# Test-Retest Reliability of the StepWatch Activity Monitor Outputs in Individuals with Chronic Stroke

Suzie Mudge, MHSc; N. Susan Stott, PhD

Department of Surgery, University of Auckland

Address for correspondence: Suzie Mudge, Department of Surgery, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. Phone: 64 9 373 7599 X 85387, Fax: 64 9 367 7159, e-mail: s.mudge@auckland.ac.nz.

Test-Retest Reliability of the StepWatch Activity Monitor Outputs in Individuals with Chronic Stroke


## Abstract

**Objective:** To examine the test-retest reliability of the StepWatch Activity Monitor outputs over two periods a week apart in participants with stroke.

**Design:** Test-retest reliability study over monitoring periods of one, two and three days

**Setting:** Participant's usual environment

**Participants:** Forty participants more than six months post stroke.

**Main measures:** StepWatch outputs: total step count, number of steps at high medium and low stepping rates, sustained activity indices, peak activity index.

**Results:** The Intraclass Correlation Coefficients were high for all StepWatch outputs and all monitoring periods but were highest for the three day monitoring period (0.930-0.989) and lowest for the one day monitoring period (0.830-0.950). The coefficient of variation ranged from 6.7-48.7% over the monitoring periods, with higher variation shown for shorter monitoring periods. The most reliable four outputs had 95% limits of agreement between three day periods that were less than 40%. These were total step count (±37.8%), highest step rate in one minute (±23.0%), highest step rate in five minutes (±38.6%) and peak activity index (±29.8%). The highest step rate in one minute was the only

StepWatch output that had 95% limits of agreement less than 40% for the two day (±31.2%) and one day (±36.7%) monitoring periods.

**Conclusions:** Total step count, highest step rate in one minute, highest step rate in five minutes and peak activity index have good test-retest reliability over a three day monitoring period, with lower reliability shown by the other StepWatch outputs. In general, monitoring over one or two days is less reliable.

**Key Words:** Cerebrovascular accident; Gait; Rehabilitation; Activity

Stroke is the most common cause of severe disability in adults[1]. Although 70% of people with stroke are able to walk independently in a clinical setting[2], fewer walk independently in the community[3]. In line with current recommendations[4], the focus is shifting from clinic based measures to measures of walking in usual environments[5], such as activity monitoring. One such monitor is the StepWatch Activity Monitor[6], which consists of an accelerometer and microprocessor. The StepWatch can be worn on the ankle for extended periods and has a range of outputs available.

Total step count is the most commonly used output, but sustained activity measures are also available for 1, 5, 20, 30 and 60 minutes. These measures are calculated by scanning the accumulated 24 hour data to determine the maximum number of steps taken during continuous intervals of 1, 5, 20, 30 and 60 minutes.  The peak activity index represents the average step rate of the fastest 30 minutes over 24 hours, regardless of when they occurred. The peak activity index is thus a non-continuous interval in contrast to the sustained activity indices which are continuous intervals. The number of steps at high (above 60 steps/min), medium (between 30 and 60 steps/min) and low (below 30 steps/min) step rates can also be calculated.

Of all the available outputs, only total step count has been tested for test-rest reliability in the stroke population[7, 8]. These two studies reported excellent test-retest reliability with Intraclass Correlation Coefficients above 0.96. One study used the StepWatch over a short period of less than an hour[7] but the other monitored participants for two days[8]. Although these results are promising, the current literature provides little guidance for clinicians or researchers regarding the optimal monitoring period in patients with stroke or information on the reliability of many of the available 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) over two three day periods at least one week apart in subjects with chronic stroke. A secondary aim was to assess test-retest reliability over shorter monitoring periods of one day (the first day of each three day monitoring period) and two days (the first two days of each three day monitoring period).

## Methods

A convenience sample of 40 individuals with chronic stroke was recruited from the hospital stroke service and newspaper advertising. Subjects were eligible for inclusion if they were at least six months post stroke, were able to walk independently but with some residual difficulty confirmed by the physical functioning scale of the 36 Item Short Form Health Survey[9] and walked in the community at least once a week. Individuals were excluded if they had fallen more than twice in the previous six months, had another serious health problem affecting walking or if they were unable to complete the testing for another reason. All participants gave written informed consent, and the study was approved by the Northern Regional Ethics Committee.

All subjects attended the Neuro Rehab Results clinic for initial testing. A StepWatch activity monitor (Orthocare Innovations, 6405 218th St SW, Suite 100, Mountlake Tce, WA 98043-2180, US) was calibrated and attached to the lateral side of the ankle of the non-paretic leg with a strap or cuff. The sensitivity and cadence settings were adjusted for each participant so that the monitor recognised every step during fast, slow and self selected walking speeds. Self-selected gait speed was tested over 10 metres[10] and the Rivermead Mobility Index[11] was administered.

Subjects were instructed to wear the monitor for three days and for the same three days the following week, removing it for sleeping and showering. Subjects were given an instruction sheet with details about the care of the StepWatch and a follow up appointment was made to pick up the monitor.

## Statistical analyses

Intraclass correlation coefficients (ICC) were calculated to assess test-retest reliability between the means of each three day period for each StepWatch output. An ICC of above 0.75 was considered to indicate excellent test-retest reliability[12]. The 95% limits of agreement between the means of each three day period for each StepWatch output were also calculated as a coefficient of repeatability[13]. Bland and Altman advocate plotting the difference between the two measurements against the mean of the two measurements for each subject and calculate 95% limits of agreement as the range of differences falling within the mean difference $\pm$ 1.96 standard deviations[13, 14]. 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 three-day period[15].

To assess the reliability of one and two day monitoring periods, the same statistical tests were used to determine the level of agreement between the first day and the first two 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**

Fifty-four participants enrolled in the study. One participant withdrew during the testing period due to an unrelated medical complication. A further 13 participants were excluded from the analysis as they did not wear the StepWatch for six complete days. The remaining 40 participants with a mean (SD) age of 69.2 (12.6) years completed the study. There were 23 men and 17 women with a mean (SD) 5.1 (5.1) years after stroke. Fifteen participants had right sided paresis. The mean (SD) gait speed was 0.67 (0.34) m/s. All participants walked independently with a mean (SD) score of 13 (1.8) on the Rivermead Mobility Index. The mean and standard deviation of each StepWatch output are shown in Table 1.

<Table 1>

ICCs were generally high for all StepWatch outputs and all monitoring periods but were lowest for the one day monitoring period (Table 1). The ICCs ranged from 0.830-0.989, indicating excellent test-retest reliability.

The coefficient of variation ranged from 6.7-37.6% for the three day period, 8.4-43.9% for the two day period and 8.5-48.7% for the one day period, indicating more variation with a shorter monitoring period (Table 1).

Bland Altman analysis showed that the four most reliable outputs had 95% limits of agreement between three day periods that were less than 40% (Table 1). These were total step count (±37.8%), highest step rate in one minute (±23.0%) (Figure 1A), highest step rate in 5 minutes (±38.6%) and peak activity index (±29.8%). The seven other StepWatch outputs had 95% limits of agreement between three day periods greater than 40% (48.4-153.3%). Figure 1B shows the Bland Altman plot for three day test-retest reliability for highest step rate in 30 minutes as an example of wide limits of agreement. The highest step rate in one minute was the only StepWatch output that had 95% limits of agreement less than 40% for the two day monitoring period (±31.2%) and for the one day monitoring period (±36.7%).

**Discussion**

This study has shown that three day monitoring of total step count by the StepWatch shows excellent test-retest reliability, with an ICC of 0.989 and 95% limits of agreement of less than 40%. Three other StepWatch outputs (highest step rate in one minute, peak activity index and highest step rate in five minutes) also have high ICCs and 95% limits of agreement less than 40% when measured over three days. This suggests that these are also reliable outputs for a three day monitoring period. In contrast, the variation of the majority of StepWatch outputs over one or two days is higher, indicating lower reliability with shorter monitoring periods.

Our results of monitoring over three days compare favourably to the study of seven day monitoring by Busse et al[16] and suggest that a three day monitoring period of total step count appears to be at least as reliable as a seven day monitoring period. Although the ICCs were high for total step count over the two and one day monitoring periods and comparable to previous work in stroke[7, 8], the 95% limits of agreement for total step count revealed considerable day to day variability. Thus we would advocate at least a three day monitoring period for total step count.

Both the highest step rate in one minute and the peak activity index had better test-retest reliability than total step count over a three day monitoring period in this study. The lower reliability of total step count may, in part, be due to the natural variation that occurs from day to day. Besides physical capacity, there are other potential influences on the total steps an individual takes in one day. For example, it is likely that behavioural, personal, environmental and social factors also impact the number of steps taken in one day. Highest step rate in one minute and peak activity index are both based on rate rather than amount of stepping and may be more reflective of maximal physical performance.

Of all the measures, the highest step rate in one minute has the best test-retest reliability and the lowest coefficient of variation over three days. The daily variability is also low. Therefore, this output could be used if it were only possible to monitor for a day. The high daily variation for all other StepWatch outputs is similar to the findings of Busse et al[16] and we would concur that the variation in most StepWatch outputs is too high for one day to be recommended as a monitoring period.

Reliability studies provide important direction for clinicians, who need to determine how much change in a measure following an intervention is required to reflect a true change. 95% limits of agreement derived from Bland Altman analyses provide one estimate of true change and frequently highlight variability not apparent with ICCs. In general, wide limits of agreement for the same measure between different weeks means that, either changes due to an intervention will be obscured by the measurement variability (caused by measurement inaccuracy or natural daily subject variability) or, alternatively, changes due to the variability of the measure will be misinterpreted as a true change in the patient. In this study, we have selected activity measures that are repeatable to within 40% as outputs that are more clinically relevant based on other studies that have shown changes of up to 40% in other parameters such as gait velocity after exercise programmes in subjects with stroke[17]. However, this does not rule out use of SAM outputs with higher variability as long as the variability is recognised.

Of note, Hopkins argues that 95% limits of agreement are too large to use as a threshold for deciding if a true change has occurred. Instead he suggests that 1.5 to 2.0 times the typical error (or coefficient of variation) as a more realistic threshold[15]. If we use 2.0 as the more conservative estimate of true change, then at least a 21% change in total step count and highest step rate in five minutes, 14% in highest step count in one minute and 16% in peak activity index represent a true change over a three day monitoring period.

The clinical relevance of activity monitoring is also an important question. In individuals with stroke, total step count is correlated to gait speed[18], the FIM motor score[19] and the Berg Balance Scale (r=0.58)[18]. In patients with other neurological conditions, total step count correlates to the physical functioning index of the 36 Item Short Form Health Survey[20] and the Rivermead Mobility

Index[21]. However, there is no published work investigating the relationship between the other StepWatch outputs and clinical tests. The high test-retest reliability of the highest step rate in one minute, peak activity index and highest step rate in five minutes supports the need for further work in this area to investigate the clinical relevance of these additional outputs.

Another limitation of the use of sustained activity indices is that there is little information published about normal values. Busse et al reported a mean peak activity index of 59 steps/minute in ten healthy subjects, but did not report on highest step rate in one minute[16]. The authors did comment that the sustained activity indices were more variable in a healthy population, compared to patients with neurological disorders. Further research is needed to establish normal values for the activity indices.

The results of this study are limited in their applicability to those with chronic stroke who are able to walk independently and who walk in the community at least once weekly. Thus the findings may not apply to those with more limited mobility or acute stroke. Likewise the results may not apply to those with a diagnosis other than stroke. It is recommended that reliability is assessed in other population groups.

**Clinical Messages**

- Total step count has excellent test-retest reliability when used for three days in individuals with stroke.
- Highest stepping rate in one minute is the most reliable output of the StepWatch.
- Monitoring for less than a three day period is not recommended due to high variability.

**Acknowledgements**

We gratefully acknowledge the contributions of Pat Bennett and Alan Barber, MD, of the Auckland District Health Board for subject recruitment.

**Contributors**

Both authors were responsible for framing the hypothesis, study design, including analysis strategy and data interpretation. SS assisted with the ethics application and SM was responsible for data collection and handling. SM took the position of lead author & SS assisted with writing and framing the paper.

## References


1. Adamson J, Beswick A, Ebrahim S. Is stroke the most common cause of disability? J Stroke Cerebrovasc Dis 2004;13(4):171-7.

2. Kelly-Hayes M, Beiser A, Kase CS, Scaramucci A, D'Agostino RB, Wolf PA. The influence of gender and age on disability following ischemic stroke: The Framingham study. J Stroke Cerebrovasc Dis 2003;12(3):119-26.

3. Lord SE, McPherson K, McNaughton HK, Rochester L, Weatherall M. Community ambulation after stroke: how important and obtainable is it and what measures appear predictive? Arch Phys Med Rehabil 2004;85(2):234-9.

4. International Classification of Functioning, Disability and Health: ICF. Geneva: World Health Organization; 2001.

5. Patla A, Shumway-Cook A. Dimensions of mobility: defining the complexity and difficulty associated with community walking. J Aging Phys Activ 1999;7:7-19.

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(1):8-18.

7. Macko RF, Haeuber E, Shaughnessy M, Coleman KL, Boone DA, Smith GV et al. Microprocessor-based ambulatory activity monitoring in stroke patients. Med Sci Sports Exerc 2002;34(3):394-9.

8. 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(12):1997-2001.

9. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30(6):473-83.

10. Bohannon RW. Walking after stroke: comfortable versus maximum safe speed. Int J Rehabil Res 1992;15(3):246-8.

11. Collen FM, Wade DT, Robb GF, Bradshaw CM. The Rivermead Mobility Index: a further development of the Rivermead Motor Assessment. Int Disabil Stud 1991;13(2):50-4.

12. McDowell I. Measuring health: a guide to rating scales and questionnaires. Third ed. New York: Oxford University Press; 2006.

13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-10.

14. Bland JM, Altman DG. Measurement error. BMJ 1996;312(7047):1654.

15. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;30(1):1-15.

16. 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(6):884-8.

17. Ivey FM, Hafer-Macko CE, Macko RF. Exercise rehabilitation after stroke. NeuroRx 2006;3(4):439-50.

18. Michael KM, Allen JK, Macko RF. Reduced ambulatory activity after stroke: the role of balance, gait, and cardiovascular fitness. Arch Phys Med Rehabil 2005;86(8):1552-6.

19. Shaughnessy M, Michael KM, Sorkin JD, Macko RF. Steps after stroke: capturing ambulatory recovery. Stroke 2005;36(6):1305-7.

20. Smith DG, Domholdt E, Coleman KL, Del Aguila MA, Boone DA. Ambulatory activity in men with diabetes: Relationship between self-reported and real-world performance-based measures. J Rehabil Res Dev 2004;41(4):571-80.

21. Busse ME, Wiles CM, van Deursen RW. Community walking activity in neurological disorders with leg weakness. J Neurol Neurosurg Psychiatry 2006;77(3):359-62.

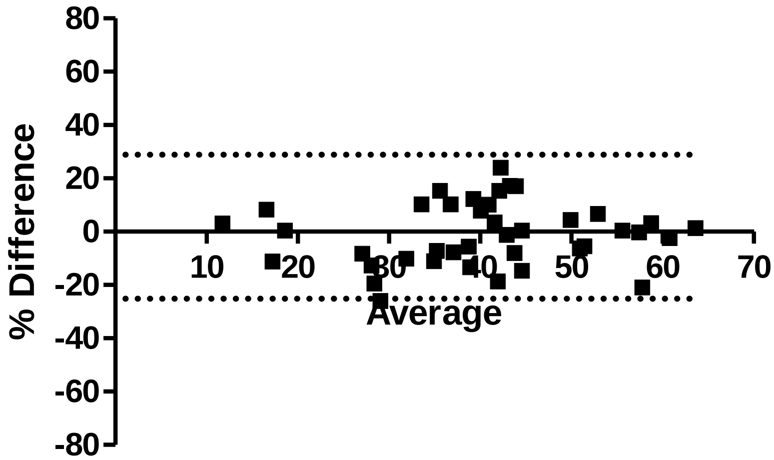
**Table 1. Mean, standard deviation and reliability statistics for SAM outputs for one day, two day and three day periods one week apart.**

| SAM 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 |
| Total step count | 6247 (4439) | 0.928 | 0.986 | 0.989 | 22.1 | 14.9 | 10.7 | 2507 (91.1) | 2028 (62.0) | 1801 (37.8) |
| Number of steps at medium rate (>30 and <60 steps/minute) | 2241 (1681) | 0.950 | 0.978 | 0.984 | 31.8 | 22.5 | 17.8 | 1549 (139) | 988 (101) | 836 (87.1) |
| Number of steps at high rate (>60 steps/minute) | 1614 (2252) | 0.873 | 0.971 | 0.926 | 48.7 | 43.9 | 37.6 | 3401 (186) | 1449 (206) | 1750 (153) |
| Number of steps at medium and high rate (>30 steps/minute) | 3855 (3604) | 0.923 | 0.985 | 0.969 | 35.2 | 24.1 | 20.9 | 4147 (148) | 1673 (109) | 1814 (90.7) |
| Number of steps at low rate (>30 steps/minute) | 2338 (1051) | 0.871 | 0.940 | 0.953 | 18.5 | 14.6 | 11.1 | 1777 (75.0) | 1112 (53.0) | 1643 (63.6) |
| Peak activity index (steps/min) | 58 (22) | 0.861 | 0.953 | 0.955 | 12.8 | 9.4 | 7.9 | 34.7 (60.6) | 19.7 (43.4) | 18.7 (29.8) |
| Highest step rate in 60 minutes (max 60) (steps/min) | 25 (18) | 0.852 | 0.948 | 0.957 | 25.7 | 18.9 | 14.8 | 31.5 (104) | 15.3 (75.3) | 14.2 (55.0) |
| Highest step rate in 30 minutes (max 30) (steps/min) | 33 (20) | 0.889 | 0.940 | 0.939 | 21.5 | 16.3 | 14.4 | 30.3 (91.5) | 19.2 (62.8) | 19.7 (52.8) |
| Highest step rate in 20 minutes (max 20) (steps/min) | 39 (22) | 0.903 | 0.929 | 0.930 | 19.3 | 15.2 | 13.5 | 31.6 (81.0) | 23.7 (55.9) | 23.4 (48.4) |
| Highest step rate in 5 minutes (max 5) (steps/min) | 59 (24) | 0.830 | 0.918 | 0.934 | 16.0 | 12.4 | 10.2 | 41.9 (63.9) | 27.6 (47.9) | 24.5 (38.6) |
| Highest step rate in 1 minute (max 1) (steps/min) | 81 (25) | 0.912 | 0.941 | 0.964 | 8.5 | 8.4 | 6.7 | 29.9 (36.7) | 23.7 (31.2) | 18.3 (23.0) |

* expressed as absolute value (percentage)

Figure 1A&B. Bland-Altman plot with 95% limits of agreement expressed as percentages for a three day monitoring period of (A) highest step rate in one minute and (B) highest step rate in 30 minutes.

A




B