

## Test-Retest Reliability of the StepWatch Activity Monitor Outputs in Healthy Adults

Suzie Mudge, Denise Taylor, Oliver Chang, and Rosita Wong

**Background:** Activity Monitors give an objective measure of usual walking performance. This study aimed to examine the test-retest reliability of the StepWatch Activity Monitor outputs (mean steps/day; peak activity index; sustained activity indices of 1, 5, 20, 30, 60 minutes; steps at high, medium, and low stepping rates).

**Methods:** Thirty healthy adults age 18 to 49 years wore the StepWatch for 2 3-day periods at least 1 week apart. **Results:** The intraclass correlation coefficients of the StepWatch outputs ranged from 0.44 to 0.91 over 3 days. The coefficient of variation ranged from 3.0% to 51.3% over the monitoring periods, with higher variation shown for shorter monitoring periods. The most reliable 5 outputs had 95% limits of agreement between 3-day periods that were less than 40%. These were mean steps/day ( $\pm 39.1\%$ ), highest step rate in 1 ( $\pm 17.3\%$ ) and 5 ( $\pm 37.4\%$ ) minutes, peak activity index ( $\pm 25.6\%$ ), and percentage of inactive time ( $\pm 9.52\%$ ).

**Conclusions:** Mean steps/day, highest step rate in 1 and 5 minutes, peak activity index, and percentage of inactive time have good test-retest reliability over a 3-day monitoring period, with lower reliability shown by the other StepWatch outputs. Monitoring over 1 or 2 days is less reliable.

**Keywords:** gait, rehabilitation, activity

Activity monitors are used to measure physical activity in an individual's natural environment. They contain a microprocessor and at least 1 accelerometer and can be worn continuously for extended periods. Thus, they are able to provide objective information about rate, amount, and patterns of physical activity, which may give insight into a person's usual performance.<sup>1</sup>

The StepWatch Activity Monitor (Orthocare Innovations, 6405 218th St SW, Suite 100, Mountlake Terrace, WA 98043–2180, US) is an example of an accelerometer based activity monitor that has been used widely in different population groups.<sup>2</sup> The StepWatch is small (75×50×20 mm) and lightweight (38 g) and is worn at the ankle. The monitor contains a custom sensor that uses a combination of acceleration, position, and timing to detect steps. Thus the outputs of the StepWatch are based on the amount, rate, and pattern of walking. The StepWatch is calibrated based on each individual's height and gait pattern and the sensitivity can be adjusted for individuals with altered gait patterns.

There have been 6 recent studies investigating test-retest reliability of the StepWatch in participants with stroke,<sup>3–5</sup> diabetic peripheral neuropathy with or without

amputations,<sup>6</sup> spinal cord injury,<sup>7</sup> control participants, and participants with a range of neurological disorders.<sup>2</sup> All of the studies measured mean steps/day, 3 of them over a short time period of less than an hour,<sup>3,6,7</sup> 1 each for 2 days,<sup>4</sup> 3 days,<sup>5</sup> and 7 days.<sup>2</sup> Each of the studies compared 2 separate time periods within 3 weeks although one study did not report statistics.<sup>6</sup> Reported intraclass correlation coefficients (ICCs) for mean steps/day ranged from 0.86 to 0.99 representing excellent test-retest reliability.

An advantage of the StepWatch is the range of outputs available. In addition to mean steps/day, the data can also be presented in numerous other forms. The peak activity index is the average step rate of the highest 30 nonconsecutive minutes over the included time in 1 day. Sustained activity measures are also available for 1, 5, 20, 30, and 60 minutes and are calculated by scanning the chosen number of consecutive minutes over the included time in 1 day and extracting the maximum number of steps achieved in this continuous interval. Thus the peak activity index and the sustained activity indices reflect rate of walking. In addition the number of steps or the duration of time at each step rate (high, medium, and low) can be calculated.

The reliability of the additional StepWatch outputs has been reported twice previously.<sup>2,5</sup> High test-retest reliability was reported particularly for mean steps/day, highest stepping rate in 1 and 5 minutes, and the peak activity index in 40 adults with stroke over a 3-day monitoring period (coefficients of variation 10.7%, 6.7%, 10.2%, and 7.9%, respectively).<sup>5</sup> The reliability of the peak activity index, and sustained activity indices for 20, 30, and 60

---

Mudge and Taylor are with the Health and Rehabilitation Research Centre, Auckland University of Technology, Auckland, New Zealand. Chang and Wong are with the Dept of Physiotherapy, Auckland University of Technology, Auckland, New Zealand.



minutes were also reported for 10 adults with a variety of neurological disorders and 10 healthy controls<sup>2</sup> over a 7-day period. Of the reported StepWatch outputs, the coefficient of variation was lowest for the peak activity index in both the control group and participants with neurological disorders (8.5%, 30.6%, respectively). Thus reliability testing in healthy participants is limited to only 10 participants and there is no published information on the activity indices for 1 and 5 minutes, which have been shown to be repeatable in participants with stroke.<sup>5</sup> Further investigation of test-retest reliability in healthy participants is warranted, particularly for the additional StepWatch outputs.

Thus, the major aim of this study was to assess the test-retest reliability of the StepWatch outputs (total step count; peak activity index; sustained activity indices of 1, 5, 20, 30, 60 minutes; steps at high, medium, and low stepping rates) during 2 3-day periods at least 1 week apart in healthy adults. A secondary aim was to assess test-retest reliability over shorter monitoring periods of 1 day (the first day of each 3-day monitoring period) and 2 days (the first 2 days of each 3-day monitoring period).

## Methods

A convenience sample of 30 healthy adults was recruited from advertising at AUT University. Participants were eligible for inclusion if they were 18 years of age or over. Individuals were excluded if they had a health condition that might impact their ability to participate, such as a recent injury limiting activity levels (eg, sprained ankle, back pain). All participants gave written informed consent, and the study was approved by the AUT University Ethics Committee.

All participants attended the Health and Rehabilitation Research Centre (HRRC) for initial testing. A StepWatch activity monitor was calibrated and attached to the lateral side of the ankle of the right leg with a Velcro strap. The monitor has an infrared light that flashes with every step, which was matched to a manual count of steps during overground walking at 3 walking speeds (fast, slow, and self selected) for 5 meters each. The sensitivity and cadence settings were adjusted, if necessary, until the flashes corresponded exactly with the manual count during the 3 walking speeds. Participants were instructed to wear the monitor for 3 days and for the same 3 days the following week, removing it for sleeping, swimming, and showering. Participants were given an instruction sheet with details about the care of the StepWatch. Participants returned to the HRRC for a subsequent session where the data were downloaded.

## Statistical Analyses

ICCs were calculated to assess test-retest reliability between the means of each 3-day period for each StepWatch output. An ICC of above 0.75 was considered to indicate excellent test-retest reliability.<sup>8</sup> The 95% limits of agreement between the means of each 3-day period

for each StepWatch output were also calculated as a coefficient of repeatability.<sup>9</sup> Bland and Altman advocate plotting the difference between the 2 measurements against the mean of the 2 measurements for each participant and then calculating 95% limits of agreement as the range of differences falling within the mean difference  $\pm$  1.96 standard deviations.<sup>9,10</sup> The 95% limits of agreement represent the repeatability of the measure from week to week and can be expressed either as absolute numbers or percentage differences between the first and second testing sessions. In this paper, the 95% limits of agreement are reported both as absolute numbers and percentage differences to allow comparison of repeatability between the different outputs. The coefficient of variation was also calculated (standard deviation expressed as a percentage of the mean) between the means of each 3-day period.<sup>11</sup>

To assess the reliability of 1- and 2-day monitoring periods, the same statistical tests were used to determine the level of agreement between the first day and the first 2 days of each monitoring period.

Bland-Altman calculations were performed using GraphPad Prism (Version 4.03; GraphPad Software Inc, 11452 El Camino Real, #215 San Diego, CA 92130, US), ICCs were calculated by SPSS (Version 14.0.0; SPSS Inc. Headquarters, 233 S. Wacker Drive, Chicago, IL 60606, US) and the coefficient of variation was calculated in Excel 2003 (Microsoft Corporation, One Microsoft Way, Redmond, WA 98052-7329, US).

## Results

Thirty participants enrolled in the study. Half of the participants wore the StepWatch for 2 periods of 3 days as instructed, however the other 15 (50%) only completed 2 monitoring periods of 2 days. Of the 30 participants with a mean (SD) age of 27.7 (8.9) years, 15 were men. Twelve (40%) participants were students, 9 (30%) individuals were employed, and 9 (30%) were not employed. Chi-square tests for independence indicated no significant difference of sex ( $P = .47$ ) or occupation ( $P = .08$ ) between participants who completed 6 days and those who completed 4 days of monitoring. Likewise, there was no difference in age (Mann-Whitney  $U$  test;  $P = .75$ ). The mean and standard deviation of each StepWatch output are shown in Table 1.

For the 15 participants who completed the 2 3-day monitoring periods, the ICCs ranged from 0.438 to 0.912 (Table 1). The majority of the ICCs (27/33) was above 0.6, and thus indicates good or excellent test-retest reliability.

The coefficient of variation ranged from 3.0% to 36.9% for the 3-day period, 4.1% to 33.8% for the 2-day period, and 6.5% to 51.3% for the 1-day period, indicating more variation with a shorter monitoring period (Table 1).

Bland Altman analysis showed that the 5 most reliable outputs had 95% limits of agreement between 3-day periods that were less than 40% (Table 1). These were mean steps/day ( $\pm 39.1\%$ ), highest step rate in 1 minute ( $\pm 17.3\%$ ) (Figure 1A), highest step rate in 5 minutes ( $\pm 37.4\%$ ), peak activity index ( $\pm 29.8\%$ ), and percentage



**Table 1 Mean, Standard Deviation, and Reliability Statistics for StepWatch Outputs for 1-Day, 2-Day, and 3-Day Periods 1 Week Apart (n = 15)**

StepWatch output	Mean (SD)	ICC			Coefficient of variation (%)			±95% limits of agreement*		
		Day 1	Days 1 & 2	All 3 days	Day 1	Days 1 & 2	All 3 days	Day 1	Days 1 & 2	All 3 days
Mean steps/day	8609 (2625)	0.512	0.838	0.895	28.8	16.4	11.8	9131 (101)	4611 (61.1)	3341 (39.1)
Number of steps at medium rate (>30 and <60 steps/minute)	4051 (1408)	0.472	0.735	0.854	32.9	20.4	13.0	4914 (112)	3137 (82.1)	2111 (53.5)
Number of steps at high rate (>60 steps/minute)	1984 (1272)	0.790	0.823	0.744	51.3	33.8	36.9	2984 (184)	2202 (102)	2521 (122)
Number of steps at medium and high rate (>30 steps/minute)	5928 (2205)	0.590	0.863	0.904	31.8	20.3	12.4	6939 (110)	3538 (71.9)	2132 (42.0)
Percentage time inactive (no steps)	77.5 (6.3)	0.612	0.820	0.912	6.5	4.1	3.0	(26.0)	(13.9)	(9.52)
Peak activity index (steps/min)	82.3 (12.7)	0.809	0.742	0.779	10.7	9.7	6.6	31.4 (39.7)	25.5 (31.4)	23.4 (25.6)
Highest step rate in 60 minutes (max 60) (steps/min)	35.4 (10.7)	0.675	0.813	0.603	31.2	21.9	17.8	35.8 (99.2)	23.9 (66.5)	26.4 (67.3)
Highest step rate in 30 minutes (max 30) (steps/min)	46.1 (11.6)	0.715	0.784	0.489	25.9	20.8	17.8	36.6 (88.6)	27.0 (59.0)	32.6 (61.1)
Highest step rate in 20 minutes (max 20) (steps/min)	52.8 (12.6)	0.766	0.699	0.438	25.0	20.7	18.0	36.2 (85.1)	33.3 (61.7)	37.2 (62.4)
Highest step rate in 5 minutes (max 5) (steps/min)	85.1 (15.1)	0.727	0.701	0.677	16.3	13.7	10.7	45.6 (60.3)	35.2 (42.1)	33.6 (37.4)
Highest step rate in 1 minute (max 1) (steps/min)	111.8 (9.3)	0.484	0.769	0.719	6.9	5.7	4.8	27.1 (25.8)	18.2 (17.0)	19.3 (17.3)

\* Expressed as absolute value (percentage).

Abbreviations: SD, standard deviation; ICC, intraclass correlation coefficient.

of time with no steps ( $\pm 9.52\%$ ). The 6 other StepWatch outputs had 95% limits of agreement between 3-day periods greater than 40% (42.0% to 122%). Figure 1B shows the Bland Altman plot for 3-day test-retest reliability for highest step rate in 20 minutes as an example of wide limits of agreement. There were 3 StepWatch outputs that had 95% limits of agreement less than 40% for both the 2-day monitoring period (percentage of time with no steps,  $\pm 13.9\%$ ; highest step rate in 1 minute,  $\pm 17.0\%$ ; peak activity index,  $\pm 31.4\%$ ) and the 1-day monitoring period (percentage of time with no steps,  $\pm 26.0\%$ ; highest step rate in 1 minute,  $\pm 25.8\%$ ; peak activity index,  $\pm 39.7\%$ ).

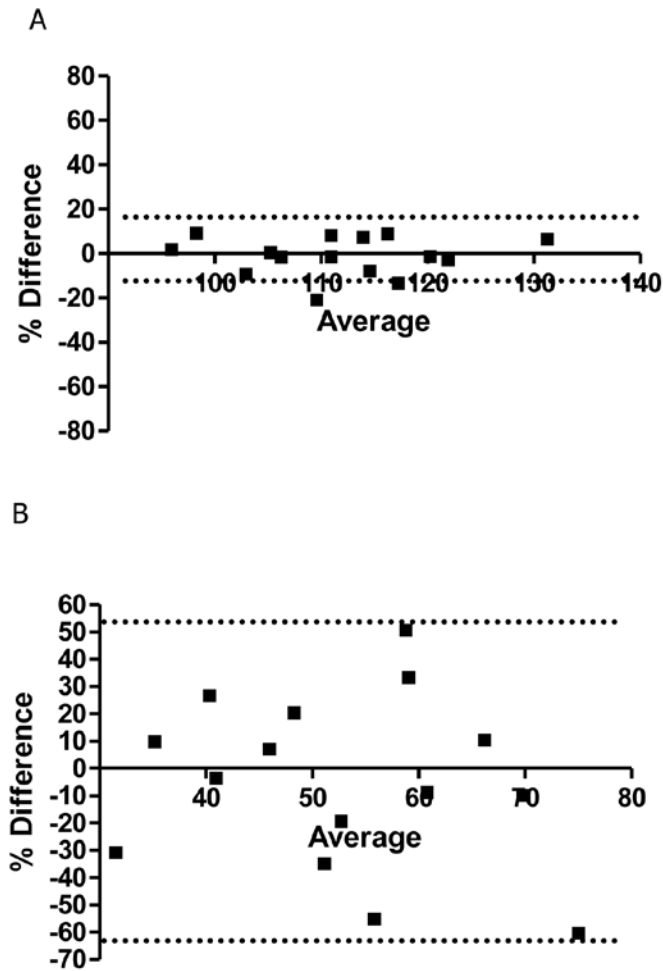
There were 15 participants who were not included in the 3-day reliability analysis, but had sufficient data to test for reliability over 2 days. Thus, Table 2 shows the reliability for a 1-day and 2-day monitoring period for all 30 participants. The ICCs ranged from 0.317 to 0.861

(Table 2) and the majority (16/22) was above 0.6, and thus indicates good or excellent test-retest reliability. The coefficient of variation ranged from 3.4% to 58.8% for the 2-day period and 5.4% to 73.6% for the 1-day period, again indicating more variation with a shorter monitoring period (Table 2). There were 2 StepWatch outputs that had 95% limits of agreement less than 40% for both the 2-day monitoring period (percentage of time with no steps,  $\pm 12.0\%$ ; highest step rate in 1 minute,  $\pm 31.9\%$ ) and the 1-day monitoring period (percentage of time with no steps,  $\pm 22.1\%$ ; highest step rate in 1 minute,  $\pm 39.9\%$ ).

## Discussion

This study has shown that 3-day monitoring of total step count by the StepWatch shows excellent test-retest reliability, with an ICC of 0.895 and 95% limits of agreement of less than 40%. Four other StepWatch outputs





**Figure 1** —Bland-Altman plot with 95% limits of agreement expressed as percentages for a 3-day monitoring period of (A) highest step rate in 1 minute and (B) highest step rate in 20 minutes.

**Table 2 Reliability Statistics for StepWatch Outputs for 1-Day and 2-Day Periods 1 Week Apart (n = 30)**

StepWatch output	ICC		Coefficient of variation (%)		±95% limits of agreement*	
	Day 1	Days 1 & 2	Day 1	Days 1 & 2	Day 1	Days 1 & 2
Mean steps/day	0.452	0.852	33.9	17.4	10310 (118)	4666 (64.3)
Number of steps at medium rate (>30 and <60 steps/minute)	0.595	0.797	34.9	23.9	4424 (124)	2965 (88.1)
Number of steps at high rate (>60 steps/minute)	0.317	0.700	73.6	58.8	6131 (235)	3106 (211)
Number of steps at medium and high rate (>30 steps/minute)	0.429	0.841	39.6	26.0	9036 (134)	4196 (93.8)
Percentage time inactive (no steps)	0.703	0.861	5.4	3.4	(22.1)	(12.0)
Peak activity index (steps/min)	0.790	0.825	15.7	12.4	45.2 (56.9)	32.2 (43.5)
Highest step rate in 60 minutes (max 60) (steps/min)	0.531	0.752	33.9	21.3	47.4 (113)	27.5 (74.3)
Highest step rate in 30 minutes (max 30) (steps/min)	0.593	0.782	30.6	21.2	56.7 (105)	32.6 (69.8)
Highest step rate in 20 minutes (max 20) (steps/min)	0.634	0.745	30.8	22.3	61.7 (102)	41.9 (73.9)
Highest step rate in 5 minutes (max 5) (steps/min)	0.768	0.758	18.8	16.8	54.6 (68.7)	46.0 (58.7)
Highest step rate in 1 minute (max 1) (steps/min)	0.734	0.841	10.7	8.9	37.9 (39.9)	31.1 (31.9)

\* Expressed as absolute value (percentage).

Abbreviations: SD, standard deviation; ICC, intraclass correlation coefficient.



(percentage of inactive time, highest step rate in 1 minute, peak activity index, and highest step rate in 5 minutes) also have high ICCs, low coefficients of variation and 95% limits of agreement less than 40% when measured over 3 days. This suggests that these are also reliable outputs for a 3-day monitoring period. In contrast, the variation of the majority of StepWatch outputs over 1 or 2 days is higher, indicating lower reliability with shorter monitoring periods.

Our reliability results in healthy participants are similar to previous work in participants with stroke.<sup>5</sup> Furthermore, these results compare favorably to Busse et al's study of monitoring with the StepWatch over 7 days<sup>2</sup> and suggest that a 3-day monitoring period of total step count and peak activity index appears to be at least as reliable as a 7-day monitoring period.

An inherent challenge in measuring the reliability of daily activity levels is that there is a natural variation that occurs from day to day, both for the number of steps taken and for the rate of stepping. It can be difficult for one reliability study to differentiate between this natural variation and measurement error. However, previous work has shown consistently that the StepWatch is accurate in step counting,<sup>3,6,12–16</sup> so the variation detected in this study likely reflects true daily variation in activity levels. A potential source of error in the StepWatch is interpretation of certain activities such as cycling or leg swinging as steps.<sup>17</sup> However the total steps accrued with such movements is likely to be negligible over a 24 hour period.

Although all participants wore the StepWatch for at least 2 periods of 2 days, only half wore it for 2 periods of 3 days, as instructed. This compliance rate is considerably lower than previous work in stroke where 74% (40/54) of participants wore the monitor for the full period. Reasons for missing days included forgetting to wear the monitor, wearing the monitor upside down or deliberately not wearing the monitor for fear of damage (eg, during skiing or contact sports). The lower compliance rate in these participants raises the issue that compliance over longer periods or in people who engage in contact sports may be problematic.

This study is limited by the number of participants who did not wear the monitor for the 3-day monitoring period, so the results are based on a smaller number of participants, half the number than originally intended. However, a separate analysis of reliability for the 30 participants over 2 and 1-day period corroborates the findings of lower reliability for these shorter monitoring periods, with only 2 StepWatch outputs (percentage of time inactive and highest stepping rate in 1 minute) showing 95% limits of agreement of less than 40%.

The study is limited by its relatively small sample size. In addition, the high proportion of students in our sample is unlikely to be representative of healthy adults. This may limit the generalizability of the findings of this study to the general population.

Both the highest step rate in 1 minute and the peak activity index had better test-retest reliability than total

step count over a 3-day monitoring period in this study. Highest step rate in 1 minute and peak activity index are both based on rate rather than amount of stepping and although they may be reflective of maximal physical performance in individuals with stroke,<sup>18</sup> healthy young adults could be expected to achieve cadences of over 150 steps/min if running. In this study, the mean fastest minute of the day was  $112 \pm 9.3$  steps/min, which implies that most, if not all, participants did not run during the monitoring period. Thus for healthy young adults in this study, neither the peak activity index nor the highest stepping rate in 1 minute seem to reflect maximal physical performance.

The mean steps taken per day by the healthy young adults in this study was  $8609 \pm 2625$  steps, which falls outside the 95% confidence interval of 9216 to 10,377 steps calculated by Bohannon in a meta-analysis of average daily steps taken by adults under the age of 65.<sup>19</sup> This may also relate to the lower than expected range of values for peak activity index and highest stepping rate in 1 minute found in this study.

## Conclusions

Mean steps/day, highest step rate in 1 and 5 minutes, peak activity index, and percentage of inactive time have good test-retest reliability over a 3-day monitoring period, with lower reliability shown by the other StepWatch outputs in healthy young adults. Monitoring over 1 or 2 days is less reliable, however compliance over longer periods may be problematic for some individuals.

## References

1. International Classification of Functioning. *Disability and Health: ICF*. Geneva: World Health Organization; 2001.
2. Busse ME, Pearson OR, Van Deursen R, Wiles CM. Quantified measurement of activity provides insight into motor function and recovery in neurological disease. *J Neurol Neurosurg Psychiatry*. 2004;75:884–888.
3. Macko RF, Haeuber E, Shaughnessy M, et al. Microprocessor-based ambulatory activity monitoring in stroke patients. *Med Sci Sports Exerc*. 2002;34:394–399.
4. Haeuber E, Shaughnessy M, Forrester LW, Coleman KL, Macko RF. Accelerometer monitoring of home- and community-based ambulatory activity after stroke. *Arch Phys Med Rehabil*. 2004;85:1997–2001.
5. Mudge S, Stott NS. Test-retest reliability of the StepWatch Activity Monitor outputs in individuals with chronic stroke. *Clin Rehabil*. 2008;22:871–877.
6. Coleman KL, Smith DG, Boone DA, Joseph AW, del Aguila MA. Step activity monitor: long-term, continuous recording of ambulatory function. *J Rehabil Res Dev*. 1999;36:8–18.
7. Bowden M, Behrman A. Step activity monitor: accuracy and test-retest reliability in persons with incomplete spinal cord injury. *J Rehabil Res Dev*. 2007;44:355–362.
8. McDowell I. *Measuring health: a guide to rating scales and questionnaires*. 3rd ed. New York: Oxford University Press; 2006.



9. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
10. Bland JM, Altman DG. Measurement error. *BMJ*. 1996;312:1654.
11. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30:1–15.
12. Mudge S, Stott NS, Walt SE. Criterion validity of the StepWatch Activity Monitor as a measure of walking activity in patients after stroke. *Arch Phys Med Rehabil*. 2007;88:1710–1715.
13. McDonald CM, Widman L, Abresch RT, Walsh SA, Walsh DD. Utility of a step activity monitor for the measurement of daily ambulatory activity in children. *Arch Phys Med Rehabil*. 2005;86:793–801.
14. Foster RC, Lanningham-Foster LM, Manohar C, et al. Precision and accuracy of an ankle-worn accelerometer-based pedometer in step counting and energy expenditure. *Prev Med*. 2005;41:778–783.
15. Shepherd EF, Toloza E, McClung CD, Schmalzried TP. Step activity monitor: increased accuracy in quantifying ambulatory activity. *J Orthop Res*. 1999;17:703–708.
16. Hartsell H, Fitzpatrick D, Brand R, Frantz R, Saltzman C. Accuracy of a custom-designed activity monitor: implications for diabetic foot ulcer healing. *J Rehabil Res Dev*. 2002;39:395–400.
17. Karabulut M, Crouter SE, Bassett DR, Jr. Comparison of two waist-mounted and two ankle-mounted electronic pedometers. *Eur J Appl Physiol*. 2005;95:335–343.
18. Mudge S, Stott NS. Timed walking tests correlate with daily step activity in individuals with stroke. *Arch Phys Med Rehabil*. 2009;90:296–301.
19. Bohannon RW. Number of pedometer-assessed steps taken per day by adults: a descriptive meta-analysis. *Phys Ther*. 2007;87:1642–1650.