In addition to this, the speed of performing a particular test, unwanted motion artefacts, breaks during a test, and varying motion amplitudes may lead to signals that are difficult to process. Filters need to be adapted to the particular recording and user, to enable the best result possible.

Participants who reported using walking aids in their everyday lives were asked not to use those during testing. This was done to enhance reliability and comparability of the motions measured by the smartwatch during the tests. This may have decreased ecological validity, as they are permitted and even encouraged to be used when testing in the field.<sup>39,40</sup> However, it is also possible that this deviation in proceedings actually increased ecological validity of the tests for some members of the population, since around one out of 4 individuals who own walking aids actually do not use them in their everyday lives,<sup>41–43</sup> which would cause testing the fall risk with walking aids to underestimate the actual fall risk. Nonetheless, it will be very interesting to re-test the automated fall risk detection with smart wearables while using walking aids in a future study.

### *Future scope*

While the results are very promising, with the high accuracy rates reflecting face validity, the next step should entail replicating the conformity with inferential statistics, by either enlarging the sample size or employing mixed model approaches. In addition, the results show that the automated assessment achieves a very high precision in those tests that were performed correctly. However, we cannot assume that the

participants always perform the tests in the expected way. They may change walking direction, move too much or simply forget what they were instructed to do. To minimize the recording of confounding effects of these breaches in adherence, the physiotherapists could be provided with more control over starting and stopping the test remotely. A solution could be provided by introducing a smartphone app to operate the recording of the smartwatch. In addition to this, algorithmic understanding of the occurrence of breaks within tests needs to be investigated further. Breaks are possible, e.g. in the 6-Minutes Walk Test, but they did not occur within this pilot.

The algorithm used in the current study provided valid information about participants' performance on the selected functional tests, which confirms our objective that it has a potential to support identification of individuals in risk of falling. The algorithms could be further designed to collect different data for different interests. The suggested additional parameters (e.g., number of steps, step impact, time to sit, time to stand, etc.) would enable a broader view on the performance of each participant. However, the definitions of ground truth and evaluation of their predictive value, precision, accuracy, and validity need to be tested. In the future, these parameters might provide valuable insights for clinicians in understanding the physical and mental state of the elderly. The benefits include a reduction of efforts for both the practitioner and the individual being tested, as well as improved objectivity and quantification of problems.

## **Conclusion**

The overall contribution of this study is the development of an objective and automated fall risk assessment. We present wearable technology and a multi-parametrical model to estimate fall risk and, consequently, to support fall prevention. The proposed approach has been evaluated and has proven to match up to observational diagnostics performed by a physiotherapist.

By introducing an automated detection (e.g., via a smartwatch), additional parameters that would currently overwhelm the examiner (e.g., physiotherapist) can be introduced. Thus, a multiparametric assessment of the users fall risk can be enabled without increasing the cognitive load and workload of the physiotherapist. Since the wrist-worn device and the automated analysis facilitate the diagnostic part, resources become available for the physiotherapists to focus on the physical training to reduce the fall risk of the patient.

### Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: MAM has received an unconditional grant from Next Step Dynamics AB (advisory and drafting). Fraunhofer IGD (MH, WC, FH, GB, MA) was contracted for expert consultancy.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project was funded by Next Step Dynamics AB.

### Guarantor

MH.


### Contributorship


MH and FH were involved in developing the software prototypes and algorithms, and together with WC, calculating and evaluating the study results, and drafting the manuscript. MAM was involved in the protocol development, evaluating the study results along with critical revision of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript

### Acknowledgements

We would like to thank the participants and physiotherapist who helped with data collection.

### ORCID iDs

Marian Haescher  <https://orcid.org/0000-0002-1242-0278>

Gerald Bieber  <https://orcid.org/0000-0003-2496-6232>

### References

1. Haescher M, Matthies DJC, Srinivasan K, et al. Mobile assisted living: smartwatch-based fall risk assessment for elderly people. In: Proceedings of the 5th international workshop on sensor-based activity recognition and interaction. New York City: ACM, 2018, pp. 1–10.
2. Rubenstein LZ, Josephson KR and Robbins AS. Falls in the nursing home. *Ann Inter Med* 1994; 121: 442.
3. Nurmi I and Lüthje P. Incidence and costs of falls and fall injuries among elderly in institutional care. *Scand Prim Health Care* 2002; 20: 118–122.
4. Forster A and Young J. Incidence and consequences of falls due to stroke: a systematic inquiry. *BMJ*. 1995. DOI:10.1136/bmj.311.6997.83.
5. Dromerick A and Reding M. Medical and neurological complications during inpatient stroke rehabilitation. *Stroke* 1994. DOI:10.1161/01.STR.25.2.358.
6. Samuelsson CM, Hansson P-O and Persson CU. Early prediction of falls after stroke: a 12-month follow-up of 490 patients in the fall study of Gothenburg (FallsGOT). *Clin Rehabil* 2019; 33: 773–783.
7. Heinrich S, Rapp K, Rissmann U, et al. Cost of falls in old age: a systematic review. 2010. DOI:10.1007/s00198-009-1100-1.
8. Hartikainen S, Lönnroos E and Louhivuori K. Medication as a risk factor for falls: critical systematic review. *J Gerontol A Biol Sci Med Sci* 2007; 62: 1172–1181.
9. Oliver D, Daly F, Martin FC, et al. Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review. *Age Ageing* 2004; 33: 122–130.
10. Tricco AC, Thomas SM, Veroniki AA, et al. Comparisons of interventions for preventing falls in older adults: a systematic review and meta-analysis. *JAMA* 2017; 318: 1687–1699.
11. Okubo Y, Schoene D and Lord SR. Step training improves reaction time, gait and balance and reduces falls in older people: a systematic review and meta-analysis. *Brit J Sports Med* 2017; 51: 586–593.
12. Aranda-Gallardo M, Morales-Asencio JM, Canca-Sanchez JC, et al. Instruments for assessing the risk of falls in acute hospitalized patients: a systematic review and meta-analysis. *BMC Health Serv Res* 2013; 13: 1, 122.
13. Lusardi MM, Fritz S, Middleton A, et al. Determining risk of falls in community dwelling older adults: a systematic review and meta-analysis using posttest probability. *J Geriatr Phys Ther* 2017; 40: 1–36.
14. Park S-H. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. *Aging Clin Exp Res* 2018; 30: 1–16.
15. Attal F, Mohammed S, Dedabrishvili M, et al. Physical human activity recognition using wearable sensors. *Sensors (Basel)* 2015; 15: 31314–31338.
16. Casale P, Pujol O and Radeva P. Human activity recognition from accelerometer data using a wearable device. In: *Iberian conference on pattern recognition and image analysis*. Berlin: Springer, 2011, pp.289–296.
17. Kumari P, Mathew L and Syal P. Increasing trend of wearables and multimodal interface for human activity monitoring: a review. *Biosens Bioelectron* 2017; 90: 298–307.
18. Johansson D, Malmgren K and Murphy MA. Wearable sensors for clinical applications in epilepsy, Parkinson's disease, and stroke: a mixed-methods systematic review. *J Neurol* 2018; 265: 1740–1752.
19. Tao W, Liu T, Zheng R, et al. Gait analysis using wearable sensors. *Sensors (Basel)* 2012; 12: 2255–2283.
20. Genovese V, Mannini A and Sabatini AM. A smartwatch step counter for slow and intermittent ambulation. *IEEE Access* 2017; 5: 13028–13037.
21. Nemati E, Suh YS, Motamed B, et al. 2016. Gait velocity estimation for a smartwatch platform using Kalman filter peak recovery. In: *2016 IEEE 13th international conference on wearable and implantable body sensor networks (BSN)*. Piscataway, NJ: IEEE, pp.230–235.
22. Johnston AH and Weiss GM. 2015. Smartwatch-based biometric gait recognition. In: *2015 IEEE 7th International Conference on Biometrics Theory,*

- Applications and Systems (BTAS)*. Piscataway, NJ: IEEE, pp.1–6.
23. Guyatt GH, Sullivan MJ, Thompson PJ, et al. The 6-Minute Walk: a new measure of exercise capacity in patients with chronic heart failure. *Can Med Assoc J* 1985; 132: 919–923.
  24. Podsiadlo D and Richardson S. The timed “up & go”: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991; 39: 142–148.
  25. Jones CJ, Rikli RE and Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. *Res Q Exercise Sport* 1999; 70: 113–119.
  26. Lord SR, Murray SM, Chapman K, et al. Sit-to-stand performance depends on sensation, speed, balance, and psychological status in addition to strength in older people. *J Gerontol A Biol Sci Med Sci* 2002; 57: M539–M543.
  27. Rossiter-Fornoff JE, Wolf SL, Wolfson LI, et al. A cross-sectional validation study of the FICSIT common data base static balance measures. *J Gerontol A Biol Sci Med Sci* 1995; 50: M291–M297.
  28. Blankevoort CG, Van Heuvelen MJ and Scherder EJ. Reliability of six physical performance tests in older people with dementia. *Phys Ther* 2013; 93: 69–78.
  29. 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: 735–738.
  30. Chetta A, Zanini A, Pisi G, et al. Reference values for the 6-min walk test in healthy subjects 20–50 years old. *Resp Med* 2006; 100: 1573–1578.
  31. Steffen TM, Hacker TA and Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. *Phys Ther* 2002; 82: 2, 128–137.
  32. Phelan EA, Mahoney JE, Voit JC, et al. Assessment and management of fall risk in primary care settings. *Med Clin North Am* 2015; 99: 281–293.
  33. Mark JA and Loomis J. The STEADI toolkit: incorporating a fall prevention guideline into the primary care setting. *Nurse Pract* 2017; 42: 50–55.
  34. Center of Disease Control. *Stay independent brochure*, <https://www.cdc.gov/steady/pdf/STEADI-Brochure-StayIndependent-508.pdf> (2019, accessed 20 October 2020).
  35. Rubenstein LZ, Vivrette R, Harker JO, et al. Validating an evidence-based, self-rated fall risk questionnaire (FRQ) for older adults. *J Saf Res* 2011; 42: 493–499.
  35. Lundin H, Sääf M, Strender L-E, et al. One-leg standing time and hip-fracture prediction. *Osteoporos Int* 2014; 25: 1305–1311.
  36. Barry E, Galvin R, Keogh C, et al. Is the timed up and go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and Meta-analysis. *BMC Geriatr* 2014; 14: 14.
  37. Beauchet O, Fantino B, Allali G, et al. Timed up and go test and risk of falls in older adults: a systematic review, the journal of nutrition. *J Nutr Health Aging* 2011; 15: 933–938.
  38. Matarese M, Ivziku D, Bartolozzi F, et al. Systematic review of fall risk screening tools for older patients in acute hospitals. *J Adv Nurs* 2015; 71: 1198–1209.
  39. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories, et al. ATS statement: guidelines for the Six-Minute Walk Test. *Am J Respir Crit Care Med* 2002; 166: 111–117.
  40. Holland AE, Spruit MA and Singh SJ. How to carry out a field walking test in chronic respiratory disease. *Breathe (Sheff)* 2015; 11: 128–139.
  41. Van der Esch M, Heijmans M and Dekker J. Factors contributing to possession and use of walking aids among persons with rheumatoid arthritis and osteoarthritis. *Arthritis Rheum* 2003; 49: 838–842.
  42. Resnik L, Allen S, Isenstadt D, et al. Perspectives on use of mobility aids in a diverse population of seniors: implications for intervention. *Disabil Health J* 2009; 2: 77–85.
  43. Edwards NI and Jones DA. Ownership and use of assistive devices amongst older people in the community. *Age Ageing* 1998; 27: 463–468.