Reliability and Validity of Visual Assessments of Gait Using a Modified Physician Rating Scale for Crouch and Foot Contact

Tishya A. L. Wren, PhD,†‡ Susan A. Rethlefsen, PT,* Bitte S. Healy, MS, PT,* K. Patrick Do, BS,* Sandra W. Dennis, PT, MS,* and Robert M. Kay, MD*†*

Abstract: This study evaluates the visual assessment of gait using portions of the Physicians' Rating Scale (PRS). Thirty children with pathologic gait were evaluated "live" and using full- and slow-speed video. Interobserver reliability (weighted kappa) was 0.57 to 0.74 for foot contact, 0.69 to 0.71 for crouch, 0.30 to 0.40 for hip flexion, 0.57 to 0.65 for knee flexion, and 0.42 to 0.52 for dorsiflexion in stance. Intraobserver reliability (comparing the three conditions) was 0.50 to 0.78 for foot contact, 0.71 to 0.80 for crouch, 0.26 to 0.44 for hip flexion, 0.60 to 0.86 for knee flexion, and 0.39 to 0.61 for dorsiflexion. Observers were correct only 12% to 32% of the time when reporting less than 0 degrees of dorsiflexion and 0% to 29% of the time when reporting more than 20 degrees of hip flexion due to overestimation of hip flexion and underestimation of ankle dorsiflexion. These errors could lead some clinicians to presume the presence of contractures that do not actually exist. Visual assessment using the PRS does not appear to accurately measure what it is most commonly used to assess: ankle position in stance.

Key Words: gait analysis, crouch, equinus, toe walking

(*J Pediatr Orthop* 2005;25:646–650)

Computerized gait analysis is becoming more widely used in evaluation and treatment planning for children with gait abnormalities.[1,2] Even with the less sophisticated technology available decades ago, computerized motion analysis was shown to be more accurate than observational gait analysis for assessment of gait deviations.[3] Computerized gait analysis is not available in all communities, since it requires specialized equipment and specially trained personnel. When computerized gait analysis is not used, gait problems are most often assessed through observation in the clinic. The Physicians' Rating Scale (PRS) is an observational tool that has been used to evaluate gait and assess the outcome of botulinum toxin injection in children with cerebral palsy.[4–9] The most useful sections of the PRS are its assessments of crouch and foot contact patterns.[4] In the PRS, crouch is defined based on combined hip, knee, and ankle position during gait,[4,6] a definition that is limited in scope and may be a source of confusion for raters.

The reliability and validity of observational tools such as the PRS have not been established. Two studies examined the interrater reliability of three components of the PRS—crouch, foot contact, and recurvatum—based on review of videotape recordings.[4,5] The other components were deemed unlikely to be discriminatory in one study of botulinum toxin injection to the plantar flexors, and the knee position (recurvatum) component showed poor interrater reliability.[4] No studies have assessed the reliability of ratings performed during live observation, and the validity of the PRS has not been established. It is particularly important that the reliability and validity of live ratings be evaluated because most observational assessments are performed "live" in the clinic.

The current study was undertaken in effort to find a valid, reliable global assessment tool for use in the absence of, or in addition to, computerized gait analysis. The PRS was used with modifications made to minimize potential sources of interrater and intrarater error. The purposes of this study were to determine the agreement among modified PRS ratings based on live observation, full-speed video, and slow-motion video; to assess the interrater reliability of these ratings; to determine whether interrater reliability is improved by using video recordings and observing gait in slow motion; and to determine the accuracy of visual joint position ratings by comparing them to corresponding three-dimensional kinematic values.

METHODS

This study involved 30 children with disabilities (ages 5–20 years) referred to our laboratory for evaluation of gait abnormalities. The study was approved by the human subjects review board at our institution. All subjects provided written assent, and their parents provided written consent to participate in the study.

Each subject made three round trips on a 15-m path in the laboratory while being observed by four raters: three experienced gait laboratory physical therapists and a gait laboratory engineer. The subjects were videotaped walking

From *Children's Orthopaedic Center, Children's Hospital Los Angeles, Los Angeles, CA; †Department of Orthopaedics, University of Southern California, Los Angeles, CA; and ‡Departments of Radiology and Biomedical Engineering, University of Southern California, Los Angeles, CA.

Study conducted at Children's Hospital Los Angeles, Los Angeles, CA.
None of the authors received financial support for this study.

Reprints: Susan A. Rethlefsen, PT, Children's Orthopaedic Center, Children's Hospital Los Angeles, 4650 Sunset Blvd., MS 69, Los Angeles, CA 90027 (e-mail: srethlefsen@chla.usc.edu).

across the laboratory. For 13 subjects, the videotaped walks and the "live" observation walks were the same. For the other 17 subjects, the videotaped walks occurred during gait testing within 1 hour of the live observations. For these 17 subjects, the live and videotaped walks were different since their live walks were not videotaped.

Each rater scored each subject for foot contact pattern and crouch according to a modified version of the PRS[4] (Table 1). The "gait pattern" and "degree of crouch" sections of the PRS were used. In addition to rating crouch qualitatively, the raters gave separate ratings for components of crouch (minimum hip flexion, minimum knee flexion, and maximum ankle dorsiflexion during stance). This breakdown of the "crouch" category of the PRS was included to allow for variations of crouched gait (eg, excessive hip/knee flexion with ankle plantarflexion, or excessive hip/knee flexion and ankle dorsiflexion) that would not be included in the narrower definition of "crouch" in the PRS.[4,6] In addition, raters were instructed to evaluate minimum hip and knee flexion and maximum ankle dorsiflexion during the stance phase of gait. The PRS does not provide any guidelines regarding specific angles to be evaluated and when in the gait cycle they should be assessed.

Each rater provided scores during live observation. The same raters watched the videotape recordings at least 1 month later at normal speed and in slow motion and provided scores at each speed. One rater was not present for live observation of three subjects but provided video ratings for all subjects, including these three. For the video assessments, ratings were given first at normal speed and then at slow speed for all subjects to minimize recall of previous ratings. All subjects underwent computerized gait analysis using a VICON (Oxford Metrics, Oxford, UK) three-dimensional motion analysis system. The kinematic data were used to calculate minimum hip flexion, minimum knee flexion, and maximum ankle dorsiflexion angles during the stance phase of gait. These angles were compared with the corresponding scores provided by the raters to evaluate how well individual joint angles can be assessed through observation.

Weighted kappa scores were used to assess interrater agreement and the agreement in intrarater scores based on live observation and full- and slow-motion video. Agreement was considered almost perfect for kappa values 0.81 to 1.00, substantial for 0.61 to 0.80, moderate for 0.41 to 0.60, fair for 0.21 to 0.40, and slight for 0.00 to 0.20.[10] One-way analysis of variance (ANOVA) was used to determine whether the kinematic values obtained from computerized gait analysis differed when different visual ratings were given. The percentage of ratings for which the kinematic values fell within the intended range was also calculated.

TABLE 1. Physicians' Rating Scale for Foot Contact and Crouch and Scoring System for Hip, Knee, and Ankle Position During the Stance Phase of Gait

	Description	Score
Foot contact	Toe-toe	0
	Toe-heel	1
	Flat foot	2
	Occasional heel-toe	3
	Heel-toe	4
Crouch	Severe	0
	Moderate	1
	Mild	2
	None	3
Minimum hip flexion in stance	>20°	0
	5–20°	1
	0–5°	2
	<0°	3
Minimum knee flexion in stance	>20°	0
	5–20°	1
	0–5°	2
	<0°	3
Maximum dorsiflexion in stance	>20°	0
	5–20°	1
	0–5°	2
	<0° (plantarflexed)	3

RESULTS

Agreement Among Live Observation, Full-Speed Video, and Slow-Motion Video

The intrarater agreement among ratings based on live observation, full-speed video, and slow-speed video for each rater is presented in Table 2. There was substantial agreement in the ratings for crouch among the three observation techniques for all raters. For crouch, 179 of the 234 (77%) ratings were exactly the same between live observation and full-speed video, 172 of the 234 (74%) were the same between live observation and slow-motion video, and 199 of the 240 (83%) were the same between full-speed and slow-motion video. When the ratings differed, crouch tended to be rated as worse during video-based assessment (both full and slow speed) compared with live observation. This tendency was not due to deterioration of gait between live observation and subsequent videotaping, since similar results were found for the subset of subjects for whom the videotaped walks were the same as those observed live.

The ratings for foot contact were slightly more variable among observation conditions than the ratings for crouch (see Table 2). Agreement among ratings based on the different observation conditions was moderate to substantial for all raters. For foot contact, 163 of the 234 (70%) ratings were exactly the same between live observation and full-speed video, 131 of the 234 (56%) were the same between live

TABLE 2. Weighted Kappa Results Comparing Ratings for Live Observation, Full-Speed Video, and Slow-Motion Video for Each Rater

Rater	Foot Contact	Crouch	Hip	Knee	Ankle
1	0.7835	0.8019	0.4424	0.8583	0.6065
2	0.6715	0.7535	0.5896	0.6510	0.3935
3	0.5012	0.7055	0.2599	0.5982	0.5801
4	0.6731	0.7438	0.3507	0.6308	0.5742

observation and slow-motion video, and 144 of the 240 (60%) were the same between full- and slow-speed video. When the ratings differed, foot contact based on slow-motion video tended to be rated as less abnormal than those based on live observation or full-speed video.

For individual joint positions, agreement among ratings from live observation, full-speed video, and slow-motion video was substantial to almost perfect for the knee, moderate for the ankle, and fair to moderate for the hip (see Table 2). The ankle tended to be rated as less dorsiflexed on slow-motion video compared with live observation or full-speed video. The hip tended to be rated as less flexed on slow-motion video compared with full-speed video and, for three of the four raters, compared with live observation. There were no discernable differences in the ratings given for the knee between the three observation conditions.

Interrater Reliability

Agreement among raters for each of the observation conditions is presented in Table 3. Agreement among the four raters was substantial for crouch and moderate to substantial for foot contact. For crouch, most ratings were either the same for all raters (generally for scores of 1 or 4) or distributed among two adjacent scores (ie, 1 or 2, 2 or 3, or 3 or 4). The ratings for foot contact exhibited greater variability. There were no discernable differences between the ratings given by the engineer and those given by the physical therapists. Use of video did not improve interrater reliability for crouch. However, video (both full and slow speed) did appear to improve interrater reliability for foot contact, particularly when slow-motion video was used.

Agreement among raters using the different observation techniques to assess individual joint positions was moderate to substantial for the knee, moderate for the ankle, and fair for the hip (see Table 3). Use of video appeared to improve interrater reliability for the knee ratings, and slow-motion video improved interrater reliability for the ankle ratings. However, agreement for the hip deteriorated when video was used.

Accuracy/Validity

Table 4 shows the averaged kinematic values for hip, knee, and ankle position obtained from computerized gait analysis for each of the visual ratings given by the raters under each observation condition. ANOVA found significant differences among the kinematic values associated with the different visual ratings for all joints ($P < 0.0001$), indicating that true differences existed between the ratings assigned by the raters. The accuracy was greatest for minimum knee flexion in stance and was much less for minimum hip extension and maximum ankle dorsiflexion in stance. For minimum knee flexion in stance, average kinematic values fell within the expected range for all observation techniques. The same was true for the ankle for subjects judged to have greater than 5 degrees of dorsiflexion in stance. However, when the raters judged ankle dorsiflexion to be less than 5 degrees, they underestimated the amount of dorsiflexion actually present. This underestimation was less when the slow-motion video was used. Minimum hip flexion in stance was overestimated (judged to be more flexed than measured kinematically) in all categories except extension past neutral. Use of slow-speed video did not appear to improve the raters' subjective estimates of hip joint position.

TABLE 3. Interrater Agreement (Weighted Kappa Results Comparing Raters) for Each Observation Condition

Condition	Foot Contact	Crouch	Hip	Knee	Ankle
Live observation	0.5700	0.7133	0.4033	0.5650	0.4217
Video, full speed	0.6302	0.6863	0.3019	0.6540	0.4377
Video, slow motion	0.7391	0.6932	0.3635	0.6445	0.5176

The percentage of cases in which raters accurately assigned ratings for hip, knee, and ankle position (based on kinematic data) using the different observation techniques is given in Table 5. At the knee, raters appropriately assigned ratings in 73% to 88% of cases at extreme positions (>20 degrees of flexion and knee hyperextension), 61% to 75% of cases when between 5 and 20 degrees of flexion, and only 11% to 20% of cases when the knee was close to neutral. At the ankle, ratings were accurate in 68% to 80% of cases when ankle dorsiflexion was rated 5 degrees or greater, but much less accurate when less dorsiflexion was present. At the hip, extension beyond neutral was identified accurately in 80% to 83% of cases, but ratings were assigned accurately in only 0% to 29% of cases when the hip was flexed. In categories where accuracy was especially poor, greater accuracy was obtained when viewing the slow-speed video.

DISCUSSION

In this study, visual assessments of gait using the modified PRS showed the best interrater reliability for crouch, foot contact pattern, and knee position; interrater reliability was worse for hip and ankle position. These findings held true regardless of the observation technique used. The reasons for lower interrater agreement for the hip and ankle may include differences in interpretation of the instructions, even though more specific instructions were given than in previous studies involving the PRS. Although raters were instructed to assess maximum dorsiflexion in stance, some raters focused on single limb stance, while others looked for maximum dorsiflexion in stance regardless of when it occurred. Assessment of foot contact involves one point in the gait cycle (initial contact), and crouch is a global assessment of leg position, which may explain the more repeatable ratings for these items. Another possible source of variability is inconsistency in how the child walked.

Previous assessments of the PRS using videotape recordings found moderate agreement for crouch and moderate to substantial agreement for foot contact. Corry et al reported weighted kappa values of 0.55 for crouch and 0.50 to 0.67 for foot contact in evaluations of 20 children with spastic equinus and cerebral palsy.[4] Kay et al reported weighted kappa values of 0.72 before surgery and 0.82 after surgery for foot contact based on slow-motion video of 54 children with cerebral palsy who underwent surgical lengthening of the triceps surae.[5] These interrater results are similar to the agreement found in this study.

TABLE 4. Kinematic Results From Computerized Gait Analysis for Each Rating Category

Variable	Observation Condition	Visual Rating			
		>20	5–20	0–5	<0
Min. hip flexion	Live	4.2 ± 4.9 (n = 15)	0.1 ± 9.0 (n = 59)	−0.1 ± 10.0 (n = 69)	−6.8 ± 9.7 (n = 91)
	Fast	2.6 ± 7.4 (n = 12)	2.4 ± 11.4 (n = 72)	−2.8 ± 8.1 (n = 91)	−8.4 ± 7.5 (n = 65)
	Slow	10.0 ± 12.6 (n = 14)	1.5 ± 10.8 (n = 49)	−0.2 ± 7.0 (n = 79)	−8.1 ± 7.6 (n = 98)
Min. knee flexion	Live	30.2 ± 14.1 (n = 40)	13.3 ± 9.4 (n = 64)	2.1 ± 8.2 (n = 93)	−5.7 ± 7.3 (n = 37)
	Fast	34.5 ± 12.0 (n = 33)	14.7 ± 8.3 (n = 70)	2.0 ± 7.0 (n = 98)	−7.3 ± 6.8 (n = 39)
	Slow	31.5 ± 12.8 (n = 39)	15.3 ± 8.0 (n = 65)	2.1 ± 6.8 (n = 85)	−6.3 ± 6.6 (n = 51)
Max. ankle dorsiflexion	Live	22.5 ± 9.7 (n = 34)	16.4 ± 6.1 (n = 98)	12.9 ± 7.7 (n = 45)	7.1 ± 6.1 (n = 57)
	Fast	24.2 ± 5.8 (n = 43)	16.1 ± 5.4 (n = 98)	12.0 ± 6.5 (n = 48)	4.8 ± 6.1 (n = 51)
	Slow	23.8 ± 5.4 (n = 47)	15.9 ± 5.6 (n = 114)	8.8 ± 5.3 (n = 42)	3.7 ± 6.5 (n = 37)

Data are given as mean ± SD.

Our results suggest that visual assessments of knee position are reasonably accurate. However, when maximum knee extension in stance was estimated to be between 0 and 5 degrees during live gait, this assessment agreed with sagittal kinematic data in only 11% of cases. Visual assessments of hip and ankle position during gait were also reasonably accurate if close to normal (hip extension past neutral and dorsiflexion beyond 5 degrees). However, the modified PRS does not appear to be an accurate tool for assessing excessive hip flexion and equinus, deviations commonly seen in pathologic gait. In fact, the common errors in visual assessment of the hip (overestimating hip flexion) and ankle (overestimating equinus) could lead some clinicians to presume the presence of contractures that do not actually exist. Therefore, visual assessment of hip and ankle position should not be considered an adequate substitute for computerized gait analysis testing, even when the assessment is done in a structured manner as with the PRS.

Raters had more difficulty categorizing knee position when it was close to normal (0–5 degrees), possibly because the 0-to-5-degree range is small, leaving little room for error. The other rating categories involved either a large range (5–20 degrees) or an open-ended range (<0 degrees or >20 degrees), leaving more room for error.

The results comparing intrarater ratings from the different observation conditions showed similar trends to those outlined above. The best agreement was observed for crouch, foot contact, and knee position for all raters.

Use of slow-motion video was helpful in some cases. It improved interrater agreement for ratings of foot contact and knee and ankle position but did not improve agreement for crouch or hip position. Accuracy of joint position ratings was improved slightly by use of the video.

This study was conducted under "ideal" conditions. The raters were experienced in gait assessment, having worked in a gait laboratory for 5 to 22 years each. Live observations were done in a quiet, open space with the subject walking three times down the length of the laboratory and back. Accuracy and reliability of visual assessments are likely to be lower for raters with less experience in gait assessment, or when assessments are done based on fewer walks in a crowded, noisy clinic.

In summary, this study suggests that visual assessment using the modified PRS is valid and reliable for assessing crouch, foot contact pattern, and maximum knee extension in stance. It appears to be less appropriate for assessing other deviations frequently seen in pathologic gait, such as excessive hip flexion or equinus. In some cases, use of slow-motion video can improve the accuracy and reliability of observational gait assessments, particularly at the ankle. However, such assessments are not an adequate substitute for computerized gait analysis testing due to their inaccuracy.

TABLE 5. Frequency and Percentage of Ratings in Which Kinematic Results from Computerized Gait Analysis Fell Within Range of Assigned Subjective Rating

Variable	Observation Condition	Visual Rating			
		>20	5–20	0–5	<0
Min. hip flexion	Live	0/15 (0%)	13/59 (22%)	10/69 (14%)	73/91 (80%)
	Fast	0/12 (0%)	19/72 (26%)	13/91 (14%)	53/65 (82%)
	Slow	4/14 (29%)	14/49 (29%)	15/79 (19%)	81/98 (83%)
Min. knee flexion	Live	29/40 (73%)	39/64 (61%)	10/93 (11%)	30/37 (81%)
	Fast	27/33 (82%)	46/70 (66%)	16/98 (16%)	34/39 (87%)
	Slow	31/39 (79%)	49/65 (75%)	17/85 (20%)	45/51 (88%)
Max. ankle dorsiflexion	Live	23/34 (68%)	67/98 (68%)	7/45 (16%)	7/57 (12%)
	Fast	31/43 (72%)	75/98 (77%)	7/48 (15%)	11/51 (22%)
	Slow	35/47 (74%)	91/114 (80%)	13/42 (31%)	12/37 (32%)

REFERENCES

1. Kay RM, Dennis S, Rethlefsen S, et al. Impact of postoperative gait analysis on orthopaedic care. *Clin Orthop*. 2000;374:259–264.
2. Kay RM, Dennis S, Rethlefsen S, et al. The effect of preoperative gait analysis on orthopaedic decision making. *Clin Orthop*. 2000;372:217–222.
3. Saleh M, Murdoch G. In defence of gait analysis. Observation and measurement in gait assessment. *J Bone Joint Surg [Br]*. 1985;67:237–241.
4. Corry IS, Cosgrove AP, Duffy CM, et al. Botulinum toxin A compared with stretching casts in the treatment of spastic equinus: a randomised prospective trial. *J Pediatr Orthop*. 1998;18:304–311.
5. Kay RM, Rethlefsen SA, Ryan JA, et al. Outcome of gastrocnemius recession and tendo Achilles lengthening in ambulatory children with cerebral palsy. *J Pediatr Orthop B*. 2004;13:92–98.
6. Koman LA, Mooney JF 3rd, Smith B, et al. Management of cerebral palsy with botulinum-A toxin: preliminary investigation. *J Pediatr Orthop*. 1993;13:489–495.
7. Koman LA, Mooney JF 3rd, Smith BP, et al. Botulinum toxin type A neuromuscular blockade in the treatment of lower extremity spasticity in cerebral palsy: a randomized, double-blind, placebo-controlled trial. Botox Study Group. *J Pediatr Orthop*. 2000;20:108–115.
8. Koman LA, Brashear A, Rosenfeld S, et al. Botulinum toxin type A neuromuscular blockade in the treatment of equinus foot deformity in cerebral palsy: a multicenter, open-label clinical trial. *Pediatrics*. 2001;108:1062–1071.
9. Slawek J, Klimont L. Functional improvement in cerebral palsy patients treated with botulinum toxin A injections: preliminary results. *Eur J Neurol*. 2003;10:313–317.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.

