Japanese Journal of Comprehensive Rehabilitation Science (2019)

Original Article

Influence of clinical experience and instruction on typical cases on the inter-rater reliability of observational gait analysis

Hiroki Tanikawa, RPT, DMSc,[1] Kei Ohtsuka, RPT, DMSc,[1] Junya Yamada, RPT, MS,[2] Masahiko Mukaino, MD, DMSc,[3] Fumihiro Matsuda, RPT, DMSc,[1] Hitoshi Kagaya, MD, DMSc,[3] Eiichi Saitoh, MD, DMSc,[3] Yoshikiyo Kanada, RPT, DMSc,[1] Shuji Hashimoto, DMSc[4]

[1]Faculty of Rehabilitation, School of Health Sciences, Fujita Health University, Toyoake, Aichi, Japan
[2]Department of Rehabilitation, Fujita Health University Hospital, Toyoake, Aichi, Japan
[3]Department of Rehabilitation Medicine I, School of Medicine, Fujita Health University, Toyoake, Aichi, Japan
[4]Department of Hygiene, School of Medicine, Fujita Health University, Toyoake, Aichi, Japan

ABSTRACT

Tanikawa H, Ohtsuka K, Yamada J, Mukaino M, Matsuda F, Kagaya H, Saitoh E, Kanada Y, Hashimoto S. Influence of clinical experience and instruction on typical cases on the inter-rater reliability of observational gait analysis. Jpn J Compr Rehabil Sci 2019; 10: 14–20.

Objective: To investigate the influence of clinical experience and instruction by observing typical cases beforehand on the inter-rater reliability of observational gait analysis.

Methods: Thirty physical therapists were divided into the instructed group and the non-instructed group. Each group was then divided into two subgroups according to clinical experience. All the groups viewed videos of 30 hemiplegic patients walking on a treadmill, and then rated the severity of abnormalities on a scale of five levels. The agreement of judgments was calculated. The instructed group viewed videos of typical cases for each severity before the rating.

Results: Agreement between raters was low in the non-instructed group and slightly higher in the instructed group. Almost all pairs of raters had a significant rank correlation. The proportion of pairs with a significant rank correlation of the rating as well as no significant difference in the ratings according to the Wilcoxon signed-rank test was larger in the instructed group than in the non-instructed group.

Conclusions: The reliability of observational judgments was low even among experienced raters. This was due to the difference in the raters' subjective scale. Observation of typical cases is effective for improving the reliability.

Key words: observational gait analysis, hemiplegia, reliability, clinical experience, instruction

Introduction

In rehabilitation medicine, gait analysis is used to quantify the degree of gait abnormality, determine the treatment, and evaluate the effectiveness of the treatment. Instrumented gait analysis is used predominantly for research. Observational gait analysis (OGA) is widely used in clinical settings as it is cost-efficient, quick, and easy [1–3]. However, OGA is subjective and uses an ordinal scale. Some studies have reported that the inter-rater reliability of OGA is poor [4–11], whereas others have reported that it is high [12–15]. The results differ depending on the type of subjects being observed, the number of raters, the raters' occupations, the number of grades in the rating scale, the part of the body being observed, and the method of observation (live or videotape). Moreover, those reliability assessment studies investigated the agreement between raters when judging abnormalities based only on the joint angle.

In analyzing the gait of stroke patients, identification of certain gait characteristics assists in understanding the mechanisms of the gait abnormality and could provide suggestions for treatment [16]. Although several OGA scales have been developed to prevent omissions and improve the reliability of OGA [4, 12, 17–19], they mainly assess the joint angles in each phase of the gait cycle. Some of these scales also simultaneously assess certain abnormal gait patterns but do not record their severity. No studies have investigated the reliability of OGA when the raters

Correspondence: Hiroki Tanikawa, RPT, DMSc
Faculty of Rehabilitation, School of Health Sciences, Fujita Health University, 1–98 Dengakugakubo, Kutsukake, Toyoake, Aichi 470–1192, Japan.
E-mail: tanikawa@fujita-hu.ac.jp
Received: September 7, 2018;Accepted: December 10, 2018.
Conflict of interest: The authors have no conflict of interest in this study. This study did not receive financial support.

doi.org/10.11336/jjcrs.10.14

judge the severity of an abnormal gait pattern in stroke patients. It is also not clear whether clinical experience has any positive influence on reliability [3, 13, 15]. Furthermore, we previously mentioned that setting a clear standard for judging the severity of abnormal gait patterns could approximate the raters' subjective scales and then consequently improve the inter-rater reliability of OGA [20]. However, there are no studies that examined specific methods to improve the reliability of OGA. As a similar study, only one demonstrated that a training program improved the observational accuracy of the smoothness of upper extremity movements in stroke patients [21].

The purpose of this study was to investigate the inter-rater reliability of physical therapists judging the severity of abnormal gait patterns in hemiplegic patients based on observation, and to investigate the influence of clinical experience and instruction by observing typical cases beforehand on the inter-rater reliability.

Methods

1. Selecting typical cases for instruction

Fifty hemiplegic patients were recruited in this study. Inclusion criteria were hemiplegia, circumduction gait and/or knee extensor thrust observed during gait, and the ability to walk without assistance. Circumduction gait was defined as "abduction and lateral rotation of the hip from initial swing to mid-swing, and adduction and medial rotation of the hip from mid-swing to terminal swing, following a semicircular trajectory" [22–24]. Knee extensor thrust was defined as "a rapid knee jerking movement towards a posterior angle from the loading response to terminal stance" [24] and cases showing hyperextension of the knee on the affected side during the whole stance phase were excluded. Among the 50 hemiplegic patients, 27 had right hemiplegia and 23 had left hemiplegia, the age was 57 ± 15 years, height was 163.4 ± 9.5 cm, weight was 59.2 ± 10.5 kg, and time after onset was 1,452 ± 3,136 days (mean ± SD). Their median lower limb Brunnstrom recovery stage was IV (range, I–VI). To enable easy observation of the body movements, the patients wore skin-tight legwear, and colored markers (30 mm in diameter) were placed bilaterally on the acromia, iliac crests, great trochanters of the femurs, lateral epicondyles of the femurs, lateral malleoli, and heads of the fifth metatarsal bones. The patients were instructed to walk on a treadmill (ADAL 3D; Tecmachine, Andrézieux-Bouthéon, France) at their normal overground gait velocity. Use of the handrail and brace was allowed. Their walking was recorded from the affected side and the back using digital video cameras (HDR-FX1, Sony).

Using the treadmill walking videos of the 50 hemiplegic patients, four experienced physical therapists (mean duration 15.8 years, range: 8.0–34.0 years) selected a patient as a typical case that demonstrated gait pattern abnormalities (circumduction gait and knee extensor thrust) for each of the five levels of severity. The five levels were normal, slight, mild, moderate, and severe based on their clinical experience. Before the therapists selected the typical cases, the two abnormal gait patterns were defined.

2. Reliability of OGA and the influence of clinical experience and instruction on the reliability

2.1 Subjects

Thirty physical therapists (21 men and 9 women) from three different facilities were recruited as raters. Their mean duration of clinical experience was 5.8 years (range: 0.4–14.4 years). The raters were divided into two groups approximately equal in average years of clinical experience. One group was defined as the instructed group and the other as the non-instructed group. The instructed group viewed videos of typical cases before making their judgments. The non-instructed group did not watch the videos and served as the control group. The instructed group and the non-instructed group were then divided into subgroups according to clinical experience (≥5 years and <5 years of clinical experience) (Figure 1).

2.2 Procedure

The procedure is shown in Figure 2. The definitions

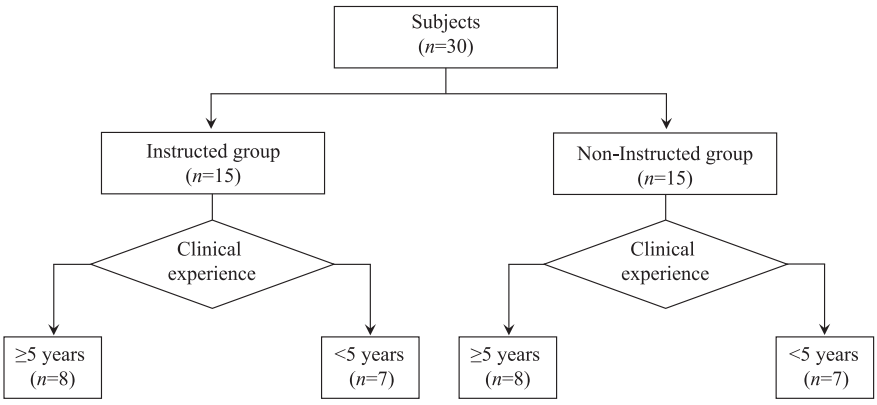

Figure 1. Flowchart of subject group assignment.

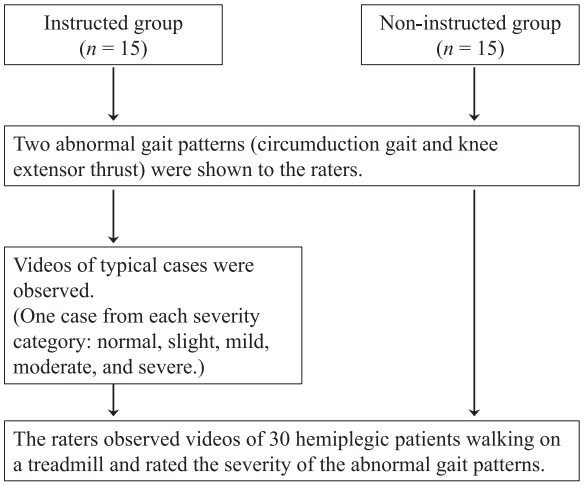

Figure 2. Procedure in this study.

of the two abnormal gait patterns were explained to both groups. The instructed group then viewed videos of typical cases selected by the four experienced physical therapists. The raters in the instructed group were allowed to watch the videos repeatedly so that they would be able to recognize the severity of the gait pattern abnormalities. Then the raters in both groups viewed videos of 30 hemiplegic patients while treadmill walking. The 30 hemiplegic patients were selected from the 50 hemiplegic patients by excluding the 5 patients who were used as typical cases and 15 patients with no relevant abnormalities (considered by the authors). The videos were projected onto a screen (182 cm in height, 244 cm in width). The duration of the video was 15 seconds for each case. Pausing, viewing in slow motion, or discussion were not allowed. The physical therapists then rated the severity of the abnormal gait pattern on a scale of five levels while viewing the same video a second time. Each group was examined on a different day, and this procedure was repeated for every abnormal gait pattern.

2.3 Data analysis

The severity categories were assigned numerical values from 1 (severe) to 5 (normal) for analysis. The median severity score was calculated for each patient (n = 30). To clarify the agreement for judging only the presence/absence of abnormality rather than the severity, we calculated the percentage of pairs showing that both raters judged a 5 (normal) or 1–4 (slight, mild, moderate, or severe) to the possible pairs in the instructed group and the non-instructed group. In addition, to determine the influence of clinical experience and instruction by viewing videos of typical cases beforehand on the inter-rater reliability of OGA, the agreement for each possible pair of raters in the four subgroups in the instructed and non-instructed groups was calculated using Cohen's kappa coefficient (κ) and weighted kappa ($w\kappa$) coefficient, and Spearman's rank correlation test and Wilcoxon's signed-rank test were performed. For the weighted kappa coefficient, one difference in the severity rating was weighted as 0.50. Statistical analysis was carried out using JMP 7 software (SAS Institute Inc., Cary, NC, USA). p-Values less than 0.05 were considered statistically significant.

This study was approved by the Institutional Review Board of Fujita Health University (13–069) and written informed consent was obtained from all the participants.

Results

The raters' median severity ratings for each case were evenly distributed for circumduction gait and knee extensor thrust (Figure 3). The raters in the instructed group were allowed to view the videos of typical cases repeatedly; however they did not need to watch the videos more than four times. The agreement in judging whether a patient had an abnormality or a normal gait was 81.3% in the non-instructed group and 83.2% in the instructed group for circumduction gait. For knee extensor thrust, agreement was 77.6% in the non-instructed group and 84.9% in the instructed group. Cohen's kappa coefficient and weighted kappa coefficient for the instructed and non-instructed groups

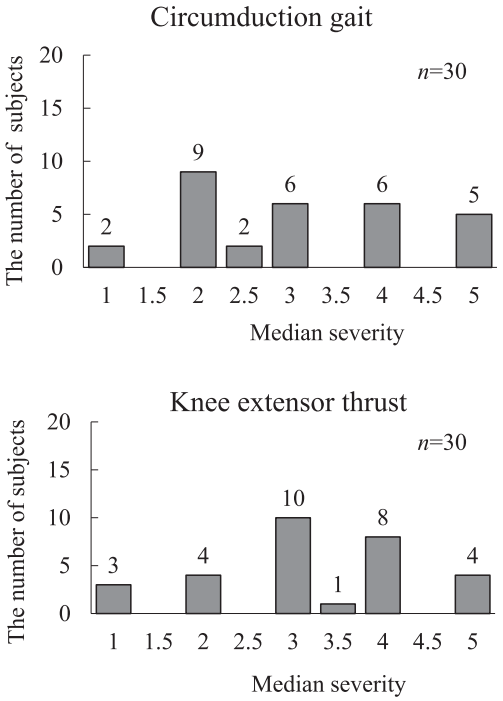

Figure 3. Raters' median rating for each case.
Histograms of the number of patients rated for gait abnormities by thirty physical therapists are shown. The horizontal axis indicates the median of the rated severity, and the vertical axis indicates the number of patients.

are shown in Table 1. The coefficient for the instructed group was higher than that for the non-instructed group for both abnormal gait patterns.

The proportion of pairs that had a significant rank correlation is shown in Table 2. Almost all pairs showed a significant rank correlation.

Table 3 shows the results of Spearman's rank correlation test and Wilcoxon's signed-rank test for each possible pair of raters. For circumduction gait, the proportion of pairs that had both a significant rank correlation and no significant difference in the observational ratings was the same in the subgroup with less than five years of clinical experience from both the non-instructed group and the instructed group. On the other hand, the proportion of pairs that had both a significant rank correlation and no significant difference in the observational ratings was greater in the subgroup from the instructed group that had five or more years of clinical experience compared to the subgroup from the non-instructed group that had five or more years of clinical experience. For knee extensor thrust, the proportion of pairs that had both a significant rank correlation and no significant difference in the observational ratings was greater in the subgroup with five or more years of clinical experience compared to those with less than five years of clinical experience in

Table 1. Agreement of the observational ratings.

	Circumduction gait		Knee extensor thrust	
	Kappa (range)	Weighted kappa (range)	Kappa (range)	Weighted kappa (range)
Instructed group				
≥5 years' experience	0.24 (0.00−0.49)	0.40 (0.00−0.64)	0.23 (0.00−0.46)	0.35 (0.13−0.51)
<5 years' experience	0.26 (0.07−0.68)	0.43 (0.15−0.78)	0.20 (0.03−0.38)	0.31 (0.06−0.43)
Non-instructed group				
≥5 years' experience	0.21 (0.00−0.61)	0.35 (0.04−0.67)	0.18 (0.00−0.49)	0.32 (0.02−0.63)
<5 years' experience	0.16 (0.00−0.61)	0.26 (0.00−0.71)	0.19 (0.00−0.59)	0.29 (0.04−0.64)

Table 2. Proportion of pairs with a significant rank correlation in the raters' judgments.

	Circumduction gait	Knee extensor thrust
Instructed group		
≥5 years' experience	28 (100%)	27 (96%)
<5 years' experience	21 (100%)	21 (100%)
Non-instructed group		
≥5 years' experience	28 (100%)	28 (100%)
<5 years' experience	21 (100%)	19 (90%)

These values indicate the number of pairs. The values in parentheses indicate the percentage of all pairs in that subgroup (n = 21 for <5 years' clinical experience, n = 28 for ≥5 years' clinical experience).

Table 3. Significance of Spearman's rank correlation test and Wilcoxon's signed-rank test in the pairs of raters.

Spearman's rank correlation coefficient	Wilcoxon's signed-rank test	Circumduction gait				Knee extensor thrust			
		Non-instructed group		Instructed group		Non-instructed group		Instructed group	
		<5 years' experience	≥5 years' experience	<5 years' experience	≥5 years' experience	<5 years' experience	≥5 years' experience	<5 years' experience	≥5 years' experience
Significant correlation	No significant difference	7 (33%)	6 (21%)	7 (33%)	15 (54%)	5 (24%)	12 (43%)	9 (43%)	20 (71%)
	Significant difference	14 (67%)	22 (79%)	14 (67%)	14 (46%)	14 (66%)	16 (57%)	12 (57%)	8 (25%)
No significant correlation	No significant difference	0	0	0	0	1 (5%)	0	0	1 (4%)
	Significant difference	0	0	0	0	1 (5%)	0	0	0

These values indicate the number of pairs. The values in parentheses indicate the percentage of all pairs in that subgroup (n = 21 pairs for <5 years' experience, n = 28 for ≥5 years' experience).

both the instructed and non-instructed groups. Moreover, for both gait abnormalities, the difference in the proportion of pairs that had both a significant rank correlation and no significant difference in the observational ratings was greater between instructed and non-instructed raters with five or more years of clinical experience compared to the difference between instructed and non-instructed raters with less than five years of clinical experience.

Discussion

This study investigated the inter-rater reliability of OGA when physical therapists rated the severity of characteristic abnormal gait patterns in stroke patients into five categories. In the gait rehabilitation of stroke patients, assessing the severity of abnormal gait patterns is important for determining the treatment method and evaluating the treatment's effectiveness [16, 25]. In the present study, the agreement in judging the presence or absence of these two abnormal gait patterns was high. However, the agreement in judging the severity of the abnormality was low. Typically, a κ value of 0.00−0.20 is taken to indicate slight reliability, 0.21−0.40 fair reliability, 0.41−0.60 moderate reliability, 0.61−0.80 substantial agreement, and 0.81−1.00 almost perfect agreement [26]. In the present study, the non-instructed group's inter-rater reliability for judging the severity was low (slight to fair) despite the raters' clinical experience. Some studies have investigated the inter-rater reliability of observational motion analysis for movement patterns other than gait. Tohara et al. found low inter-rater reliability when rating the function despite repeated observations, and this was because the observations were subjective and rater bias could easily occur [27]. However, in the present study, almost all pairs of raters had a significant rank correlation indicating that there was good agreement on the order of severity assigned to patients. This means that the raters did not agree on the specific severity for each patient but they did agree on the sequence assignment. The cause of the low inter-rater reliability for judging the severity was the difference in the base point of the subjective scale, and the range of each grade on the subjective scale between raters and in each rater.

In this study, we also examined the influence of viewing videos of typical cases prior to rating and the influence of clinical experience on the inter-rater reliability of OGA. Although highly reliable gait analysis is important for making clinical judgments, OGA is widely used in clinical settings [5]. Thus, a training program to improve its reliability is necessary [13, 20, 21]. By clarifying the influence of clinical experience and instruction by viewing videos of typical cases beforehand on the inter-rater reliability of OGA, the present study provides a suggestion for improving the reliability.

The number of pairs that showed a significant rank correlation as well as no significant difference in the ratings was larger in the instructed group than in the non-instructed group. This indicates that observing typical cases before making OGA judgments calibrated the raters' subjective scale. Moreover, the difference in the proportion of pairs that had both a significant rank correlation and no significant difference in the observational ratings between the instructed and non-instructed raters with five or more years of clinical experience was higher than that between the instructed and non-instructed raters with less than five years of clinical experience. These results indicate that clinical experience and instruction combined have an effect on inter-rater reliability. One study has shown that clinical experience has no influence on the ability to judge joint angle abnormalities using OGA but does have a positive influence on the rater's ability to predict the cause of abnormal gait patterns [28]. In the present study, clinical experience had a minor influence on the inter-rater reliability. However, clinical experience had a positive influence on understanding the instruction and calibrating the subjective scale.

Our results showed that observing typical cases prior to judgment improved the inter-rater reliability of OGA regardless of the clinical experience of the raters. In OGA, using videos, having a defined standard, ensuring consistency between subjective and objective measures, a training program, and a discussion after the observation all improve inter-rater reliability [21, 27, 29, 30]. Our results indicate that a training program observing and understanding typical cases before using OGA in the clinical setting could improve the reliability. Furthermore, clearly defined standard cases that demonstrate different severity levels of abnormal gait patterns are necessary to standardize the raters' subjective scale and improve the inter-rater reliability. In this present study, a subjective procedure was used for selecting typical cases for instruction. It would be better to use quantitative measures showing the difference from the normal with a three-dimensional motion analysis system to objectively select the cases.

There is no standard scale for judging the severity of abnormal gait patterns, so the severity was categorized into five levels in this study. The definitions of the levels were decided by referring to previous studies [7, 19, 20], and were considered not to overlap one another and to clarify the order or the severity. Although the reliability may be higher than five levels for low-severity gait abnormalities, the detection accuracy is lower and insufficient for use in clinical settings. Once the number of scale points exceeds seven, the difference between the points becomes considerably less clear [31]. This indicates that a five-level scale was appropriate. Furthermore, raters viewed videos of treadmill walking in the present study. One advantage of a video assessment is that it

allows repeated viewing of the gait pattern without inducing patient fatigue [14]. On the other hand, a disadvantage is that the two-dimensional data and lack of high-definition images may make it difficult to assess joints [6]. To resolve this problem, we ensured that anatomical landmarks were as visible as possible by attaching colored markers to the landmarks. Using a treadmill for gait analysis has other merits: walking speed can be controlled, multiple steps can be analyzed easily, and the subject walks within the same space [22]. In the present study, the videos of treadmill walking allowed us to place the subjects in the center of the screen and adjust the viewing speed despite the gait velocity. There are a few differences in joint angles and temporal factors between overground and treadmill gait [32, 33]. However, this was not the cause of the low agreements because the raters observed the same subjects and judged the severity. The limitation of this study is the lack of investigation into intra-rater reliability. Some studies have reported high intra-rater reliability but low inter-rater reliability for OGA [3, 4, 8]. Therefore, we consider that the cause of the low inter-rater reliability was not inconsistent judgment but a difference in the raters' subjective scales.

Conclusions

OGA is the most widely used method of clinical gait assessment. In the gait analysis of stroke patients, identification of abnormal gait patterns and judgment of the severity are important. This study demonstrated that physical therapists could detect abnormal gait patterns and rank their severity using OGA. However, agreement in judging the severity was low, caused by a difference in the raters' subjective scales. The inter-rater reliability of OGA was low even if the raters had substantial clinical experience, and it had a slight influence on the reliability of OGA. However, the observation of typical cases before making the assessment improved the inter-rater reliability regardless of clinical experience. A training program like this should be carried out in advance when physical therapists judge the severity of abnormal gait patterns based on observation in the clinical setting. Also, a clear standard for judgment showing the severity of abnormal gait patterns that is consistent with the objective measures using a technique such as three-dimensional motion analysis should be defined.

References

1. Robinson JL, Smidt GL. Quantitative gait evaluation in the clinic. Phys Ther 1981; 61: 351–3.
2. Stuberg WA, Colerick VL, Blanke DJ, Bruce W. Comparison of a clinical gait analysis method using videography and temporal-distance measures with 16-mm cinematography. Phys Ther 1988; 68: 1221–5.
3. Brunnekreef JJ, van Uden CJ, van Moorsel S, Kooloos JG. Reliability of videotaped observational gait analysis in patients with orthopedic impairments. BMC Musculoskelet Disord 2005; 17: 1–9.
4. Goodkin R, Diller L. Reliability among physical therapists in diagnosis and treatment of gait deviations in hemiplegics. Percept Mot Skills 1973; 37: 727–34.
5. Eastlack ME, Arvidson J, Snyder-Mackler L, Danoff JV, McGarvey CL. Interrater reliability of videotaped observational gait-analysis assessments. Phys Ther 1991; 71: 465–72.
6. Keenan AM, Bach TM. Video assessment of rearfoot movements during walking: a reliability study. Arch Phys Med Rehabil 1996; 77: 651–5.
7. Maathuis KG, van der Schans CP, van Iperen A, Rietman HS, Geertzen JH. Gait in children with cerebral palsy: observer reliability of physician rating scale and Edinburgh visual gait analysis interval testing scale. J Pediatr Orthop 2005; 25: 268–72.
8. Krebs DE, Edelstein JE, Fishman S. Reliability of observational kinematic gait analysis. Phys Ther 1985; 65: 1027–33.
9. Skaggs DL, Rethlefsen SA, Kay RM, Dennis SW, Reynolds RA, Tolo VT. Variability in gait analysis interpretation. J Periatr Orthop 2000; 20: 759–64.
10. Miyazaki S, Kubota T. Quantification of gait abnormalities on the basis of continuous foot-force measurement: correlation between quantitative indices and visual rating. Med Biol Eng Comput 1984; 22: 70–6.
11. Kawamura CM, de Morais Filho MC, Barreto MM, de Paula Asa SK, Juliano Y, Novo NF. Comparison between visual and three-dimensional gait analysis in patients with spastic diplegic cerebral palsy. Gait Posture 2007; 25: 18–24.
12. Lord SE, Halligan PW, Wade DT. Visual gait analysis: the development of a clinical assessment and scale. Clin Rehabil 1988; 12: 107–19.
13. McGinley JL, Goldie PA, Greenwood KM, Olney SJ. Accuracy and reliability of observational gait analysis data: judgments of push-off in gait after stroke. Phys Ther 2003; 83: 146–60.
14. Mackey AH, Lobb GL, Walt SE, Stott NS. Reliability and validity of the observational gait scale in children with spastic diplegia. Dev Med Child Neurol 2003; 45: 4–11.
15. Bernhardt J, Bate PJ, Matyas TA. Accuracy of observational kinematic assessment of upper-limb movements. Phys Ther 1988; 78: 259–70.
16. Olney SL, Griffin MP, McBride ID. Temporal, kinematic, and kinetic variables related to gait speed in subjects with hemiplegia: a regression approach. Phys Ther 1994; 74: 872–85.
17. Zimbelman J, Daly JJ, Roenigk KL, Butler K, Burdsall R, Holcomb JP. Capability of 2 gait measures for detecting response to gait training in stroke survivors: gait assessment and intervention tool and the tinetti gait scale. Arch Phys Med Rehabil 2012; 93: 129–36.
18. Hughes KA, Bell F. Visual assessment of hemiplegic gait following stroke: pilot study. Arch Phys Med Rehabil

1994; 75: 1100−7.
19. Rodriques AA, Black PO, Kile KA, Sherman J, Stellberg B, McCormik J, et al. Gait training efficacy using a home-based practice model in chronic hemiplegia. Arch Phys Med Rehabil 1996; 77: 801−5.
20. Tanikawa H, Ohtsuka K, Saitoh E, Itoh N, Yamada J, Muraoka Y, et al. Inter-rater reliability of visual inspection about abnormal gait. Sogo Rehabil 2010; 38: 1175−81. Japanese.
21. Bernhardt J, Bate PJ, Matyas TA. Training novice clinicians improves observation accuracy of the upper extremity after stroke. Arch Phys Med Rehabil 2001; 82: 1611−8.
22. Itoh N, Kagaya H, Saitoh E, Ohtsuka Y, Yamada J, Tanikawa H, et al. Quantitative assessment of circumduction, hip hiling, and forefoot contact gait using Lissajous figures. Jpn J Compr Rehabil Sci 2012; 3: 78−84.
23. Kerrigan DC, Frates EP, Rogan S, Riley PO. Hip hiking and circumduction: quantitative definitions. Am J Phys Med Rehabil 2000; 79: 247−52.
24. Perry J. Gait Analysis: Normal and Pathological Function. Thorofare: SLACK; 1992.
25. Tanikawa H, Ohtsuka K, Mukaino M, Inagaki K, Matsuda F, Teranishi T, et al. Quantitative assessment of retropulsion of the hip, excessive hip external rotation, and excessive lateral shift of the trunk over the unaffected side in hemiplegia using three-dimensional treadmill gait analysis. Top Stroke Rehabil 2016; 23: 311−7.
26. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159−74.
27. Tohara H, Chiba Y, Nakane A, Goto S, Ouchi Y, Teranaka S, et al. Inter-and intra-rater reliability in videofluoroscopic swallowing study. Jpn J Dysphagia Rehabil 2005; 9: 139−47. Japanese.
28. Bonkohara S, Yamamoto S. Reliability and accuracy of observational gait analysis. Rigakuryoho Kagaku 2008; 23: 747−52. Japanese.
29. McCullough GH, Wertz RT, Rosenbek JC, Mills RH, Webb WG, Ross KB. Inter- and intrajudge reliability for videofluoroscopic swallowing evaluation measures. Dysphagia 2001; 16: 110−8.
30. Stoeckli SJ, Huisman TA, Seifert B, Martin-Harris BJ. Interrater reliability of videofluoroscopic swallow evaluation. Dysphagia 2003; 18: 53−7.
31. Krosnick JA, Presser S. Number of points on rating scale. In: Marsden PV, Wright JD, editors. Handbook of Survey Research. 2nd ed. Bingley: Emerald Group Publishing; 268−75.
32. Alton F, Baldey L, Caplan S, Morrissey MC. A kinematic comparison of overground and treadmill walking. Clin Biomech 1998; 13: 434−40.
33. Riley PO, Paolini G, Della Croce U, Paylo KW, Kerrigan DC. A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects. Gait Posture 2007; 26: 17−24.