Inter-rater reliability of a newly developed gait analysis and motion score

Christina Dürregger[1], Klemens A Adamer[1], Michael Pirchl[1,2]
and Michael J Fischer[1,2,3] ✉

Abstract

Background: Gait analysis constitutes an essential part of orthopedic rehabilitation assessment. Previous studies indicate that observational-based gait analysis lacks reliability and requires extensive clinical training. Therefore, gait analysis in the clinical practice heavily relies on technical aids. The aim of the present study is to develop a reliable gait analysis assessment tool that can accurately assess clinically relevant gait cycle parameters in daily clinical practice. **Methods:** In this pilot study, a new gait analysis and motion score (GAMS), comprising 10 observational and 5 technically measured (e.g. pressure plate) gait parameters, was developed. The parameters were dichotomously operationalized, reflecting pathological versus physiological manifestations of the parameters. The rating algorithm was administered by 12 raters using videotaped treadmill sessions of 10 orthopedic subjects ($n = 120$ ratings). Inter-rater reliability was calculated using the intraclass correlation coefficient (ICC) and the percentage of rating agreement. **Results:** The mean (standard deviation (SD)) GAMS ratings ranged from 10.0 (1.1) to 21.5 (1.3) points. The overall GAMS ICC was 0.98 (95% confidence interval (CI) 0.96–1.00), whereas the ICC of observational parameters alone was 0.97 (95% CI 0.93–0.99). The mean (SD) percentage of rating agreement was 86.1% (3.3%). For the observational parameters, the mean (SD) rating agreement was 82.5% (4.5%). **Conclusion:** This new GAMS shows excellent overall inter-rater reliability for a continuum of functional gait statuses. The new score may be an appropriate clinical tool to objectively evaluate patients' gait patterns. Furthermore, the GAMS may find application as a clinician-reported outcome measure in orthopedic rehabilitation. Further studies are required to verify the validity and accuracy of the new GAMS and its functionality in assessing clinical changes in gait patterns.

Keywords

Clinician-reported outcome measures, gait analysis, inter-rater reliability, orthopedic rehabilitation

Introduction

Since the introduction of the World Health Organization's (WHO) International Classification of Functioning (ICF), outcome measurement has evolved as a core requirement for evidence-based rehabilitation.[1,2] There are currently a plethora of patient- and clinician-reported outcome measures that describe disease-specific aspects of a patient's health status.[3,4]

Based on the ICF, the WHO developed core sets for rehabilitation to standardize how "functioning" is clinically assessed.[5] The WHO ICF generic set contains seven key indicators that differentiate functioning levels among all patients.[5] One of these indicators is the category walking (d450 walking). The assessment of the functional walking status and its change over time should be considered as an indispensable part of outcome measurement in orthopedic rehabilitation.

---

[1] Vamed Rehabilitation Center Kitzbuehel, Kitzbühel, Austria
[2] Ludwig Boltzmann Institute for Rehabilitation Research, Vienna, Austria
[3] Department of Rehabilitation Medicine, Hanover Medical School, Hannover, Germany

Corresponding author:
Michael J Fischer, Vamed Rehabilitation Center Kitzbühel, Hornweg 32, A-6370 Kitzbühel, Austria.
Email: michael.fischer@reha-kitz.at

The functional walking status as an everyday activity (d450 walking) can be obtained by patient-reported outcome measures, such as questionnaires, whereas the underlying physiological function and gait pattern (b770 gait pattern functions) require clinician-reported outcome measures, including technical assessments, to be objectively quantified.[5]

Because the human gait is a highly complex movement, there are numerous parameters, which potentially reflect a patient's functional gait status.[6–9] The field of clinical gait analysis aims to identify relevant parameters and to reliably and objectively assess their manifestation at a certain point of time. Furthermore, gait analysis enables the identification of pathological gait patterns, which substantially guides diagnosis, intervention planning, and outcome prognosis.[9,10]

Unfortunately, clinical gait analysis increasingly relies on the employment of technical aids, the access of which often constitutes an economic hurdle in routine medical care.[1,2] Existing scores like the gait deviation index and the gait profile score are reliable measures to assess the functional gait status but require many measurements and calculations' while others like the Wisconsin gait scale are easy to assess but are specifically designed for the evaluation of gait quality for specific conditions like stroke patients.[11–13]

Previous studies indicate that observational gait analysis, without technical support, only shows slight to moderate reliability and requires extensive clinical training.[14–16] It has also been suggested that accurate assessment of treatment success cannot be achieved with videotape ratings alone.[17]

The aim of the present study was to establish a new gait analysis and motion score (GAMS) as a clinician-reported outcome measure for rehabilitation, based on the following required criteria: (1) it should distinguish pathological gait patterns from nonpathological movements; (2) all key aspects of the gait cycle should be taken into account; (3) it should have high objectivity; (4) it should be a user-friendly rating algorithm enabled by an appropriate parameter operationalization; (5) it should have the potential to assess the patient's functional gait status; and (6) it should detect minimal clinical changes. In this first study, we emphasize the evaluation of the inter-rater reliability of the new score.

Materials and methods

This study was previously approved by the institutional review board and complied with the Declaration of Helsinki. Each patient was informed that data from the case would be submitted for publication and gave their consent.

Parameter selection and operationalization

A multidisciplinary team of nine physical therapists and two medical doctors screened the standard reference gait

Table 1. Percentage of rating agreement.[a]

Sum of ratings	Percentage of agreement
0	100.0%
1	91.7%
2	83.3%
3	75.0%
4	66.7%
5	58.3%
6	50.0%
7	58.3%
8	66.7%
9	75.0%
10	83.3%
11	91.7%
12	100.0%

[a]Calculation of rating agreement in percent. Given 12 raters and a dichotomous item operationalization, a summative rating score of 0 and 12 represents 100% agreement, whereas a score of 6 represents the lowest agreement of only 50%.

analysis literature[6–9,17–21] and available scores[11,12,22–24] for relevant gait parameters to be used in the development of the new GAMS. Potential parameters were included in the GAMS development if they met the following criteria: (1) the clinical manifestation of a selected parameter can change over time; (2) it reflects a key element of the gait cycle; and (3) it is operationalizable into a dichotomous parameter, separating clinically relevant from nonpathological manifestations (Online Supplemental Table 1). In a Delphi process, out of 37 potential parameters identified, 15 items (including 10 items with a dual rating for being applied to both the left and right side) met the inclusion criteria and were further operationalized to dichotomous parameters and included in the GAMS (Figure 1). Each dichotomous parameter was rated by the numbers 0 or 1. A parameter score of 0 indicates a pathological parameter manifestation, whereas a parameter score of 1 indicates that it does not meet clinical significance. Due to the dual rating of 10 parameters, the final score ranges from 0 to 25 points. A lower total score at the end (sum of all parameters) was defined as a more pathological gait by definition.

Participant characteristics and rating procedure

Based on 12 available raters (see below), the required sample size was calculated according to Walter et al.[25] Assuming a reliability of 0.7 or higher as acceptable (H_0: $p = 0.7$) and expecting a reliability of 0.9 (H_1: $p = 0.9$), due to the dichotomous item operationalization, with a significance level of $\alpha = 0.05$, a power of $1 - \beta = 0.8$, and a possible dropout of 10%, a required sample size of 10 subjects was calculated resulting in a total number of 120 observations ($n = 12$ raters, $k = 10$ subjects, $kn = 120$ ratings).

A medical doctor and a physical therapist, who both did not participate in the GAMS parameter selection process, retrospectively chose videotaped treadmill sessions of 10

Gait Analysis and Motion Score (GAMS)						
No.	Parameter	Rating Instruction		Source	Rating	
					L	R
1	Initial Contact	0	flatfoot/toe contact	O/P		
		1	heel strike			
2	Heel lift	0	delayed/early/no contact	O/P		
		1	normal			
3	Hind foot	0	>5° deviation	O		
		1	<5° deviation			
4	Clearance	0	reduced/additional movements	O		
		1	appropriate			
5	Knee position sagittal plane	0	hyperextension/-flexion	O		
		1	appropriate			
6	Knee position frontal plane	0	deviation medial/lateral	O		
		1	no deviation			
7	Hip extension	0	no extension	O		
		1	>5° extension			
8	Trendelenburg's sign	0	>8° deviation	O		
		1	no deviation			
9	Duchenne's sign	0	>10° deviation	O		
		1	no deviation			
10	Trunk position	0	flexion	O		
		1	appropriate			
11	Center of pressure course	0	asymmetric	Z		
		1	symmetric			
12	Step length difference	0	>5% difference	Z		
		1	< 5% difference			
13	Stride width	0	<5 cm or >13 cm	Z		
		1	5-13 cm			
14	Speed	0	<2.5km/h or use of the handles	Z		
		1	>2.5 km/h without the handles			
15	Foot rotation	0	<5° or >15°	Z		
		1	5°-15°			
L = left leg, R = right leg O = videotaped observation, P = pressure plate Z = ZEBRIS FDM Treadmill report				Sum L/R		
				SCORE		

Figure 1. Gate analysis and motion score rating form.

orthopedic patients of the study site (of all recorded treadmill sessions generated as part of the medical examination between years 2014 and 2017), based on the following selection criteria: (1) the selected subjects should display various orthopedic conditions, which best reflect the patient population; (2) the level and type of the gait status impairment should differ among subjects; and (3) only subjects with no need of walking frames (crutches, etc.) may be selected.

The treadmill sessions were conducted as part of the patients' medical examination at the study site using a Zebris FDM-T treadmill ergometer with an integrated pressure plate and two SYNCLightCam modules for video recording (Zebris Medical GmbH, Isny, Germany) to measure their pressure distribution.

The videotaped sessions and Zebris treadmill protocols were retrospectively rated by 12 independent, blinded assessors (2 general medical doctors and 10 physical therapists, all of whom had at least 1-year experience in gait analysis, who were not involved in the subjects' medical treatment and did not conduct the treadmill sessions with the study

Table 2. Patient characteristics.[a]

ID no.	Sex	Age (years)	Medical condition	Side	GAMS rating (mean ± SD)
1	Male	57	Ankle fracture	Right	16.8 ± 2.0
2	Male	75	Lumbar spine fusion surgery	Left	18.6 ± 2.0
3	Male	59	Hip arthroplasty	Right	20.5 ± 1.8
4	Female	56	Hip arthroplasty	Left	20.7 ± 1.9
5	Male	64	Ankle fracture	Right	21.5 ± 1.3
6	Male	41	Hip fracture (femoral neck)	Right	17.4 ± 2.2
7	Male	39	Lumbar spine fusion surgery		11.9 ± 2.1
8	Male	36	Knee arthroplasty	Right	13.8 ± 2.1
9	Female	69	Ankle fracture	Right	14.4 ± 2.0
10	Female	69	Thoracic fusion surgery		10.0 ± 1.1

GAMS: gait analysis and motion score; SD: standard deviation.
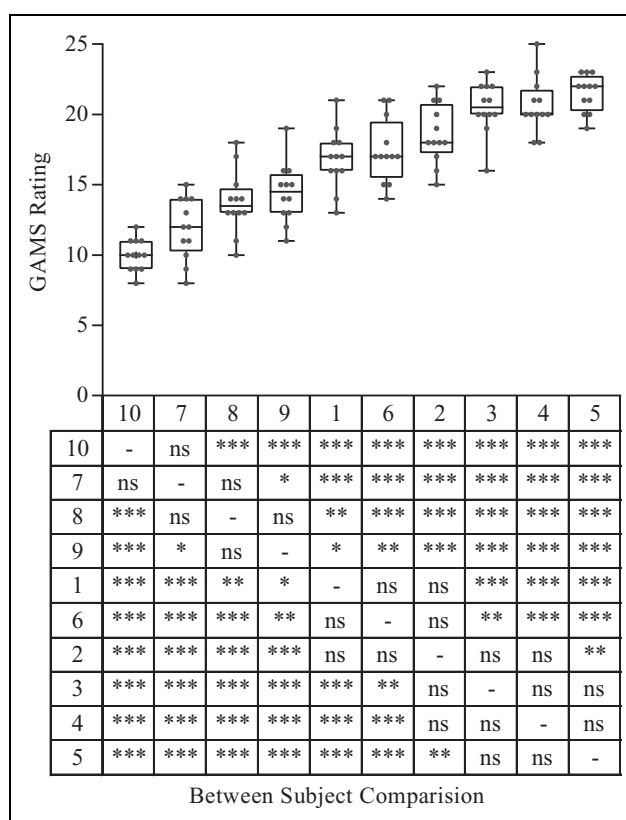[a]Patient characteristics and mean GAMS rating of 12 raters.

subjects), using the GAMS protocol (Figure 1, Online Supplemental Table 1). Every assessor underwent a 1-h GAMS training prior to rating the sessions. The blinded assessors reviewed a standardized Zebris FDM treadmill report, a video clip of three performed gait cycles (from the back and the left side), and a Zebris pressure plate report of each subject. In addition to rating each GAMS parameter, the blinded assessors also graded the subjective rating difficulty of each item, based on a score ranging from 1 to 5, with 1 = very easy and 5 = very difficult to rate.

Statistical analysis

All statistics were calculated using IBM® SPPS 25® (Armonk, New York, USA). Descriptive statistics was used. Inter-rater reliability was calculated using the intraclass correlation coefficient (ICC), as ICC has been shown to be a suitable measurement of reliability of dichotomous data (the presence or absence of a trait) in cases, where more than two raters are involved.[26–28] Inter-rater reliability was analyzed using a two-way mixed, average score intraclass correlation (C, k), with 95% confidence interval (CI) calculated. The ICCs were ranked by reliability classification of Koo and Li, whereby <0.50 had poor reliability, 0.50–0.75 had moderate reliability, 0.75–0.90 had good reliability, and >0.90 had excellent reliability.[29] Given that differences in means among assessors' ratings decrease the ICC, further analysis was necessary.[27] The percentage of rating agreement among the assessors for each parameter was calculated for each subject (joint probability of agreement), as given in Table 1. The significance of differences between subject ratings was calculated by one-way analysis of variance with a Tukey's multiple comparisons post hoc test.

Results

Among the 10 subjects (3 females and 7 males), their mean age was 52.8 ± 13.5 years and 7 had lower limb surgery and 3 had spinal surgery (Table 2).

	10	7	8	9	1	6	2	3	4	5
10	-	ns	***	***	***	***	***	***	***	***
7	ns	-	ns	*	***	***	***	***	***	***
8	***	ns	-	ns	**	***	***	***	***	***
9	***	*	ns	-	*	**	***	***	***	***
1	***	***	**	*	-	ns	ns	***	***	***
6	***	***	***	**	ns	-	ns	**	***	***
2	***	***	***	***	ns	ns	-	ns	ns	**
3	***	***	***	***	***	**	ns	-	ns	ns
4	***	***	***	***	***	***	ns	ns	-	ns
5	***	***	***	***	***	***	**	ns	ns	-

Between Subject Comparision

Figure 2. Mean GAMS ratings.

Mean GAMS ratings per subject

The mean GAMS ratings scored by the 12 assessors for each of the 10 subjects ranged from 10.0 (standard deviation (SD) 1.1) to 21.5 (SD 1.3) points (Table 2). Figure 2 shows the mean subject ratings ranked from highest to lowest and the significance of differences between subject ratings.

Inter-rater reliability

The ICC value of all 15 parameters (right and left leg rated separately) ranged from 0.47 to 1.00, as given in Table 3.

Table 3. Inter-rater reliability of the gait analysis and motion score.

Item	Parameter	ICC (95% CI) L/R	Reliability classification L/R	Rating agreement L/R	Rating difficulty
1	Initial contact	0.98 (0.96–1.00)/0.98 (0.96–0.99)	Excellent/excellent	94%/96%	2
2	Heel lift	0.85 (0.67–0.96)/0.84 (0.64–0.95)	Good/good	79%/79%	2
3	Hind foot	0.81 (0.58–0.95)/0.81 (0.57–0.94)	Good/good	77%/78%	3
4	Clearance	0.90 (0.78–0.97)/0.91 (0.80–0.97)	Good/excellent	88%/88%	2
5	Knee position sagittal plane	0.57 (0.03–0.87)/0.77 (0.47–0.93)	Moderate/good	74%/76%	2
6	Knee position frontal plane	0.94 (0.87–0.98)/0.93 (0.83–0.98)	Excellent/excellent	88%/87%	2
7	Hip extension	0.83 (0.62–0.95)/0.73 (0.40–0.92)	Good/moderate	73%/70%	3
8	Trendelenburg's sign	0.88 (0.72–0.96)/0.47 (0.19–0.84)	Good/poor	82%/88%	2
9	Duchenne's sign	0.76 (0.46–0.93)/0.92 (0.82–0.98)	Moderate/excellent	83%/87%	2
10	Trunk position	0.92 (0.82–0.98)	Excellent/excellent	93%	2
11	Center of pressure course	0.97 (0.93–0.99)	Excellent/excellent	99%	1
12	Step length difference	1.00 (0.99–1.00)	Excellent/excellent	99%	1
13	Stride width	0.99 (0.99–1.00)	Excellent/excellent	98%	1
14	Speed	0.99 (0.97–1.00)	Excellent/excellent	95%	1
15	Foot rotation	1.00 (0.99–1.00)/1.00 (1.00–1.00)	Excellent/excellent	99%/100%	1
	Observational GAMS	0.97 (0.93–0.99)	Excellent	82.5% ± 4.5%	
	Overall GAMS	0.98 (0.96–1.00)	Excellent	86.1% ± 3.3%	

CI: confidence interval; ICC: intraclass correlation; L: left leg; R: right leg; observational GAMS: items 1–10, reliability classification according to Koo and Li; rating difficulty 1–5 (1: very easy and 5: very difficult).

Only parameter 8 showed a poor ICC for the right side. The overall GAMS ICC was 0.98 (95% CI 0.96–1.00), whereas the ICC of just the observational parameters (parameters 1–10) was 0.97 (95% CI 0.93–0.99). The mean percentage of rating agreement was 86.1% (SD 3.3%). For just the observational parameters, the mean percentage of rating agreement was 82.5% (SD 4.5%).

Rating difficulty

The mean rating difficulty for all Zebris report parameters (parameters 11–15) was estimated as "very easy to rate" by the examiners (Table 3). Eight of the 10 observational parameters were estimated as "easy to rate," whereas parameter 3 (hind foot) and parameter 7 (hip extension) were estimated as "moderately difficult to rate."

Discussion

We introduce a new score for the clinical assessment of the functional gait status of orthopedic patients. Our data indicate that the new GAMS is a highly reliable screening tool, which considers key aspects of the gait cycle. The distribution of our mean subject scores (Figure 2) further indicates that different grades of gait pathology are captured by the GAMS. In our dataset, the mean subject ratings range from 10 to 21.5 points, which can be explained by the subjects' characteristics (Table 2). We did not expect lower scores, because only patients capable of performing a treadmill session and with no need of walking frames participated in the study. Higher scores may be expected in subjects with minimal to no gait pathology. Our findings indicate that the new GAMS may be able to detect minimal clinical changes. Therefore, the GAMS may be a suitable clinician-reported outcome measure for orthopedic rehabilitation. However, further studies are required to verify the validity of our new score.

Concerning the inter-rater agreement, the overall GAMS shows excellent inter-rater reliability, which also holds true for the observational parameters. This is in contrast to previous studies[2,6,7] and may likely be due to the dichotomous operationalization of these gait parameters. The examiners mainly estimated the parameters as "easy to rate." Parameters with moderate rating difficulty should be further operationalized in future versions of the score to obtain an easier rating algorithm.

In addition to lack of validation, a main methodological drawback was the one-sided videotaping of the treadmill sessions, which may have decreased the reliability of some parameters and increased their rating difficulty (e.g. hip extension). The main goal of further developments of the score should be the replacement of parameters, where technical support is required (parameters 11–15), with observational parameters. This would highly facilitate the use of the GAMS and make the assessment less costly. Furthermore, items with low inter-rater reliability and high rating difficulty should be further operationalized to enhance usability. In future studies, we aim to investigate the validity (preliminary data indicate high correlation with other patient-reported and clinician-reported outcome measures) and the test–retest reliability of the score to evaluate its ability to detect clinical changes. The current version of the GAMS may provide useful information to assist the planning of the rehabilitation process. The main benefit of our new score, compared to existing ones, is its easy dichotomous rating algorithm, which allows for an efficient

screening of patients without the requirement of an extensive training of the therapeutic staff.

Conclusion

The new GAMS is a highly reliable tool to assess the functional gait status of orthopedic patients. Pending further study, development, and validation, it may be applied as a screening instrument, as well as a clinician-reported outcome measure.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iD

Michael J Fischer [ID] https://orcid.org/0000-0002-6345-0813

Supplemental material

Supplemental material for this article is available online.

References

1. Abrams D, Davidson M, Harrick J, et al. Monitoring the change: current trends in outcome measure usage in physiotherapy. *Man Ther* 2006; 11: 46–53.
2. Reid JC, Kho ME, and Stratford PW. Outcome measures in clinical practice: five questions to consider when assessing patient outcome. *Curr Phys Med Rehabil Reps* 2015; 3: 248–254.
3. Howe TE, Dawson LJ, Syme G, et al. Evaluation of outcome measures for use in clinical practice for adults with musculoskeletal conditions of the knee: a systematic review. *Man Ther* 2012; 17: 100–118.
4. Oner C, Sadiqi S, Lehr AM, et al. The need of validated disease-specific outcome instruments for spine trauma. *J Orthop Trauma* 2017; 31(4): S33–S37.
5. Bickenbach JE. *ICF core sets: manual for clinical practice*. 1st ed. Toronto: Hogrefe Pub, 2012, pp. 14–21.
6. Carollo JJ and Matthews D. Strategies for clinical motion analysis based on functional decomposition of the gait cycle. *Phys Med Rehabil Clin N Am* 2002; 13: 949–977.
7. Dicharry J. Kinematics and kinetics of gait: from lab to clinic. *Clin Sports Med* 2010; 29: 347–364.
8. Kharb A, Saini V, Jain YK, et al. A review of gait cycle and its parameters. *IJCEM Int J Comput Eng Manage* 2011; 13: 78–83.
9. Baker RW. *Measuring walking: a handbook of clinical gait analysis*. 1st ed. London: Wiley, 2013, pp. 120–135.
10. Switaj TL and O'Connor FG. Chapter 43: gait analysis. In: Seidenberg PH and Beutler AI (eds) *The Sports Medicine Resource Manual*. Philadelphia: W.B. Saunders, 2008, pp. 536–542.
11. Schwartz MH and Rozumalski A. The gait deviation index: a new comprehensive index of gait pathology. *Gait Posture* 2008; 28: 351–357.
12. Baker R, McGinley JL, Schwartz MH, et al. The gait profile score and movement analysis profile. *Gait Posture* 2009; 30: 265–269.
13. Rodriquez AA, Black PO, Kile KA, et al. Gait training efficacy using a home-based practice model in chronic hemiplegia. *Arch Phys Med Rehabil* 1996; 77: 801–805.
14. Eastlack ME, Arvidson J, Snyder-Mackler L, et al. Interrater reliability of videotaped observational gait-analysis assessments. *Phys Ther* 1991; 71: 465–472.
15. Brunnekreef JJ, Van Uden CJ, Van Moorsel S, et al. Reliability of videotaped observational gait analysis in patients with orthopedic impairments. *BMC Musculoskeletal Disord* 2005; 6: 17.
16. Krebs DE, Edelstein JE, and Fishman S. Reliability of observational kinematic gait analysis. *Phys Ther* 1985; 65: 1027–1033.
17. Robinson JL and Smidt GL. Quantitative gait evaluation in the clinic. *Phys Ther* 1981; 61: 351–353.
18. Gage JR and Russman BS. *Gait analysis in cerebral palsy*. 1st ed. London: Cambridge University Press, 1991, pp. 21–32.
19. Götz-Neumann K. *Gehen verstehen: Ganganalyse in der Physiotherapie*. 3rd ed. Stuttgart: Georg Thieme Verlag, 2011, pp. 25–100.
20. Gronley JK and Perry J. Gait analysis techniques. rancho los amigos hospital gait laboratory. *Phys Ther* 1984; 64: 1831–1838.
21. Ludwig O.*Ganganalyse in der Praxis: Anwendung in Prävention, Therapie und Versorgung*. 2nd ed. Geislingen: C. Maurer, 2015, pp. 70–140.
22. Cretual A, Bervet K, and Ballaz L. Gillette gait index in adults. *Gait Posture* 2010; 32: 307–310.
23. Lord SE, Halligan PW, and Wade DT. Visual gait analysis: the development of a clinical assessment and scale. *Clin Rehabil* 1998; 12: 107–119.
24. Read HS, Hazlewood ME, Hillman SJ, et al. Edinburgh visual gait score for use in cerebral palsy. *J Pediatr Orthop* 2003; 23: 296–301.
25. Walter SD, Eliasziw M, and Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998; 17: 101–110.
26. Bartko JJ and Carpenter WT. On the methods and theory of reliability. *J Nerv Ment Dis* 1976; 163: 307–317.
27. Liu J, Tang W, Chen G, et al. Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Arch Psychiatry* 2016; 28: 115–120.
28. Shrout PE and Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420–428.
29. Koo TK and Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15: 155–163.