Department of Rehabilitation and Physical Medicine
Kagoshima University
Graduate School of Medical Sciences

Intra-rater and inter-rater reliability of gait analysis using portable gait rhythmogram for post-stroke hemiparetic patients

RYUJI MIYATA

Kagoshima, Japan
2017

Contents

NOTE

This thesis is originally based on the below paper published in Journal of Physical Therapy Science.

Ryuji Miyata, Shuji Matsumoto, Seiji Miura, Kentaro Kawamura, Tomohiro Uema, Kodai Miyara, Ayana Niibo, Tadashi Ogura, Megumi Shimodozono

Intra-rater and inter-rater reliability of the portable gait rhythmogram in post-stroke patients.

Journal of Physical Therapy Science 29: 874-879, 2017.

ABSTRACT

[Purpose] Gait analysis, such as portable gait rhythmogram (PGR) provides objective information that helps in the quantitative evaluation of human locomotion. The purpose of this study was to assess the reliability of PGR in post-stroke patients.

[Subjects and Methods] Two raters (A and B) examined 44 post-stroke patients. To assess intra-rater reliability, rater A tested subjects on three separate occasions (Days 1, 2, and 3). To assess inter-rater reliability, raters A and B independently tested participants on the same occasion (Day 3). The PGR procedure was first explained and demonstrated by the rater, and then subjects performed one bout of the PGR protocol (two practice trials, followed by four recorded trials). There was a 1-minute break between trials for data collection to avoid fatigue. Next, subjects performed one bout of the PGR protocol on Day 2. Finally, subjects performed two bouts of the PGR protocol, with one administered by rater A and one by rater B, on Day 3. While wearing the PGR device, each patient performed the 10-m walk test twice at a comfortable speed followed by two timed trials at a maximum speed. Changes in walking time, gait cadence, and the peak absolute value of acceleration vectors were examined during the 10-m walk test. Data from the two trials at each speed were averaged for statistical analysis.

[Results] There was no significant systematic bias between test occasions or raters. Intraclass correlation coefficient values were 0.93–0.97 for intra-rater reliability at both the comfortable speed and maximum speed, and 0.97–0.98 (comfortable speed) and 0.87–0.99 (maximum speed) for inter-rater reliability. The standard error was 1.25–1.49 (comfortable speed) and 1.62–1.77 (maximum speed) for intra-rater investigation, and 1.04–1.32 (comfortable speed) and 0.91–1.26 (maximum speed) for inter-rater investigation. At the 90% confidence level, the minimum detectable change ranged from 2.9–4.1%, and the error of an individual's score at a given time point ranged from ±2.1–2.9%.

[Conclusion] Based on this excellent reliability of the PGR in post-stroke patients, it can be recommended as a simple test of gait analysis in this population.

Key words: portable gait rhythmogram, reliability, stroke

INTRODUCTION

Patients with stroke have gait and balance problems related to joint mobility and stability, muscle tone, muscle weakness and endurance, and loss of proprioception[1]. Gait analysis provides objective information that helps in the quantitative evaluation of human locomotion. The walking patterns of post-stroke patients were extensively analyzed quantitatively in the 1990s using gait analysis systems based on stereophotogrammetry, force platforms, electromyography, and metabolimeters[2]. Classical gait analysis provides a huge quantity of data, usually focused on hip, knee, and ankle angular kinematics and kinetics, but it is expensive and time-consuming in clinical settings. Thus, there is a need for a few meaningful indices to assess deterioration or amelioration of patients' gait[3].

More recently, many researchers have started to use less-expensive wearable and optoelectronic devices[4], devoting more attention to the entire locomotor system, including the upper body (not only lower limbs), and highlighting the need to measure a few clinically meaningful benchmarks of human gait, such as stability, symmetry, and harmony[5]. The wearable accelerometer system is frequently used today for clinical applications such as diagnosis and treatment planning in people suffering from neurological disorders such as stroke. With the development of microelectromechanical system technology, body-fixed inertial sensors such as accelerometers and gyroscopes have increasingly been used for gait analysis.

Recently, Yoneyama et al. developed the portable gait rhythmogram (PGR) system, which allows the precise extraction of gait-related acceleration from overall motion-related acceleration, for the assessment of both normal and Parkinson's disease gait in an ambulatory setting over long-term monitoring[6] (Figure 1A, 1B). In their study, as an essential component of this system, a robust methodology must be developed for the identification and characterization of gait signal[6]. A novel peak detection algorithm has been proposed that may be implemented in the rhythmogram to isolate every stride event automatically from acceleration signals[6]. This algorithm works self-adaptively, requiring no user-specific parameters[6].

To be useful in clinical practice for post-stroke patients, the PGR must demonstrate adequate measurement properties, including reliability, validity, and responsiveness. Reliability is often investigated first, being a prerequisite for the other properties. For performance-based measures, where measurements are taken by a rater, both intra-rater reliability, the extent to which measurements taken by the same rater are consistent, and inter-rater reliability, the extent

Fig 1A. Portable Gate Rhythmogram (PGR)

Fig 1B. Subject with PGR

to which measurements taken by different raters are similar, are important[7].

No previously published study has adequately evaluated the intra-rater or inter-rater reliability of the PGR for short-term monitoring in post-stroke patients. Additionally, the possibility of speed-related differences in reliability should be considered. Accordingly, the primary objective of this study was to evaluate the intra-rater and inter-rater reliability of the PGR in post-stroke patients. The secondary objective was to evaluate reliability at each speed.

SUBJECTS AND METHODS

A repeated-measures design was used. Intra-rater reliability was evaluated by comparing PGR scores (10-m walking time, gait cadence [steps per minute], and the peak absolute value of acceleration vectors) taken by the same rater (rater A) on three separate test occasions (labeled Days 1, 2, and 3), a maximum of 7 days apart. Three test occasions were used to allow for the possibility of a learning effect between Days 1 and 2. Day 1 was considered a familiarization day, and reliability was calculated using data from Days 2 and 3 only. Inter-rater reliability was evaluated by comparing PGR scores taken by two different raters (raters A and B) on the same test occasion (Day 3). This measurement took place on the final test occasion, allowing inter-rater reliability to be analyzed independent of any learning effect[8].

The procedures used in this study complied with the 1964 Declaration of Helsinki, as revised in 2013. Informed consent was obtained from each subject according to the ethical guidelines of the hospital after they fully

understood the study purpose and methodology. This study was carried out with the permission of the ethical committee of Kagoshima University (#27-95).

According to the Sturges' rule, a power calculation determined that 19 subjects of each gender were required for a reliability analysis involving two time points or raters, to distinguish ρ_0=0.7 from ρ_1=0.9 at α=0.05 and β=0.2[9]. Allowing for a 10% dropout rate, 47 post-stroke patients were recruited from inpatients and outpatients admitted to the Kirishima Rehabilitation Centre of Kagoshima University Hospital in Japan. Two women and one man subsequently withdrew from the study before its completion due to scheduling difficulties (n=2) and illness (n=1), leaving 44 subjects (23 men, 21 women). The mean (± standard deviation: SD) age of the post-stroke patients was 49.6±14.8 years (range: 33–68 years). The diagnosis of stroke was based on computed tomography (CT) scanning or magnetic resonance imaging (MRI), as well as neurological function. Of the 44 post-stroke patients, 32 were diagnosed with cerebral infarction, and 12 were diagnosed with cerebral hemorrhage. The mean time from onset was 25.6±10.2 weeks (range: 12–46 weeks). The median clinical Brunnstrom stage of the hemiplegic lower limb was 4 (range: 3–6): 4 patients were stage 3, 25 patients were stage 4, 13 patents were stage 5, and 2 patients were stage 6. Eighteen patients had right hemiplegia, and 26 had left hemiplegia.

Subjects were excluded if they reported a history of injury or surgery to the legs or lumbar spine or any neurological disorders before the stroke episode, since these could affect neuromuscular performance[8]. Those aged over 70 years or with a cognitive disorder were also excluded due to the potential frailty and inability to perform the study procedure. To facilitate the generalizability of findings and comparison with other studies, subject activity levels were recorded using the Barthel index, which evaluates activities of daily living (ADL) involving mobility, stairs, and transfers[10]. A mean (± SD) Barthel index score of 72.4±13.6 was obtained, which is approximately equivalent to walking independently (but may use any aid, for example, an ankle-foot orthosis or cane) > 50 yards[10].

Subjects performed the test on three occasions (Day 1, 2, and 3), with or without the use of an ankle-foot orthotic or a T-cane. The mean (± SD) interval between Days 1 and 2 was 3.2 ± 1.4 days, with 4.6 ± 1.8 days between Days 2 and 3. All three test occasions took place at the same time of day. Testing was administered by rater A on Day 1. The PGR procedure was first explained and demonstrated by the rater, and then subjects performed one bout of the PGR protocol (two practice trials, followed by four recorded trials). There was a 1-minute break between trials for data collection to avoid fatigue. Next, subjects performed one bout of the PGR protocol (two practice trials, followed by four

recorded trials) on Day 2. Finally, subjects performed two bouts of the PGR protocol, with one administered by rater A and one by rater B, on Day 3. Rater testing was preassigned according to a counterbalanced, randomized ordering across consecutive subjects. Raters were blind to each other's findings. The first bout of the PGR comprised two practice trials, followed by four recorded trials. After 10 minutes of rest, subjects performed the second bout of the PGR (one practice trial, followed by four recorded trials). The additional practice trial was performed to prevent a performance decrease resulting from the rest period.

The PGR is a small device (8 × 6 × 2 cm; weight, 80 g) that three-dimensionally (a_x,a_y,a_z) measures the accelerations (1) accompanying limb and trunk movements and (2) induced by step-in and kick-off during gait (LSI Medience Corporation, Tokyo, Japan)[11,12]. The PGR is attached to the waist of the patient and records the above signals at a sampling rate of 100 Hz. The data are automatically stored in a micro-SD card. When recording is completed, the absolute value of acceleration vectors (a; $a^2 = a^2_x + a^2_y + a^2_z$) is calculated and graphically displayed on a PC. Using the mathematical method of "pattern matching," the acceleration vectors caused by stepping can be distinguished from those of other limb and trunk movements or by unexpected artifacts, as reported previously[11,12]. While wearing the PGR device, each patient performed the 10-m walk test twice at a comfortable speed followed by two timed trials at a maximum speed[13]. Changes in walking time, gait cadence, and the peak absolute value of acceleration vectors were examined during the 10-m walk test. Data from the two trials at each speed (comfortable and maximum) were averaged for statistical analysis[14].

Statistical analysis was conducted using the Statistical Package for the Social Sciences (SPSS; SPSS Inc. Chicago, Illinois version 18.0 for Windows) with the level of statistical significance set a priori at 0.05. Separate analyses were conducted for each trial speed (comfortable and maximum). For both intra-rater and inter-rater reliability, we calculated an intraclass correlation coefficient (ICC) with absolute agreement and a 95% confidence interval[7]. The ICC is the most commonly used reliability index for continuous data[15]. An ICC ≥0.7 indicates "good" reliability[16] and is considered sufficient for using a measure in research. An ICC ≥0.9 indicates "excellent" reliability[16] and is considered sufficient for making clinical decisions regarding individual patients.

For both the intra-rater and inter-rater analyses, we estimated the standard error of measurement (SEM) as the square root of the mean square error term from the analysis of variance produced during the ICC calculation. The 95% confidence interval for the SEM was calculated using the method of Stratford and Goldsmith[17]. The SEM represents

the amount of error associated with a measure, expressed in actual units of measurement[18]. Using the SEM from the intra-rater analysis, the minimum detectable change at the 90% confidence level (MDC_{90}) was estimated as: SEM × √2 × 1.64[18,19]. The MDC_{90} is the smallest change in an individual's score considered to be a true change and not measurement error[18]. Additionally, we estimated the error in an individual's score at a given point in time at the 90% confidence level as ±SEM × 2.32[18,19].

RESULTS

In the intra-rater analysis, mean (± SD) PGR scores for Days 1, 2, and 3 are shown in Table 1. Repeated-measures analyses of variance indicated no significant systematic bias between testing occasions (Table 1). Reliability and measurement error statistics are presented in Table 2. ICC values were excellent, while SEM values ranged from 1.25 to 1.77, demonstrating similar reliability at both the comfortable speed and maximum speed. For the total sample, the error associated with an individual's score at a given point in time at the 90% confidence level ranged was very small and the MDC_{90} ranged from 2.9% to 4.1%.

In the inter-rater analysis, mean (± SD) PGR scores for raters A and B are shown in Table 3. There was no significant systematic bias between raters (Table 3). Reliability and measurement error statistics are presented in Table 4. ICC values ranged from 0.87 to 0.99, while SEM values ranged from 0.91 to 1.32, demonstrating similar reliability and measurement error for both the comfortable speed and maximum speed.

Table 1. PGR scores for Days 1, 2, and 3 (intra-rater analysis)

	PGR scores (mean ± SD)		
	Day 1	Day 2	Day 3
10-m walking time (sec)			
Comfortable speed	13.2 ± 3.4	13.1 ± 2.9	13.0 ± 2.6
Maximum speed	11.7 ± 4.1	11.5 ± 3.4	11.3 ± 3.6
Gait cadence (steps/min)			
Comfortable speed	86.6 ± 14.4	87.0 ± 13.4	87.8 ± 13.3
Maximum speed	97.2 ± 12.6	97.8 ± 12.8	98.6 ± 16.4
Peak absolute value of acceleration vectors (m/sec^2)			
Comfortable speed	33.2 ± 9.4	35.0 ± 8.2	34.6 ± 8.4
Maximum speed	44.3 ± 8.2	45.6 ± 9.4	45.1 ± 8.1

Values are expressed as mean ± standard deviation.
PGR: portable gait rhythmogram; SD: standard deviation

Table 2. Intra-rater reliability and measurement error statistics

	ICC (95% CI)	SEM, % (95% CI)	Error in an individual's score, %	MDC_{90}, %
10-m walking time (sec)				
Comfortable speed	0.96 (0.92–0.98)	1.49 (1.22–1.92)	± 2.4	3.5
Maximum speed	0.93 (0.86–0.97)	1.68 (1.35–2.19)	± 2.8	3.9
Gait cadence (steps/min)				
Comfortable speed	0.96 (0.93–0.98)	1.46 (1.26–1.75)	± 2.4	3.4
Maximum speed	0.94 (0.89–0.98)	1.77 (1.44–2.30)	± 2.9	4.1
Peak absolute value of acceleration vectors (m/sec^2)				
Comfortable speed	0.97 (0.93–0.99)	1.25 (1.02–1.63)	± 2.1	2.9
Maximum speed	0.96 (0.93–0.97)	1.62 (1.40–1.93)	± 2.7	3.8

Error in an individual's score estimated at 90% confidence level.
ICC: intraclass correlation coefficient; 95% CI: 95% confidence interval; SEM: standard error of measurement; MDC_{90}: minimum detectable change at 90% confidence level

Table 3. PGR scores for raters A and B (inter-rater analysis)

	PGR scores (mean ± SD)	
	Rater A	Rater B
10-m walking time (sec)		
Comfortable speed	13.0 ± 2.6	12.9 ± 2.4
Maximum speed	11.3 ± 3.6	11.1 ± 3.1
Gait cadence (steps/min)		
Comfortable speed	87.8 ± 13.3	88.4 ± 13.4
Maximum speed	98.6 ± 16.4	99.6 ± 17.4
Peak absolute value of acceleration vectors (m/sec^2)		
Comfortable speed	34.6 ± 8.4	36.6 ± 10.4
Maximum speed	45.1 ± 8.1	47.2 ± 9.8

PGR: portable gait rhythmogram; SD: standard deviation

Table 4. Inter-rater reliability and measurement error statistics

	ICC (95% CI)	SEM, % (95% CI)
10-m walking time (sec)		
Comfortable speed	0.97 (0.92–0.99)	1.21 (0.92–1.80)
Maximum speed	0.98 (0.93–0.99)	1.09 (0.82–1.61)
Gait cadence (steps/min)		
Comfortable speed	0.97 (0.92–0.99)	1.32 (1.00–1.95)
Maximum speed	0.99 (0.97–1.00)	0.91 (0.69–1.33)
Peak absolute value of acceleration vectors (m/sec^2)		
Comfortable speed	0.98 (0.97–0.99)	1.04 (0.85–1.35)
Maximum speed	0.87 (0.78–0.93)	1.26 (1.02–1.62)

ICC: intraclass correlation coefficient; 95% CI: 95% confidence interval; SEM: standard error of measurement

DISCUSSION

The PGR demonstrated excellent intra-rater and inter-rater reliability at both comfortable and maximum speeds in post-stroke patients. Most ICC values exceeded 0.9, suggesting sufficient reliability for making clinical decisions regarding individual patients[16]. Such high reliability is uncommon for gait performance-based measures in post-stroke patients, who typically demonstrate greater inconsistency than healthy volunteers. Reliability values exceeding 0.9 have not been previously reported for the 10-m walk test or 6-minute walk test in post-stroke patients[20,21]. Further, since the reliability of gait acceleration-based measures has not been widely investigated, ICC values exceeding 0.9 have never been reported for accelerometry measurements in post-stroke patients[22].

The reliability of an acceleration measurement device similar to the PGR, i.e., the smartphone, has been evaluated in post-stroke patients by several authors[23,24]. The previous studies comparable to the present investigation have calculated the reliability of smartphone-based gait analysis tools. Cerrito et al. found that intra-rater reliability was "good" (ICC, range: 0.86–0.93) for subjects' vertical ground reaction forces and vertical acceleration, but the reliability of smartphone-derived parameter was "poor" (ICC=0.43)[25].

The inter-rater reliability of another acceleration system was evaluated in a study by Hartmann et al, with all ICCs less than 0.9 (range 0.12–0.88)[26]. Direct comparison with our investigation is difficult, because in our study, different raters simultaneously measured the same session of the acceleration system, rather than independently testing subjects, possibly reducing performance variation. A shared finding with our investigation is that inter-rater reliability was superior to intra-rater reliability.

The present study has a number of clinically relevant findings. (1) There was no significant learning effect (systematic bias) between test occasions; therefore, no additional familiarization day is required before using the PGR. (2) The error in an individual's score at a given point in time was very small. (3) The MDC_{90}, the smallest change in an individual's score considered to reflect true change and not measurement error[19]. (4) The excellent inter-rater reliability suggests that a clinician who has not used the PGR before can become proficient with the measure following a single familiarization session. (5) The excellent inter-rater reliability demonstrates that a 10-minute rest period can be used between the practice and recorded trials. Unlike accelerometry measurement using a smartphone[23,24], the PGR does not currently allow such a rest period.

The present study was designed in accordance with recommendations for conducting a reliability study,

considering such factors as sample size, blinding, representativeness of raters, systematic bias, appropriate statistical analysis, and clinical relevance[9,15]. However, there are one main limitations that should be considered when interpreting its results. To exclude the effects of motor learning, the inter-rater reliability investigation did not take place until Day 3. Thus, participants had already received instructions from rater A, the more experienced rater. Therefore, any variability resulting from the initial instructions being given by different raters was removed, possibly inflating inter-rater reliability. However, given that the PGR is simple to perform and instructions were standardized, any such inflation is likely to be small.

In conclusion, the PGR system demonstrated excellent levels of intra-rater and inter-rater reliability in post-stroke patients, with no significant between-session learning effect. Reliability and measurement error were similar at comfortable and maximum speed. The PGR system demonstrated excellent intra-rater and inter-rater reliability in post-stroke patients, providing a basis for future investigations to establish its clinical utility.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

REFERENCES

1) Beyaert C, Vasa R, Frykberg GE: Gait post-stroke: Pathophysiology and rehabilitation strategies. Neurophysiol Clin, 2015, 45: 335-355.
2) De Quervain IA, Simon SR, Leurgans S, et al.: Gait pattern in the early recovery period after stroke. J Bone Joint Surg Am, 1996, 78: 1506-1514.
3) Schutte LM, Narayanan U, Stout JL, et al.: An index for quantifying deviations from normal gait. Gait Posture, 2000, 11: 25–31.
4) Iosa M, Morone G, Fusco A, et al.: Seven capital devices for the future of stroke rehabilitation. Stroke Res Treat, 2012, 2012, 187965.
5) Torricelli D, Mizanoor RSM, Gonzalez J, et al.: Benchmarking human-like posture and locomotion of humanoid robots: A preliminary scheme. Biomimetic and Biohybrid Systems, 2014, 8608, 320–331.
6) Yoneyama M, Kurihara Y, Watanabe K, et al.: Accelerometry-based gait analysis and its application to

Parkinson's disease assessment- part 2: a new measure for quantifying walking behavior. IEEE Trans Neural Syst Rehabil Eng, 2013, 21: 999-1005.

7) Batterham AM, George KP: Reliability in evidence-based clinical practice: a primer for allied health professionals. Phys Ther Sport, 2003, 4: 122–128.

8) Hertel J, Miller SJ, Denegar CR: Intratester and intertester reliability during the Star Excursion Balance Tests. J Sport Rehabil, 2000, 9: 104–116.

9) Walter SD, Eliasziw M, Donner A: Sample size and optimal designs for reliability studies. Stat Med, 1998, 17: 101–110.

10) Mahoney F, Barthel D: Functional evaluation: the Barthel index. Md State Med J, 1965, 14: 61-65.

11) Mitoma H, Yoneyama M, Orimo S: 24-hour recording of parkinsonian gait using a portable gait rhythmogram. Intern Med, 2010, 49: 2401–2408.

12) Utsumi H, Terashi H, Ishimura Y et al.: Quantitative assessment of gait bradykinesia in Parkinson's disease using a portable gait rhythmogram. Acta Med Okayama, 2012, 66: 31–40.

13) Bohannon RW: Comfortable and maximum walking speed of adults aged 20-79 years: reference values and determinants. Age Ageing, 1997, 26: 15-19.

14) Bohannon RW, Andrews AW, Thomas MW: Walking speed: reference values and correlates for older adults. J Orthop Sports Phys Ther, 1996, 24: 86-90.

15) Karanicolas PJ, Bhandari M, Kreder H, et al.: Evaluating agreement: conducting a reliability study. J Bone Joint Surg Am, 2009, 91A: 99–106.

16) Hahn EA, Cella D, Chassany O, et al.: Precision of health-related quality-of-life data compared with other clinical measures. Mayo Clin Proc, 2007, 82: 1244–1254.

17) Stratford PW, Goldsmith CH: Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. Phys Ther, 1997, 77: 745–750.

18) Bent NP, Wright CC, Rushton AB, et al.: Selecting outcome measures in sports medicine: a guide for practitioners using the example of anterior cruciate ligament rehabilitation. Br J Sports Med, 2009, 43: 1006–1012.

19) Reid A, Birmingham TB, Stratford PW, et al.: Hop testing provides a reliable and valid outcome measure during rehabilitation after anterior cruciate ligament reconstruction. Phys Ther, 2007, 87: 337–349.

20) Maeda A, Yuasa T, Nakamura K, et al.: Physical performance tests after stroke: reliability and validity. Am J Phys Med Rehabil, 2000, 79: 519-525.

21) Graser JV, Letsch C, van Hedel HJ. Reliability of timed walking tests and temporo-spatial gait parameters in youths with neurological gait disorders. BMC Neurol, 2016, 16: 15.

22) Saremi K, Marehbian J, Yan X, et al.: Reliability and validity of bilateral thigh and foot accelerometry measures of walking in healthy and hemiparetic subjects. Neurorehabil Neural Repair, 2006, 20: 297-305.

23) Capela NA, Lemaire ED, Baddour N, et al.: Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants. J Neuroeng Rehabil, 2016, 13: 5.

24) Isho T, Tashiro H, Usuda S: Accelerometry-based gait characteristics evaluated using a smartphone and their association with fall risk in people with chronic stroke. J Stroke Cerebrovasc Dis, 2015, 24: 1305-1311.

25) Cerrito A, Bichsel L, Radlinger L, et al.: Reliability and validity of a smartphone-based application for the quantification of the sit-to-stand movement in healthy seniors. Gait Posture, 2015, 41: 409-413.

26) Hartmann A, Murer K, de Bie RA, et al.: Reproducibility of spatio-temporal gait parameters under different conditions in older adults using a trunk tri-axial accelerometer system. Gait Posture, 2009, 30: 351-355.

ACKNOWLEDGEMENT

The author thanks Dr. S. Matsumoto for his persistent guidance and help throughout the study, Dr. S. Miura and Dr. K. Kawamura for their valuable data interpretation, Mr. T. Uema, Mr. K. Miyara, and Ms. A. Niibo for their excellent advice on data collection, and Dr. T. Ogura and Prof. M. Shimodozono for their drafting the article and encouragement throughout the study.