Average in-home gait speed: Investigation of a new metric for mobility and fall risk assessment of elders

Erik Stone a,*, Marjorie Skubic b,1, Marilyn Rantz c,2, Carmen Abbott d,3, Steve Miller c

a Department of Electrical and Computer Engineering, University of Missouri – Columbia, 349 Engineering Building West, Columbia, MO 65211, United States
b Department of Electrical and Computer Engineering, University of Missouri – Columbia, United States
c Sinclair School of Nursing, University of Missouri – Columbia, United States
d Department of Physical Therapy, University of Missouri – Columbia, United States

A R T I C L E   I N F O
Article history:
Received 13 December 2013
Received in revised form 1 August 2014
Accepted 30 August 2014

Keywords:
In-home gait
Kinect
Fall risk
Mobility
Elderly

A B S T R A C T

A study was conducted to assess how a new metric, average in-home gait speed (AIGS), measured using a low-cost, continuous, environmentally mounted monitoring system, compares to a set of traditional physical performance instruments used for mobility and fall risk assessment of elderly adults. Sixteen participants were recruited from a local independent living facility. In addition to having their gait monitored continuously in their home for an average of eleven months, the participants completed a monthly clinical assessment consisting of a set of traditional assessment instruments: Habitual Gait Speed, Timed-Up and Go, Short Physical Performance Battery, Berg Balance Scale – short form, and Multidirectional Reach Test. A methodology is developed to assess which of these instruments may work well with the largest subset of older adults, is best suited for detecting changes in an individual over time, and most reliably captures the true mobility level of an individual. Using the ability of an instrument to predict how an individual would score on all the instruments as a metric, AIGS performs best, having better predictive ability than the traditional instruments. AIGS also displays the best agreement between observed and smoothed values, indicating it has the lowest intra-individual test–retest variability of the instruments. AIGS, measured continuously, during normal everyday activity, represents a significant shift in assessment methodology compared to infrequently assessed, traditional physical performance instruments. Continuous, in-home data may provide a more accurate and precise picture of the physical function of older adults, leading to improved mobility and fall risk assessment.

1. Introduction

Research has shown that the parameters which describe locomotion are indispensable in the diagnosis of frailty and fall risk [1]. Additionally, studies have indicated that gait parameters may be predictive of future falls and adverse events in older adults [2–5], and that scores on certain mobility tests are good indicators of fall risk [6,7]. Studies have also shown that interventions to prevent falls among older adults, such as household modifications and exercise routines to improve physical function, could significantly reduce falls and be highly cost effective [8,9]. Despite these findings, gait parameters and mobility tests are generally assessed infrequently, if at all, through observation by a clinician with a stop watch or using expensive equipment in a physical performance lab. Furthermore, these sparse, infrequent evaluations may not be representative of a person’s true locomotion ability [10].

A variety of clinical instruments have been proposed for assessing mobility in the elderly, including the instruments completed each month by participants in this study. Each of these instruments is supported by studies in the literature indicating their validity, and, in many cases, suggesting their use as fall risk screening tools. However, the large number of instruments highlights the fact that no single instrument has proven completely sufficient. There are a number of contributing factors to this observation. First, although these instruments are validated

http://dx.doi.org/10.1016/j.gaitpost.2014.08.019
0966-6362/© 2014 Elsevier B.V. All rights reserved.
by one or more studies, many are also the subject of studies indicating their limitations, or lack sufficient evidence for conclusive evaluation in regards to fall risk [11]. In all likelihood, each instrument may work well with a specific subset of elderly adults (e.g., high functioning, frail, very old, etc.) among whom the underlying physical characteristics assessed by that instrument are best suited for discrimination. Therefore, one study may find an instrument works well whereas another study may find the same instrument does not, due largely to differences in the studied population.

Second, physical performance instruments, when used with frail, elderly populations, are generally subject to significant intra-individual test–retest variability when test sessions are more than a day or two apart, independent of actual change in physical function [12–14]. This variability limits their usefulness for detecting changes in an individual over time. Relatively large minimum detectable change (MDC) values that have been reported for these instruments with frail, elderly populations, such as 2.9 for Short Physical Performance Battery (SPPB), 8 for Berg Balance Scale (BBS), and approximately 30% for Timed-Up and Go (TUG), provide further evidence of this issue [12,15,16].

A system that monitors gait parameters continuously, during normal everyday activity, may offer significant benefits for mobility and fall risk assessment. Based on the preferences of older adults, an ideal system would not use wearable devices, and would not require active involvement from the user [17]. In [18], researchers developed a system that uses an array of passive infrared motion sensors mounted on the ceiling in a hallway to measure gait speed in home environments. Using data from a one month period centered on a participant’s first annual physical evaluation, researchers found statistically significant associations between in-home gait speed and a variety of mobility assessments, including the motor section of the Unified Parkinson’s Disease Rating Scale, stopwatch timed gait speed, and the Tinetti balance scale [10].

More recently, a low-cost, environmentally mounted, automated monitoring system based on the Microsoft Kinect has been developed that continuously monitors and reports the gait of residents, in their homes, during normal everyday activity [19]. No special procedures or movements are required of the residents. Privacy of the people living with the monitoring system is protected by only using the depth imagery from the Kinect (see Fig. 1B). As compared to traditional performance instruments assessed infrequently, this continuous in-home data may provide a more accurate, precise, and reliable picture of physical function.

The purpose of this study was to assess how a new metric, average in-home gait speed (AIGS), derived from continuous in-home gait data compares to a set of traditional physical performance instruments used for mobility and fall risk assessment. A methodology was developed to assess which instrument may work well with the largest subset of older adults, is best suited for detecting changes in an individual over time, and most reliably captures the true mobility level of an individual.

2. Design and methods

2.1. Subjects

With approval from the Institutional Review Board at the University of Missouri – Columbia, participants were recruited from a local independent living facility with the goal of maintaining continuous monitoring in 10 apartments. To meet this goal, monitoring systems were installed in 15 apartments for time periods ranging from 2 to 22 months. Three of the apartments had two residents, yielding a total of 18 study participants. However, due to very similar physical characteristics, separate in-home gait estimates could not be made for the residents of one apartment, necessitating their removal. Thus, data from 16 participants (7 male, 9 female) residing in 14 apartments were used for analysis. The average monitoring duration was approximately 11 months. Summary statistics are given in Table 1. Informed consent was obtained from all subjects.

2.2. Procedure

2.2.1. Average in-home gait speed

Kinect-based gait measurement systems were installed in the main living area of each apartment, as shown in Fig. 1A. Walking segments of four feet or greater occurring in view of the systems are automatically identified, segmented, and analyzed for gait speed, height of the individual walking, stride time, and stride length, among other attributes [19]. Using these data, a probabilistic model representing each resident’s gait is created, and then updated over time using data from the prior eight weeks [19]. Possible limitations of the Kinect-based gait system, such as use in multi-resident homes, and of the Kinect sensor itself are discussed in [19].

The average in-home gait speed (AIGS) of a resident for a given day is computed as a weighted average of gait speed from all walks identified in their apartment during the prior seven days. The

![Fig. 1. Kinect-based in-home gait system. (A) Photo of system installed in a typical apartment. The Kinect sensor is placed on small shelf above the front door. A computer is placed in the cabinet above the refrigerator. (B) Sample raw data and extracted foreground for three selected frames during a walking sequence in an apartment. (C) Three-dimensional models for the three selected frames created using depth data and extracted foregrounds.](image-url)
weight each walk receives in the average is determined by its likelihood given the resident model [19]. The number of walks used to compute an AIGS measurement depends on a variety of factors, including apartment layout, apartment clutter, system positioning, and resident behavior. Typically, 4–35 walks per day were identified with likelihood sufficient to conclude they were from a resident, not from a visitor or staff member. Therefore, each AIGS measurement was based on between 25 and 250 walks that received a meaningful weight in the average. Following installation and an initial setup procedure, which usually takes less than one hour, the process of identifying walking segments, extracting gait parameters, creating and/or updating resident models, and generating daily AIGS estimates is completely automated.

2.2.2. Monthly assessments

Participants completed a monthly clinical assessment consisting of a set of traditional instruments that have been studied and used to assess mobility and fall risk in the elderly: Habitual Gait Speed (HGS) [1,4,20–22], Timed-Up and Go (TUG) [7,23], Short Physical Performance Battery (SPPB) [24,25], Multidirectional Reach Test (MDRT) [26,27], and Berg Balance Scale – short form (BBS-SF) [28,29]. The monthly assessments were administered by research assistants who have periodic training by a physical therapist for accuracy in conducting them. Video from a subsample of the monthly assessments was also reviewed to confirm accuracy. Most months all participants completed the entire assessment. However, participant availability and preference prevented collection of all instruments from all participants every month.

HGS was assessed using a ten foot walking path and a stop watch. The walk was repeated twice and the average time of the two walks was recorded as the final score. TUG was assessed using an arm chair, a ten foot walking path, and a stop watch. The time required for the subject to complete the test was recorded as the final score. SPPB standard protocol, which consists of three separate sections: gait speed, repeated chair stands, and static and dynamic balance; was assessed using a ten foot walking path, an arm chair, and a stop watch. For the gait speed section, the fastest of the two ten foot walks completed for HGS was used. All three sections of the SPPB were scored on a scale from 0 to 4 using a standardized scoring scale, and the sum of the scores was recorded as the final score. BBS-SF, which consists of seven separate balance tests, half of the 14 comprising the full Berg Balance Scale (BBS), was assessed using a stop watch. The sum of the scores was recorded as the final score.

MDRT, as originally described, consists of four separate assessments: forward reach (also known as Functional Reach), right and left sideways reach, and backwards reach; with each reach assessed as the average of three trials. No method is given for combining the reach scores into a single measure, as each reach was independently validated. As part of the monthly assessments administered for this study, participants completed the forward reach (FR), and one sideways reach (SR) of their preference.

2.2.3. Evaluation methodology

Evaluation consisted of two steps. First, each instrument was assessed on how well a score on that instrument could predict the scores an individual would receive on all of the instruments. The assumption is that the score on each instrument is dependent on a set of underlying physical characteristics that influence mobility in elderly adults, as indicated by studies in the literature. Thus, the instrument that is able to best predict (on average) how an individual will score on all the instruments likely best captures the physical characteristics underlying all the instruments. As a result, one would expect the instrument with the best prediction ability to work well with the largest subset of elderly adults, and to most reliably capture the true mobility level of an individual.

As the instruments are subject to test–retest variability, the values observed during the monthly assessments reflect both the true, intrinsic score the individual would receive, as well as this test–retest variability inherent in the instrument. To better approximate the true score of an individual, the observed scores the individual receives on an instrument are smoothed, temporally, using a moving average. In this case, a centered moving average with window size 5. To illustrate, if a participant had an observed TUG time of 18 s for the month of March, and observed TUG times of 20, 25, 22, and 21 s for the months of January, February, April, and May, respectively. The estimated true TUG time for the individual for the month of March, obtained after smoothing, would be 21.2 s. For this analysis, the accuracy of predicted scores was measured against the estimated true scores obtained after smoothing.

Second, to evaluate the ability of each instrument to detect changes in an individual over time, each instrument was assessed on how well the observed and smoothed values of the instrument agreed. A large difference between the observed and smoothed values of an instrument implies large variation from one measurement to the next that is outside of any long term trend. Frequent, large variations are indicative of high test–retest variability leading to large MDC, and, thus, limited usefulness in detecting changes in an individual over time. Of course, an instrument that always produced the same value for an individual would have the lowest test–retest variability. However, such an instrument would be unable to accurately predict how an individual would score on other instruments, as it would not be capturing an actual measure of mobility. Thus, assessment of both prediction accuracy on all the instruments, and intra-individual test–retest variation, results in a multiple criteria evaluation that addresses these potential tradeoffs.

To allow prediction of one instrument given a score on another, a simple neural network model was used to learn a mapping between instruments. This neural network model was chosen over other methods, such as standard linear regression, as it allows simple non-linear relationships to be captured. The model uses a single hidden neuron with a non-linear activation function, and a single output neuron with a linear activation function. The simplicity of the model, with only four trainable parameters
(the input and bias weights of each neuron), increases the chance of good generalization, while maintaining the flexibility needed to approximate the simple, yet crucial, non-linear relationships that exist between the instruments. The model was trained using the Covariance Matrix Adaptive Evolutionary Strategies optimization algorithm, with the best result of 20 randomly initialized trials being selected. The input data were first normalized by subtracting the mean and dividing by the standard deviation, and the objective function to be minimized was the squared error on the training data. Example learned mappings are shown in Fig. 2.

After learning a mapping from one instrument to another, agreement of the predicted scores with the smoothed scores of an instrument was assessed using three metrics. The first was the intra-class correlation coefficient (ICC) (two-way, single measure, absolute agreement). ICC is a scaled metric that assesses the absolute agreement between two sets of measurements, with a value of 1.0 indicating perfect agreement. The second metric was root mean square deviation (RMSD). RMSD conveys the magnitude of the difference between values, as it is not scaled. Thus, to allow averaging across instruments, RMSD was normalized by dividing by the standard deviation of the observed scores of the instrument, to obtain normalized RMSD (NRMSD).

3. Results

From 16 participants, 159 complete monthly assessments were collected, along with their AIGS computed for the days the assessments were administered. (An additional 13 partially completed monthly assessments were not included.) Table 2 shows the agreement, measured using ICC, RMSD, and NRMSD, of the predicted scores with the smoothed scores of the instruments.

The average agreement of the predicted scores with the smoothed scores of only the traditional instruments shows AIGS compares favorably, ranking a close second to TUG on both ICC and NRMSD. This is a compelling finding, as the average agreement with the traditional instruments alone is inherently biased in their favor. This is due to the fact that the mapping of the traditional instruments to their own smoothed values is included in the average, which is not the case for AIGS. Furthermore, AIGS is the best predictor of smoothed TUG after observed TUG, and the best predictor of smoothed SR after observed SR. If AIGS is included alongside the traditional instruments in the calculation of the average, AIGS becomes the best predictor based on both ICC and NRMSD.

As shown in Table 2 (see the diagonal entries, shaded gray), the observed and smoothed values of AIGS show the best agreement of all the instruments, based on both ICC and NRMSD. The high level of agreement indicates that AIGS displays little intra-individual test–retest variability, clearly less than that of the traditional instruments. This is somewhat expected, however, as a single AIGS measurement is based on tens or hundreds of measurements made during normal everyday activity, while the traditional instruments are effectively a snapshot based on one, or a small number of, measurements made during an explicit performance evaluation.

Fig. 2. Example non-linear mappings learned from one instrument to another.
Table 2
Agreement of predicted values with smoothed values after non-linear mapping (N=159, 16 subjects).

<table>
<thead>
<tr>
<th><strong>ICC</strong> (Higher is better)</th>
<th><strong>TARGET INSTRUMENT (SMOOTHED)</strong></th>
<th><strong>AVERAGE W/O AIGS INCLUDED a</strong></th>
<th><strong>AVERAGE WITH AIGS INCLUDED b</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>AIGS</td>
<td>AIGS</td>
<td>0.993</td>
<td>0.949 0.949 0.818 0.857 0.718 0.690 0.733</td>
</tr>
<tr>
<td>TUG</td>
<td>0.795</td>
<td>0.969 0.880 0.819 0.750 0.710 0.771</td>
<td><strong>0.816</strong> 0.813</td>
</tr>
<tr>
<td>HGS</td>
<td>0.662</td>
<td>0.855 0.938 0.748 0.728 0.659 0.711</td>
<td>0.773 0.757</td>
</tr>
<tr>
<td>SR</td>
<td>0.744</td>
<td>0.799 0.734 0.890 0.769 0.671 0.598</td>
<td>0.743 0.743</td>
</tr>
<tr>
<td>FR</td>
<td>0.538</td>
<td>0.703 0.680 0.757 0.889 0.728 0.559</td>
<td>0.719 0.694</td>
</tr>
<tr>
<td>BBS-SF</td>
<td>0.539</td>
<td>0.550 0.555 0.581 0.643 0.834 0.673</td>
<td>0.639 0.625</td>
</tr>
<tr>
<td>SPPB</td>
<td>0.613</td>
<td>0.652 0.638 0.562 0.516 0.734 0.899</td>
<td>0.667 0.659</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>RMSD</strong> (Lower is better)</th>
<th><strong>TARGET INSTRUMENT (SMOOTHED)</strong></th>
<th><strong>AVERAGE W/O AIGS INCLUDED a</strong></th>
<th><strong>AVERAGE WITH AIGS INCLUDED b</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>AIGS</td>
<td>1.46</td>
<td>2.41 0.29 0.50 0.44 0.57 0.56 0.54</td>
<td>0.484 0.431</td>
</tr>
<tr>
<td>TUG</td>
<td>7.39</td>
<td>1.91 0.42 0.48 0.55 0.55 0.51</td>
<td><strong>0.456</strong> 0.473</td>
</tr>
<tr>
<td>HGS</td>
<td>9.00</td>
<td>2.99 0.63 0.63 0.55 0.58 0.56 0.56</td>
<td>0.506 0.533</td>
</tr>
<tr>
<td>SR</td>
<td>8.10</td>
<td>4.11 0.59 0.49 0.53 0.53 0.53 0.53 0.42</td>
<td>0.541 0.554</td>
</tr>
<tr>
<td>FR</td>
<td>10.1</td>
<td>5.24 0.63 0.63 0.54 0.39 0.53 0.65</td>
<td>0.562 0.594</td>
</tr>
<tr>
<td>BBS-SF</td>
<td>10.1</td>
<td>5.09 0.63 0.63 0.63 0.63 0.59 0.63</td>
<td>0.626 0.648</td>
</tr>
<tr>
<td>SPPB</td>
<td>9.47</td>
<td>5.56 0.66 0.66 0.68 0.70 0.53 0.36</td>
<td>0.598 0.617</td>
</tr>
</tbody>
</table>

a Average only includes columns TUG through SPPB.
b Average includes all columns, AIGS through SPPB.

4. Discussion

Results indicate that AIGS is better able to predict how an individual would score on all the instruments included in this study than any of the traditional instruments, and that the observed and smoothed values of AIGS show better agreement than those of any of the traditional instruments. This suggests that AIGS may work with the largest subset of elderly adults, be best suited for detecting changes in an individual over time, and most reliably capture the true mobility level of an individual. In addition, AIGS can be measured continuously and inexpensively, using a system that requires no active involvement of the subject. The system is also quite unobtrusive, taking up no floor space as currently installed, and using a computer that is the size of a typical paperback book.

Compared to traditional physical performance instruments, AIGS represents a significant shift in assessment methodology. This shift is illustrated by the large difference between the average AIGS and average HGS observed in this study, namely 51.0 cm/s and 65.4 cm/s, respectively. Traditional instruments are explicit evaluation snapshots, generally with specific, fixed protocols, conducted in clinical settings. On the other hand, AIGS is a more continuous measure assessed in an individual’s home, during their normal everyday activity.

A large and growing body of research has shown the importance of measuring gait [1–5]. Despite this, gait is generally measured infrequently, either in a clinician’s office or a performance laboratory. The results of this study suggest that measurement of gait during an individual’s normal, everyday activity in their own home could provide clinicians with a more reliable assessment of mobility and fall risk on a continuous basis. With detection of increasing fall risk using AIGS, clinicians can recommend strength training and other therapeutic interventions to reduce the risk of falls. The results also indicate AIGS could potentially be used across a broad range of frailty levels, with the only requirement being that an individual walks regularly while in their home. A shift to continuous, daily measurement may offer significant benefits for purposes beyond more reliable mobility and fall risk assessment. For example, early detection of health changes [30] could allow more targeted use of clinical interventions when they may be most effective. Furthermore, such an automated system could provide an unobtrusive, low cost, quantifiable method for evaluating the impact of interventions over time.
The clear limitation of this initial study is the sample size. While 159 monthly clinical assessments and thousands of in-home walks were used in the analysis, they were collected from 16 individuals residing in the same independent living facility. Although some of the individuals showed significant change over the monitoring period, others did not; thus, the effective size of the studied sample lies between 16 and 159. Given the encouraging results of this initial investigation, a larger study could provide additional confirmation of the findings, as well as a better understanding of the relationship between AIGS, mobility level, and fall risk among older adults.

Acknowledgments

This work was supported in part by the Agency for Healthcare Research and Quality under grant R01HS018477, and by the National Science Foundation under grant CNS-0931607. The content is solely the responsibility of the authors and does not necessarily represent the official views of either agency.

Conflicts of interest statement

Nothing to disclose.

References